Comparative genomics of orobanchaceous species with different parasitic lifestyles reveals the origin and stepwise evolution of plant parasitism

Yuxing Xu[a], Jingxiong Zhang[a], Canrong Ma[a,b], Yunting Lei[a], Guojing Shen[a], Jianjun Jin[c], Deren A. R. Eaton[c], Jianqiang Wu[a,b,*]

[a] Department of Economic Plants and Biotechnology, Yunnan Key Laboratory for Wild Plant Resources, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China; [b] CAS Center for Excellence in Biotic Interactions, University of Chinese Academy of Sciences, Beijing 100049, China; [c] Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10027, USA

*Jianqiang Wu

Email: wujianqiang@mail.kib.ac.cn

Author Contributions: Y.X. and J.W. designed research; J.Z., C.M., and G.S. prepared plant tissues, DNA samples, and RNA samples; Y.X., J.W., Y.L., G.S., and J.J. analyzed data; Y.X., J.W., J.J., and D.E. wrote the paper.

Keywords: Genomics, Orobanchaceae, Parasitism, Whole-Genome Duplication, Gene Loss, Gene family expansion.

Abstract

Orobanchaceae is the largest family of parasitic plants, containing autotrophic and parasitic plants with all degrees of parasitism. This makes it by far the best family for studying the origin and evolution of plant parasitism. Here we provide three high-quality genomes of orobanchaceous plants, the autotrophic *Lindenbergia luchunensis* and the holoparasitic plants *Phelipanche aegyptiaca* and *Orobanche cumana*. Phylogenomic analysis of these three genomes together with those previously published and the transcriptomes of other orobanchaceous species, created a robust phylogenetic framework for Orobanchaceae. We found that an ancient whole-genome duplication (WGD; about 73.48 Mya), which occurred earlier than the origin of Orobanchaceae, might have contributed to the emergence of parasitism. However, no WGD events occurred in any lineage of orobanchaceous parasites except for *Striga* after divergence from their autotrophic common ancestor, suggesting that, in contrast to previous speculations, WGD is not associated with the emergence of holoparasitism. We detected evident convergent gene loss in all parasites within Orobanchaceae and between Orobanchaceae and dodder *Cuscuta australis*. The gene families in the orobanchaceous parasites showed a clear pattern of recent gains and expansions. The expanded gene families are enriched in functions related to the development of the haustorium, suggesting that recent gene family expansions may have facilitated the adaptation of orobanchaceous parasites to different hosts. This study illustrates a stepwise pattern in the evolution of parasitism in the orobanchaceous parasites, and will facilitate future studies on parasitism and the control of parasitic plants in agriculture.

Introduction

Most plants are autotrophs. However, about 4750 angiosperm species are parasitic (Nickrent, 2020), and completely or partly depend on plant hosts to survive, from which they extract water and nutrients. Some parasitic plants still have chlorophyll and photosynthesize; these are known as hemiparasitic plants. Others can no longer photosynthesize, and are known as holoparasites. According to whether hosts are needed for the completion of life cycles, parasitic plants can also be classified as facultative and obligate (Westwood et al., 2010). In angiosperms, parasitic plants independently evolved 12 or 13 times from autotrophic plants (Westwood et al., 2010). In eight parasitic plant lineages, hemiparasitic species became extinct, leaving only holoparasites remaining (Westwood et al., 2010). However, the Orobanchaceae (Lamiales), the largest family of parasitic plants, comprises facultative hemiparasitic, obligate hemi- and obligate holoparasitic species (Supplemental Figure 1) (Westwood et al., 2010; Joel et al., 2013). Orobanchaceae even contains three genera of autotrophs: *Rehmannia*, *Triaenophora*, and *Lindenbergia*, with *Lindenbergia* being the closest relative of the parasitic lineages (Li et al., 2019). Furthermore, some of the orobanchaceous parasites are economically important, and are devastating pests in agriculture. For example, the annual economic loss caused by the witchweed (*Striga*) and broomrapes (*Orobanche* and *Phelipanche*) is over 10 billion US dollars in Asia and Africa (Twyford, 2018). Orobanchaceae is thus the best extant system in which to study the evolutionary transition to plant parasitism, as well as to develop genomic resources for future agricultural applications (Westwood et al., 2010).

The phylogenetic relationships among the major lineages of Orobanchaceae remain unclear. Previous research by McNeal et al. (2013) split the family into six major clades based on three nuclear and two plastid markers, while a more recent study by Li et al. (2019) analyzed five additional markers and recognized eight major clades, with incongruent relationships compared to the previous treatment. Following the Angiosperm Phylogeny Group nomenclature, the Li et al. (2019) treatment recognized Rehmannieae, Lindenbergieae, Cymbarieae, Orobancheae, *Brandisia*, Rhinantheae, Buchnereae, and

Pedicularideae; furthermore, in Li et al. (2019), the holoparasitic clade Orobancheae was placed as the sister to all other parasitic Orobanchaceae, and the authors proposed that holoparasitism may have evolved early in Orobanchaceae.

The evolution of parasitism requires major morphological and physiological innovations to perceive and recognize potential host plants and to form parasitic connections. Parasitic plants have evolved a unique organ, the haustorium, using which they can attach and penetrate host tissues (Yoshida et al., 2016). The haustoria of many hemiparasites, such as *Triphysaria* and *Striga*, fuse with the xylem of host plants, enabling these hemiparasites to obtain water and nutrients from the host xylem. Some holoparasites fuse their haustorial xylem and phloem with that of their hosts, and symplastic continuity between parasite and host phloem tissues occurs by means of plasmodesmatal connections (Yoshida et al., 2016). The obligate hemiparasites and holoparasites in the Orobanchaceae can perceive strigolactones released from host roots, which activate the germination process in the parasitic plants (Cook et al., 1966; Cook et al., 1972). Recent studies have revealed that in obligate parasites of this family, strigolactones released by the host roots are perceived by the gene *KAI2d*, which appears to have evolved from the karrikin receptor gene *KAI2* by duplication and neofunctionalization. The interaction between the strigolactones and *KAI2d* enables the parasites to detect the existence of host plants in the vicinity (Conn et al., 2015; Tsuchiya et al., 2015; Toh et al., 2015). The evolution and development of haustoria have drawn much attention (Yoshida et al., 2016). Yang et al. (2015) comparatively analyzed the transcriptomes of haustorial tissues from three orobanchaceous species *Triphysaria versicolor*, *Striga hermonthica*, and *Phelipanche aegyptiaca* and found a core set of parasitism genes, many of which underwent gene duplications or reconfiguration of transcriptional regulation. Previous studies on *Cuscuta australis* (Convolvulaceae) (Sun et al., 2018), which is a shoot parasitic plant, the root parasite *Striga asiatica* (Orobanchaceae) (Yoshida et al., 2019), and an endoparasite *Sapria himalayana* (Rafflesiaceae) (Cai et al., 2021) suggested that gene loss is common in the evolutionary histories of parasitic plants. Wicke et al. (2016) studied the plastomes of various species in the Orobanchaceae, and found that the plastomes of orobanchaceous parasites, particularly those of holoparasites,

had clearly experienced large-scale gene loss, leading to most of the photosynthesis-related genes being missing. Gene loss, such as loss of the plastome *ndh* genes, was detected even in the orobanchaceous hemiparasites (Wicke et al., 2016). The degree of gene loss is generally considered to be positively associated with the degree of parasitism (Cai et al., 2021), as a greater degree of parasitism requires higher dependency on the host and fewer physiological processes are needed by the parasite itself. Gene loss may therefore enable the parasite to synchronize its own physiology with that of the host, probably resulting in better fitness (Shen et al., 2020).

To date, genomes from five orobanchaceous species have been released: the autotrophic *Rehmannia glutinosa* (Ma et al., 2021), the obligate hemiparasite *Striga asiatica* (Yoshida et al., 2019), and the facultative hemiparasites *Euphrasia arctica* (Becher et al., 2020), *Phtheirospermum japonicum* (Cui et al., 2020) and *Pedicularis cranolopha* (Jin and Eaton, 2022). However, the genomes of orobanchaceous holoparasites are still lacking. In this study, we sequenced and annotated new reference genomes of the autotroph *Lindenbergia luchensis* and two economically important holoparasitic pests *Orobanche cumana* and *Phelipanche aegyptiaca*. Here we present the largest comparative analysis of genomes of orobanchaceous species to date. A robust phylogeny was reconstructed and the genome evolution of these orobanchaceous species was analyzed in detail, including the histories of genome duplication, gene family expansion, and gene loss, as well as the evolution of haustorium-related genes. Our analyses yielded a new scenario for the origin and evolution of parasitism in Orobanchaceae. This study fills the gap in our understanding of the evolution of plant parasitism and provides novel genomic resources for the most diverse and economically important parasitic plant clade.

Results

Sequencing, assembly, and annotation of the genomes of *Lindenbergia luchunensis*, *Orobanche cumana*, and *Phelipanche aegyptiaca*

Using a Nanopore platform, 36.47Gb data were generated for *Lindenbergia luchunensis* (Llu), and 158.30 Gb and 396.00 Gb data were generated respectively for *Orobanche*

137 *cumana* (Ocu) and *Phelipanche aegyptiaca* (Pae) on a PacBio Sequel II platform in the
138 CLR mode. Short reads (38.19 Gb, 141.20 Gb, and 329.33 Gb, respectively) were
139 produced for Llu, Ocu, and Pae using an MGI-SEQ 2000. Genome survey analysis using
140 the short reads indicated low degrees of heterozygosity (Supplemental Table 1). Long
141 reads were thus directly used for *de novo* assembly of the genomes, and the short reads
142 were used to correct the genome assemblies, forming contigs. The total length of the Pae
143 genome assembly was 3876.90 Mb (N50 9.97 Mb). Using the Hi-C data, we obtained
144 chromosome-level Llu and Ocu genomes (Llu: 20 pseudochromosomes, 212.21 Mb; Ocu:
145 19 pseudochromosomes, 1417.94 Mb, Supplemental Table 1). The accuracy of each of
146 the three genomes was found to exceed 99.9% (see Supplemental Methods and
147 Supplemental Table 1). We then annotated repetitive sequences, including simple repeats,
148 transposable elements, and RNA genes. In the genomes of Llu, Ocu, and Pae, the
149 repetitive sequences occupied 30.72%, 79.82%, and 84.40%, respectively (Supplemental
150 Table 2). Mapping the RNA-seq data (Supplemental Table 3) from different organs to the
151 respective genomes resulted in 29,669, 42,525, and 50,484 protein-coding genes in Llu,
152 Ocu, and Pae, respectively. Benchmarking universal single-copy orthologs (BUSCO)
153 analysis indicated that the completeness of these three genomes was 97.1% (Llu), 80.5%
154 (Ocu), and 76.3% (Pae) (Supplemental Table 4). The low completeness values of the
155 genes in Ocu and Pae genome is due to the large-scale gene losses seen in Ocu and Pae
156 (see below for analysis), which is associated with their holoparasitic lifestyle.
157
158 **Phylogenetic relationships in Orobanchaceae based on nuclear gene sequences**
159 Previous phylogenetic analyses suggested that there are eight major clades in the
160 Orobanchaceae, but that the phylogenetic relationships among the clades were still
161 uncertain (Schneeweiss, 2013; McNeal et al., 2013; Li et al., 2019). The nuclear genes
162 from the 19 available genomes/transcriptomes of Orobanchaceae species (Supplemental
163 Table 5), which cover seven of the eight clades of Orobanchaceae, and three outgroup
164 species (Supplemental Table 5) were used in our phylogenetic analysis (see Supplemental
165 Methods). Finally, after filtering out those orthogroups (OGs) which had low quality
166 and/or high copy numbers, 907 OGs were selected for phylogenetic reconstruction.
167 Coalescent and concatenation methods yielded identical tree topologies, with 100%

support for all the nodes (Figure 1). In this phylogenetic tree, which is consistent with previous findings (Wolfe et al., 2005; Schneeweiss, 2013; McNeal et al., 2013; Li et al., 2019), the autotrophic lineage Rehmannieae is sister to all the other orobanchaceous lineages, and another autotrophic lineage Lindenbergieae is the sister clade of all the parasitic plant lineages (Figure 1). Within the group of parasitic plants, Cymbarieae is the sister group to all the remaining lineages, which differs from the results of Li et al. (2019), which was based only on a few nuclear genes. Buchnereae and Orobancheae are sister clades, as are Pedicularideae and Rhinantheae (Figure 1). The topology of these four clades is distinct from the results of all previous studies (Wolfe et al., 2005; Schneeweiss, 2013; McNeal et al., 2013; Li et al., 2019), although it is similar to the topologies based on the *matK* and *rps2* genes (McNeal et al., 2013).

Several previous studies indicated that parasitic plants have higher nucleotide substitution rates than typical autotrophic plants (Lemaire et al., 2011; Bromham et al., 2013); however, in our phylogenetic tree, only the branch length of the holoparasitic plant *Aeginetia indica* in the clade Buchnereae is exceptionally long, while the branches of all holoparasitic plants in the clade *Orobancheae* were similar to those of the autotrophic plants (Supplemental Figure 2). We calculated the K_s values of all the lineages in Orobanchaceae, and indeed no branches showed high substitution rates (Supplemental Figure 3).

Based on this tree, we estimated the divergence times for all the nodes with a Bayesian molecular dating analysis. The estimated origin of the Orobanchaceae is about 46.51 million years ago (Mya; 95% highest posterior density (HPD), 37.91-55.39 Mya), and the emergence of parasitism predates 38.58 Mya (95% HPD, 31.27-45.81 Mya). Our analysis also indicated that holoparasites evolved independently at least three times in the Orobanchaceae (Figure 1), a result similar to that of Fu et al. (2017), and of these three lineages of holoparasites, Orobancheae appeared the earliest, at least 33.95 Mya (95% HPD, 27.41-40.43 Mya).

Genome evolution in the Orobanchaceae

199 Genome sizes and chromosome numbers vary in the Orobanchaceae, and they seem to be
200 positively associated with the degrees of parasitism (Joel et al., 2013; Lyko and Wicke,
201 2021). It is plausible that multiple rounds of WGD events occurred during the evolution
202 of orobanchaceous species (Joel et al., 2013; Lyko and Wicke, 2021). To
203 comprehensively reveal the WGD events in the Orobanchaceae, six species with
204 reference genomes (Pae, Ocu, and Llu from this study; *Striga asiatica* (Sas),
205 *Phtheirospermum japonicum* (Pja), and *Pedicularis cranolopha* (Pcr) from
206 published/released data) and four outgroup species (*Mimulus guttatus* (Mgu; Lamiales),
207 *Solanum lycopersicum* (Solanales, lamiids), *Coffea canephora* (Gentianales, lamiids), and
208 *Arabidopsis thaliana* (Brassicales, rosids)) were chosen for colinear fragment
209 comparisons. Instead of the commonly used methods based on colinear fragments
210 obtained by homology searches, an analysis based on "orthologous colinear fragments
211 (OCFs)" was specifically designed for this study (see Supplemental Methods and
212 Supplemental Figure 4), which enabled us to investigate all possible WGD events during
213 the evolution of the Orobanchaceae.
214
215 Our analysis based on OCFs from all species pairs (Figure 2) indicated that the Sas
216 lineage experienced an independent WGD event (named α^B event; Figure 2), which was
217 consistent with a previous study (Wickett et al., 2011; Yoshida et al., 2019). However, no
218 WGD events were detected in any other orobanchaceous lineages, including those
219 containing Pae and Ocu (Figure 2). The ancestor of Orobanchaceae did not experience
220 any independent WGD events, even though there was a WGD event during the evolution
221 of the common ancestor of Orobanchaceae and *Mimulus*, which is named the β^L event
222 (Figure 2). Notably, the detected WGD events in the outgroup species are consistent with
223 previous reports (Simillion et al., 2002; Sato et al., 2012; Denoeud et al., 2014),
224 supporting the reliability of our novel WGD analysis pipeline. Next, we selected four
225 nodes (S1-S4 in Figure 2) within the Lamiales clade, and all the orthologous genes
226 originating from these speciation events were selected for calculation of the K_s values
227 (Supplemental Figure 5). We found that the peak values of the K_s distributions had a
228 linear relationship with the divergence times of these nodes (R^2 = 0.9926, Figure 2).
229 Therefore, the K_s values from the paralogous genes resulting from the α^B and β^L events

were used to infer the times of these two events: α^B and β^L occurred at 32.64 ± 6.58 and 73.48 ± 6.58 Mya, respectively (Figure 2). The α^B event likely occurred in the Buchnereae soon after it diverged from the Orobancheae at 35.78 Mya (95% HPD, 28.97-42.52 Mya; Figures 1 and 2). The occurrence of the β^L event was close to the Cretaceous-Paleocene boundary, and the mass extinction event at this time (Schulte et al., 2010) when many plant lineages experienced WGD events (Fawcett et al., 2009; Wu et al., 2020). We speculated that the β^L event may have played an important role in enabling the common ancestor of Mgu and the Orobanchaceae to survive the extreme environmental changes at the Cretaceous-Paleocene boundary.

Knowledge of chromosome numbers is important for understanding the genome evolution of eukaryotes. The chromosome numbers in the Orobanchaceae vary (Joel et al., 2013; Lyko and Wicke, 2021). Based on the chromosome-level genome data from Ocu, Pcr, Llu and the outgroup species Mgu, we obtained 18 protochromosomes by reconstructing the ancestral karyotype of the Orobanchaceae (Supplemental Figure 6). Analysis of the syntenic chromosomal fragments indicated that Llu and Ocu exhibited good collinearity with the ancestral protochromosomes (Supplemental Figures 6 and 7). For example, Chromosome 9 of Ocu and Chromosome 5 of Llu appeared to be very similar to the ancestral Chromosome 8. However, Pcr has only eight chromosomes and its chromosomes are likely to have experienced at least 11 fusion events.

Orobanchaceous species have variable genome sizes, ranging from 223 Mb to 10.7 Gb, (Bai et al., 2012; Joel et al., 2013; Lyko and Wicke, 2021). Of the six orobanchaceous species we studied (Llu, Pja, Pcr, Sas, Ocu, and Pae), Llu has the smallest genome (223 Mb), while Pae has the largest (3992 Mb), 20 times bigger than the Llu genome. Even though recent WGD can rapidly increase genome size, our analysis ruled out recent WGD events in the evolution of these six species. Thus, to gain insight into the driving force behind the genome expansions in these parasites, we compared the genome sizes of Llu, Pja, Pcr, Sas, Ocu, and Pae and the repetitive elements (REs) in their genomes (Figure 3A). The large genome sizes mainly resulted from increased numbers of REs. In the genomes of Ocu, Pja, and Pcr, which are about 1 Gb, 64-79% of the genomes are

261 made up of REs; Pae has the largest genome (3.877 Gb), and 84% of the Pae genome
262 comprises REs, with the Long Terminal Repeats (LTRs) family being the most abundant.
263 Notably, analysis of the sequences of the LTRs in Llu, Pcr, Ocu, and Pae suggested that
264 the subfamilies of LTRs expanded are species-dependent, especially during the very
265 recent extreme extents in Pae and Ocu (Supplemental Figure 8). To determine whether
266 the massively increased numbers of LTRs could affect the coding regions in these
267 species, we compared the intron sizes in Pae and Llu, which have the largest and smallest
268 genomes, respectively, of the orobanchaceous species included in this study. We found
269 that 73.23% of intron length variations were less than 1-fold, regardless of genome sizes
270 (Figure 3B). Moreover, the RE contents in the gene flanking regions became larger as
271 genome size increased (Figure 3C). Dramatic changes in the flanking sequences of a
272 large number of genes may have caused genome-wide changes in gene expression. We
273 also noticed that even though no WGD events occurred, Pja, Pcr, Pae, and Ocu, which
274 have very large genomes, had many more genes than does Llu, and that this was mainly
275 due to tandem or segmental gene duplication (30.96%-56.33%; Supplemental Table 6),
276 but not LTR-assisted "copy-paste" duplication (6.55%-17.40%; Supplemental Table 6).
277 Because Chromosome 9 of Llu and Chromosome 5 of Ocu are homologous, we
278 specifically inspected the differences between these two chromosomes. Ocu's
279 Chromosome 5 is eight times larger than is Chromosome 9 of Llu (82 Mb vs 10 Mb). We
280 found many Class I retrotransposons, including LTRs, in the proximal region of
281 Chromosome 5 of Ocu, where gene densities are relatively low, supporting our
282 hypothesis that expansion of LTRs is the main driving force behind the increased genome
283 sizes in orobanchaceous parasites (Figure 3D). This finding is consistent with a previous
284 conclusion obtained by genome skimming (Piednoël et al., 2012).
285
286 **Evolution of genes and gene families in the Orobanchaceae**
287 We analyzed all the gene families in the orobanchaceous species (Pae, Ocu, Sas, Pja, Pcr,
288 and Llu), *Cuscuta australis* (Cau), and their relatives (for orobanchaceous species: Mgu
289 and *Sesamum indicum*; for Cau: *Ipomoea nil* and *Solanum lycopersicum*), as well as two
290 outgroup species, *Coffea canephora* and *Arabidopsis thaliana* (Dataset S1). We found
291 that, like Cau, all parasites in the Orobanchaceae experienced frequent gene family

292 contractions; however, many gene families in orobanchaceous parasites exhibited severe
293 expansions, which are rare in Cau (Figure 4). Based on the phylogenetic tree (Figure 1),
294 we calculated the numbers of gene families that underwent significant expansions and
295 contractions across all branches (Figure 4). Contractions were frequent when parasitism
296 and obligate parasitism first appeared in the Orobanchaceae, and thereafter, most lineages
297 appeared to have undergone independent gene family contractions. Most of the gene
298 family expansions occurred independently in different species. Of the expanded gene
299 families, genes related to transcriptional regulation, transport, cell wall formation,
300 phloem/xylem histogenesis, and lateral root formation were enriched (Dataset S2), and
301 these genes may be involved in the evolution of the haustorium.
302
303 Furthermore, we screened the transcriptomic data (Supplemental Table 7) from Pja, Sas,
304 and Pae for genes with high expression levels in the haustoria (where expression level of
305 a given gene was at least 1-fold greater in haustorium tissues than in other tissues, see
306 Supplemental Methods), and found that at least 40.92% of these haustorially highly
307 expressed genes (HHEGs) evolved through duplication after the emergence of the most
308 recent common ancestor (MRCA) of Cca and the Orobanchaceae (Figure 5A,
309 Supplemental Table 8). Next, these duplicated genes were used to place the duplication
310 events in the phylogeny (Figure 5A, Supplemental Table 9): 28.03%-36.34% of the
311 duplication events occurred in the MRCA of Mgu and the Orobanchaceae (S4 in Figure
312 2), which speciated after the β^{L} WGD event (Figure 2). To quantify how many of the
313 duplication genes among the HHEGs resulted from the β^{L} WGD event, we extracted the
314 colinear fragments where the gene duplicates were located and found that 96.91%
315 (439/453) of these duplicated genes were located in colinear fragments caused by the β^{L}
316 event. Thus, at least 11% of the HHEGs originated from the β^{L} event. Most of the other
317 duplication events occurred during the specific evolution of Pja, Sas, and Pae, but not in
318 the common ancestor of the orobanchaceous parasites (Figure 5A, Supplemental Table
319 9). Therefore, the β^{L} event was associated with the origin of many haustorium-related
320 genes.
321

322 The β^L event occurred at 73.48 ± 6.58 Mya, while the origin of parasitism was estimated
323 to be at least 38.58 Mya (95% HPD, 31.27-45.81 Mya), ~ 35 million years after the β^L
324 event (Figure 2). After a WGD event, the duplicated genes are often rapidly lost (Maere
325 et al., 2005). We used a mathematical model in which the duplicated genes are
326 exponentially lost during evolution under neutral selection (Ren et al., 2018) (see
327 Supplemental Methods) to estimate the fates of duplicated genes resulting from the β^L
328 event, and then we compared the predicted values with the retained duplicated genes
329 from the β^L event in Mgu, Llu, Pja, Sas, and Pae. The model predicted that 48.10% of the
330 duplicate genes from the β^L event would be retained when parasitism emerged, and
331 assuming neutral evolution, that about 18.23% of the duplicated genes should be retained
332 in the extant genomes, a proportion similar to those observed in the autotrophic Mgu and
333 Llu (Figure 5B, Supplemental Table 10). In the genome of the facultative hemiparasite
334 Pja, the retention rate of β^L-generated gene duplicates was 17.97%, a value similar to the
335 value predicted under neutral selection (18.23%). However, when we focused on the
336 HHEGs, the retention rate increased to 30.17% (Figure 5B, Supplemental Table 10),
337 suggesting strong positive selection. In terms of the obligate hemiparasite Sas and
338 holoparasite Pae, only 9.42% and 9.06%, respectively, of the duplicates resulting from
339 the β^L event were retained (Figure 5B, Supplemental Table 10); these retention rates are
340 much lower than the model predictions, and this might be related to the large-scale gene
341 loss in obligate parasites (see below). Similarly, the retention rates of duplicated HHEGs
342 resulting from the β^L event decreased to 17.97% and 17.54%, respectively (Figure 5B,
343 Supplemental Table 10). Thus, in both facultative and obligate parasites, the retention
344 rates of gene duplicates from the β^L event are higher in HHEGs than in genome-wide
345 gene duplicates, suggesting that the β^L event was important for the origin of the
346 haustorium and parasitism. Furthermore, more of the gene duplicates resulting from the
347 β^L event were retained in the facultative than in the obligate parasites. It is possible that
348 during the evolution of obligate parasites from their facultative ancestor, some of the
349 gene duplicates were no longer required and were therefore deleted from the genome.
350
351 Gene loss is common in parasitic plants (Sun et al., 2018; Cai et al., 2021). We split all
352 the gene families into OGs, and these OGs were thereafter used for gene loss analysis.

353 Gene family contractions were not additionally analyzed in other ways, as comparison of
354 gene loss provides detailed information on the biological significance of the missing
355 genes. A bioinformatic pipeline was also employed to screen for false positives generated
356 by errors of gene annotation and orthogrouping (see Supplemental Methods). As a
357 comparison, Cau was included in the gene loss analysis with five parasitic
358 Orobanchaceae species, as *Cuscuta* and Orobanchaceae are both Lamiids. *Sapria*
359 *himalayana* (Rafflesiaceae) (Cai et al., 2021) was not included as it is phylogenetically
360 very distant from Lamiids and is an endoparasite. We found that the severity of gene loss
361 was positively associated with the degree of parasitism: the facultative hemiparasites Pja
362 and Pcr lost 2.41% and 2.95%, respectively, of their conserved genes, the obligate
363 hemiparasite Sas lost 5.71%, and the obligate holoparasites Ocu, Pae, and Cau lost
364 13.11%, 13.85%, and 14.88%, respectively (Figure 6A and B). The history of gene loss
365 events in orobanchaceous parasites was then estimated based on the principle of
366 maximum parsimony, and the common ancestor of all orobanchaceous parasites was
367 found to have lost only 74 genes, while 191 and 777 genes were lost, respectively, during
368 the evolution of the ancestors of obligate parasites and holoparasites (Figure 6A).
369 Notably, even though hundreds of genes were lost during the species-specific evolution
370 of these parasites (Figure 6A), many were convergent (Figure 6A). By comparing the lost
371 genes in the Pedicularideae (Pja and Pcr) and the Orobancheae (Ocu and Pae), we found
372 that 74-122 genes were convergently lost, which was well supported by the
373 hypergeometric test (Figure 6A). Furthermore, in the closely related holoparasites Ocu
374 and Pae, 33.6% and 29.9% of the lost genes were species-specific, and selection analysis
375 indicated that significant portions of these Ocu- and Pae-retained genes, which had been
376 lost in the other species, were under relaxed selection (Supplemental Figure 9). It is
377 possible that Ocu and Pae are still experiencing genes loss, and these genes will be
378 convergently lost in the future due to relaxed selection.
379
380 We next compared the lost genes in the orobanchaceous parasites and Cau. Among the
381 1687 lost genes in Cau, more than half (889 genes, 57.14%) were also lost in Pae and
382 Ocu, and 75 genes were also lost in the hemiparasites Sas, Pja, and Pcr (Figure 6B). It is
383 evident that all orobanchaceous parasites and Cau convergently lost various genes,

384 suggesting that some of these gene loss events could contribute to their adaptation to the
385 parasitic lifestyle. To gain further insight into the functions of the lost genes in Cau, Pae,
386 and Ocu, which are all holoparasites, we first performed gene ontology (GO) enrichment
387 analysis on all their lost genes, and the GO terms were then used for interspecies pairwise
388 comparisons. The resulting common and unique GO terms were further categorized using
389 a network analysis (Figure 6C). Pae and Ocu-lost genes exhibited very similar GO terms,
390 while the enriched GO terms from the Cau lost genes appeared to be quite different from
391 those of Pae or Ocu (Supplemental Figure 10). In Cau, most of the GO terms for the lost
392 genes were related to ion transport in roots, ethylene biosynthesis, and response to
393 nutrient and oxygen levels, which is consistent with the rootless morphology of Cau
394 (Figure 6C). The majority of the lost genes in Pae and Ocu were involved in
395 photosynthesis, chloroplast organization, and chlorophyll biosynthesis. Although Cau
396 also lost genes related to these processes, the numbers of GO terms under these processes
397 are markedly fewer than those from Pae or Ocu (Figure 6C).
398
399 In the leaves of autotrophic plants, a complex flowering regulatory network perceives
400 endogenous and environmental cues and determines the flowering time of the plant
401 accordingly (Blümel et al., 2015). Given that parasitic plants need to interact with their
402 hosts, and that holoparasites no longer have leaves, which are the organs that determine
403 flowering (Jaeger et al., 2013), parasitic plants are likely to have evolved unique
404 mechanisms to regulate flowering. Genome sequencing of the dodder Cau and a fully
405 mycoheterotrophic orchid *Gastrodia elata* indicated that many flowering-related genes
406 were lost in both species (Sun et al., 2018; Xu et al., 2021). We therefore inspected the
407 genes responsible for the regulation of flowering in all the orobanchaceous parasites
408 included in this study, with Cau included for comparisons (Figure 6D). We found that
409 many important flowering regulatory genes were indeed missing from the holoparasites
410 Pae, Ocu, and Cau, including *CO*, *SVP*, *ELF4*, *AGL19* and *AGL16*, and *CDF1* and *CDF3*
411 (Figure 6D). In terms of the regulatory pathways of flowering, Cau lost more genes
412 related to the autonomous and hormone pathways than did Pae and Ocu, while Pae and
413 Ocu lost more genes involved in the circadian and photoperiod pathways than did Cau.
414 We suppose that this is due to the fact that Cau uses its host plant's flowering signal FT

415 to activate its own flowering (Shen et al., 2020), and that Pae and Ocu are unlikely to rely
416 on light to signaling flowering, as they spend a significant amount of time living
417 underground.
418
419 **The evolution of KAI2 gene family**
420 After perceiving the strigolactones (SLs) released from the roots of adjacent host plants,
421 the seeds of obligate parasites in the Orobanchaceae germinate (Cook et al., 1966;
422 Yokota et al., 1998; Joel et al., 2013). In normal autotrophic plants, D14, an alpha/beta
423 hydrolase protein, perceives SLs and regulates physiological processes related to
424 development and stress resistance (Waters et al., 2017). In *Striga*, however, the receptor
425 for SLs released by the host has been identified to be encoded by *KAI2d.* Importantly,
426 *KAI2d* did not evolve from *D14*, but rather from the *KAI2* gene, which encodes the
427 receptor of karrikin, through tandem duplication, large gene family expansion, and
428 neofunctionalization (Conn et al., 2015; Yoshida et al., 2019).
429
430 We analyzed the *KAI2* gene families in the Orobanchaceae and in a few autotrophic
431 species from lamiids, with *Arabidopsis* as the outgroup (Figure 7A). Only a single *KAI2*
432 was identified in *Arabidopsis*, whereas in *Coffea canephora*, *Solanum lycopersicum*, and
433 plants of the Orobanchaceae, at least two copies exist, suggesting that *KAI2* duplicated in
434 the ancestor of lamiids resulting in the two paralogs, *KAI2c* and *KAI2i* (Figure 7A). We
435 inspected the chromosomal fragments containing the *KAI2* gene in Llu and *Solanum*
436 *lycopersicum*, and found that in both species, *KAI2c* and *KAI2i* are either on different
437 chromosomes (Llu) or distantly located on the same chromosome (~ 20 Mb in *Solanum*
438 *lycopersicum*). Collinearity analysis indicated that the two chromosomal fragments
439 containing the *KAI2* gene are obviously homologous with the fragment containing *KAI2*
440 gene in *Arabidopsis* (Figure 7B). Given that no WGD events were detected in the
441 ancestor of the lamiids ((Denoeud et al., 2014) and this study, Figure 2), the duplication
442 of *KAI2* in the lamiids is a result of segmental duplication. *KAI2d* was then derived from
443 duplication of *KAI2i* by tandem duplication (Figure 7A), as in Pja, Pcr, and Sas, *KAI2d*
444 and *KAI2i* are adjacent (Figure 7B). In Pja and Pcr, the chromosome fragments
445 containing *KAI2i* and *KAI2d* show good collinearity with the chromosome fragment

446 containing *KAI2i* in Llu, while in Sas, Ocu, and Pae, the collinearity was not detected
447 (Figure 7B). Thus, we inferred that this tandem duplication event, in which *KAI2d* came
448 into existence, occurred in the common ancestor of the orobanchaceous parasitic plants.
449 In Pja and Pcr the locations of these duplicated genes remained the same, whereas in Sas,
450 Pae, and Ocu both genes translocated to other loci (Figure 7B).
451
452 **Discussion**
453
454 Orobanchaceae is the only family of plants containing species displaying all types of
455 parasitic lifestyles, and it is preferred for studying the origin and evolution of parasitism.
456 The phylogeny of the Orobanchaceae has been studied using plastome fragments (Young
457 et al., 1999; Wolfe et al., 2005), ITS (Wolfe et al., 2005; McNeal et al., 2013), and
458 individual nuclear genes (McNeal et al., 2013; Li et al., 2019). Even though most of these
459 studies identified eight monophyletic clades, the phylogenetic relationships among these
460 clades was not completely resolved. In this study, we assembled three genomes of the
461 orobanchaceous plants *L. luchunensis*, *O. cumana*, and *Phelipanche aegyptiaca* (*P.*
462 *aegyptiaca*). Together with the published/released transcriptome and genome data from
463 other orobanchaceous plants, this study covered seven out of eight major clades of the
464 Orobanchaceae. Using 907 nuclear OGs, we reconstructed a highly resolved and well-
465 supported phylogeny of the Orobanchaceae, and the divergence times were estimated
466 using a Bayesian molecular dating approach (Figure 1). Based on the phylogeny, we
467 propose that 1) the common ancestor of the Orobanchaceae, an autotrophic plant, evolved
468 parasitism (a haustorium) at least 38.58 Mya (95% HPD, 31.27-45.81 Mya); 2) the
469 Cymbarieae, which comprises only facultative hemiparasites, was the earliest diverged
470 parasitic lineage with low biodiversity; 3) by the Oligocene (33.90-23.03 Mya), a
471 facultative hemiparasitic ancestor rapidly evolved into two lineages; one diverged into
472 the Buchnereae (facultative hemiparasitic, obligate hemiparasitic, and holoparasitic
473 species) and the Orobancheae (holoparasites), and the other evolved to be the
474 Pedicularideae (facultative hemiparasitic species) and the Rhinantheae (facultative
475 hemiparasitic and holoparasitic); these four clades are the most species-rich. Grasslands
476 rapidly expanded during the Oligocene (Torsvik and Cocks, 2016), and this event may

477 have provided orobanchaceous parasites with opportunities to exploit grasses and herbs,
478 enabling the expansion of these parasites (Eriksson and Kainulainen, 2011; McNeal et al.,
479 2013). Notably, contrasting with a recent study (Li et al., 2019), our scenario suggested
480 that the hemiparasites evolved earlier than the holoparasites, which is consistent with the
481 progressive evolution of increasing dependency of hosts.
482
483 Most species in the Orobanchaceae have 8 to 20 pairs of chromosomes, but the
484 chromosome number can reach 132 in some species (Barker et al., 1988; Joel et al., 2013;
485 Lyko and Wicke, 2021). Based on the variation in chromosome number, Lyko and Wicke
486 (2021) proposed at least three ancient WGD events during the evolution of
487 Orobanchaceae, including one WGD event in *Striga* and another in the Orobancheae
488 clade. The WGD event in *Striga* was confirmed by a recent genome analysis (Yoshida et
489 al., 2019). However, our collinearity analysis of the genome data from *L. luchunensis*,
490 *Phtheirospermum japonicum* (*P. japonicum*), *Pedicularis cranolopha*, *Striga asiatica* (*S.*
491 *asiatica*), *O. cumana*, and *P. aegyptiaca* did not reveal any other WGD events in the
492 Orobanchaceae, except a single WGD event (α^B) that occurred during the evolution of the
493 ancestor of Buchnereae (32.64 ± 6.58 Mya) (Figure 2), the clade containing *S. asiatica*.
494 Moreover, the 1:1 collinearity of the chromosomes of *L. luchunensis* and *O. cumana*
495 (Supplemental Figure 5) also strongly suggests that no ploidization occurred after the
496 divergence of *L. luchunensis* and *O. cumana*. When we inspected the emergence times of
497 facultative parasitism, obligate parasitism, and later holoparasitism in the Orobanchaceae,
498 no WGD events were found to have co-occurred with these evolutionary events (Figures
499 1 and 2), suggesting that genome duplications in the Orobanchaceae were not associated
500 with the evolution of parasitism. However, by analyzing the evolutionary history of
501 HHEGs in *P. japonicum*, *S. asiatica*, and *P. aegyptiaca*, we found that the β^L WGD
502 event, which occurred long before the emergence of parasitism, was involved in the
503 origin of at least 11% of the HHEGs. In all parasites, the retention rates of β^L-duplicated
504 genes are much higher among the HHEG than genome-wide (Figure 5B, Supplemental
505 Table 10). Particularly, in the facultative hemiparasite *P. japonicum*, the retention rate of
506 the β^L duplicated genes among the HHEGs was as high as 30.17%, far exceeding the
507 expected retention rate under neutral selection (18.23%), indicating that these gene

508 duplicates were retained by positive selection (Figure 5B). The gene duplicates caused by
509 the β^L event may have allowed the ancestor of parasitic plants to acquire parasitism and a
510 haustorium through neofunctionalization and/or subfunctionalization, and many of these
511 genes remain being important in the lateral haustorium of facultative hemiparasitic plants.
512 When the terminal haustorium/obligate parasitism evolved independently in the ancestors
513 of *S. asiatica* and *P. aegyptiaca*, the HHEGs originated from the β^L seem to have become
514 less important, indicated by the highly reduced retention rates of the β^L-duplicated genes
515 among the HHEGs. Notably, the ancestors of *S. asiatica* and *P. aegyptiaca* both
516 experienced lineage-specific gene duplications in the HHEGs (Figure 5A), and these
517 lineage-specific gene duplications are likely to be the driving force behind the evolution
518 of the terminal haustorium and obligate parasitism.
519
520 The dodder *C. australis* and the endoparasite *Sapria himalayana* (*S. himalayana*) are both
521 non-orobanchaceous species. *C. australis* has only 19671 genes (Sun et al., 2018), and it
522 is among the higher plants with the fewest protein-coding genes. The *S. himalayana*
523 genome comprises 55179 genes; however, most of them are transposon-like sequences,
524 leaving only about 10000 low-copy genes (Cai et al., 2021). In contrast, the
525 orobanchaceous parasites have many more genes than do *C. australis* or *S. himalayana*
526 (Figure 4), and this is due to species-specific gene family expansions in different
527 orobanchaceous parasites (Figure 4). Furthermore, of these expanded gene families,
528 genes likely to be related to haustorium development and function were identified and
529 many of the genes highly expressed in haustoria evolved from lineage-specific
530 duplications (Figure 5). We propose that species-specifically expanded gene families may
531 be the driving force underlying the adaptation of the parasites to different hosts.
532
533 Similar to *C. australis* and *S. himalayana*, large-scale gene loss was also detected in
534 orobanchaceous parasites, especially in the holoparasites *O. cumana* and *P. aegyptiaca*.
535 The reconstructed gene-loss history indicated that only 74 genes were missing in the
536 common ancestor of the orobanchaceous parasites, when parasitism evolved; by contrast,
537 777 genes were deleted from the genome of the MRCA of *O. cumana* and *P. aegyptiaca*.
538 Thus, gene loss became more severe as the degree of parasitism increased. Our gene loss

539 data from the Orobanchaceae and the previous published data from *S. himalayana* (Cai et
540 al., 2021) and the full mycoheterotrophic orchidaceous *Gastrodia elata* (Xu et al., 2021),
541 which are all heterotrophs, all strongly support the idea that the degree of gene loss and
542 parasitism are highly positively correlated. Accordingly, heterotrophic plants could be
543 divided into four classes with increasing levels of heterotrophy and degrees of gene loss.
544 In the first class are heterotrophs with 2-3% gene loss; these plants are facultative
545 hemiparasites. In the second class are heterotrophs with 6-7% gene loss requiring hosts to
546 complete certain developmental stages, and these are obligate hemiparasites and initial
547 mycoheterotrophic plants, such as most species of Orchidaceae. In the third class are
548 heterotrophs that have lost 13-15% genes, and these are holoparasites or full
549 mycoheterotrophs with vegetative organs, such as *Cuscuta*, *Orobanche*, and *Gastrodia*.
550 The fourth class comprises plants where gene loss is over 30%; these are endoparasites
551 with only reproductive organs, such as *S. himalayana*. The lost genes in the heterotrophs
552 of the first and second classes seem to be rather random, while the genes lost in the
553 heterotrophs of the third class are related to photosynthesis, transcriptional regulation,
554 stress adaptation, and organogenesis; the lost genes in the endoparasite *Sapria*
555 *himalayana*, from the fourth class, includes almost all those lost in *P. aegyptiaca*, *O.*
556 *cumana*, *C. australis*, and *Gastrodia elata*, representing an extreme form of parasitism.
557
558 Convergent gene loss in different parasitic plants may be the result of neutral evolution
559 after the loss of purifying selection, as these genes are no longer required for parasitism.
560 However, the loss of some genes may enhance the fitness of the parasitic plant. In *Striga*,
561 the *protein phosphatase 2C* gene is mutated, leading to abscisic acid (ABA) insensitivity,
562 with the impaired ABA signaling pathway leading to very high transpiration in *Striga*,
563 and enabling this parasite to absorb nutrients from host xylem sap, even under drought
564 conditions (Fujioka et al., 2019; Yoshida et al., 2019). The leafless *Cuscuta australis* has
565 lost almost all the flowering regulatory pathways, and its *FT* gene, which encodes an
566 important flowering signal, had pseudogenized; however, *C. australis* uses the FT
567 proteins from host plants to activate its own flowering, enabling it to synchronize its
568 flowering times with those of its hosts (Shen et al., 2020). In this manner, *C. australis* can
569 parasitize a large number of host species with varying flowering times, and losing the

570 flowering genes is thus an elegant adaptive strategy in *C. australis* (Shen et al., 2020). It
571 would be interesting to study which pathways are affected by gene loss in
572 orobanchaceous parasites and, importantly, the physiological and ecological impact
573 incurred by gene loss.
574
575 Searcy and Maclnnis (1970) proposed a three-phase model of parasitic plant genome
576 evolution: first, development of haustorium which enables a physical connection with
577 host plants; secondly, after the haustorium has evolved, the parasites start to lose genes
578 that are no longer needed, and thirdly, the parasites evolve to adapt to their specific hosts.
579 The evolution of orobanchaceous parasites supports this model well. The β^L WGD event
580 in the common ancestor of the Orobanchaceae and *Mimulus* generated many duplicated
581 genes which likely contributed to the evolutionary gain of a haustorium; thereafter, gene
582 loss events occurred in orobanchaceous parasites, especially in the holoparasitic species,
583 and in parallel, expansions in specific gene families may have contributed to
584 specialization and adaptation of the parasites to the hosts.
585
586 In this study, our phylogenomic analysis constructed a robust phylogenetic framework of
587 the major lineages of the Orobanchaceae. Furthermore, genome evolution and
588 comparative genomics analyses shed new light on the origin and evolution of parasitism
589 in the Orobanchaceae. These genome resources and further molecular, genetic, and
590 genomic studies on orobanchaceous species will provide many interesting insights into
591 the evolution underlying the drastic physiological and morphological changes in these
592 fascinating plants.
593
594 **Materials and Methods**
595
596 **Estimation of genome size**
597 Genome sizes were estimated based on flow cytometry and genomic *k*-mer analysis
598 (Supplemental Table 1). A CyFlow Space-3000 (Partec, Germany) was used, and *Oryza*
599 *sativa* L. ssp. *japonica* (370 Mb) and *Zea mays* ssp. *mays* var. B73 (2.1 Gb) were selected
600 as the internal standards (Supplemental Table 1). In the genomic *k*-mer analysis, the

601 genome sizes were estimated using ~100 × coverage of short-read sequencing data
602 (Supplemental Table 1), based on the *k*-mer (*k*=17) depth frequency distribution analysis
603 embedded in the GCE software (v1.0.0) (Liu et al., 2020).
604
605 **Genome sequencing**
606 Leaves of Llu and aboveground parts of stems of Ocu and Pae were harvested and
607 cleaned with tap water before being used for DNA extraction. For short-read sequencing,
608 DNA libraries were constructed using a VAHTS Universal Plus DNA Library Prep Kit
609 for MGI (Vazyme, Nanjing, China) following the manufacturer's instructions, and the
610 libraries were sequenced on an MGI-SEQ 2000 (MGI Tech, Shenzhen, China) to
611 generate 150-bp paired-end data for each species with approximately 100 × coverage. For
612 long-read sequencing, we used the PromethION platform from Oxford Nanopore
613 Technologies for Llu, and for Ocu and Pae, we used the PacBio Sequel II platform with
614 the CLR mode. For each species, long-read data with approximately 100 × coverage were
615 generated.
616
617 Tissues of Llu and Ocu were used for the Hi-C pipeline following a previously described
618 method (Rao et al., 2014), including crosslinking, chromatin digestion with MboI (New
619 England BioLabs, MA), labeling of DNA ends, ligation, purification, shearing, and biotin
620 pull-down. The Hi-C libraries were sequenced on an MGI-SEQ 2000 (MGI Tech,
621 Shenzhen, China) platform in the 150-bp paired-end mode, resulting in approximately
622 100 × coverage.
623
624 **Genome assembly and annotation**
625 For genome assembly, we used the long-read data for *de novo* assembly, followed by
626 error correction and polishing using the short-read data to obtain high-quality contigs. For
627 Llu and Ocu, we used the Hi-C data to anchor the contigs onto chromosomes, to obtain
628 chromosome-level assemblies (See Supplemental Methods for assembly software and
629 procedures). To aid genome annotation, we obtained RNA-seq data from different tissues
630 of orobanchaceous plants (Supplemental Table 3) using a DNBSEQ-T7 platform (MGI
631 Tech, Shenzhen, China). Based on these data, we adopted an annotation pipeline

632 integrating RNA-seq data, *ab initio* predicted features, and homologous sequence
633 information to annotate repetitive sequences and protein-coding genes in the genomes
634 (Supplemental Methods for details).
635
636 See Supplemental Methods for phylogenetic analysis and comparative genomics analysis.
637
638 **Data availability**
639
640 The genome assembly, gene models, and sequence reads of Llu, Ocu, and Pae are
641 available at the National Genomics Data Center (https://ngdc.cncb.ac.cn/) under
642 BioProject number PRJCA008847. All data are also available from the corresponding
643 author upon request.
644
645 **Acknowledgments**
646
647 We thank Dr. Xiaopeng Yun (Inner Mongolia Academy of Agricultural and Animal
648 Husbandry Sciences) for providing Ocu and Pae seeds, the Germplasm Bank of Wild
649 Species in Southwest China for providing Llu seeds, and Dr. Steven Runo (Kenyatta
650 University) for providing the photograph of *Striga*. Drs. Jun Zhao and Jian Zhang (Inner
651 Mongolia Agricultural University) are thanked for the identification of Ocu. We are
652 grateful to the High-Performance Computing Facility of the Department of Economic
653 Plants and Biotechnology at the Kunming Institute of Botany for providing computing
654 services. This study was supported by the National Natural Science Foundation of China
655 (No. 32000179, Y.X.), the Strategic Priority Research Program of Chinese Academy of
656 Sciences (No. XDPB16, J.W.), the CAS "Light of West China" Program (G.S.), the
657 Special Research Assistant of Chinese Academy of Sciences (Y.X. and J.Z.), and the
658 Postdoctoral Directional Training Foundation of Yunnan Province (Y.X.).
659
660 **References**
661

662 **Bai, C., Alverson, W. S., Follansbee, A., and Waller, D. M.** (2012). New reports of
663 nuclear DNA content for 407 vascular plant taxa from the United States. *Annals of*
664 *Botany* **110**:1623–1629.
665 **Barker, W. R., Kiehn, M., and Vitek, E.** (1988). Chromosome numbers in Australian
666 *Euphrasia* (Scrophulariaceae). *Plant Systematics and Evolution* **158**:161–164.
667 **Becher, H., Brown, M. R., Powell, G., Metherell, C., Riddiford, N. J., and Twyford,**
668 **A. D.** (2020). Maintenance of Species Differences in Closely Related Tetraploid
669 Parasitic *Euphrasia* (Orobanchaceae) on an Isolated Island. *Plant Communications*
670 **1**:100105.
671 **Blümel, M., Dally, N., and Jung, C.** (2015). Flowering time regulation in crops — what
672 did we learn from Arabidopsis? *Current Opinion in Biotechnology* **32**:121–129.
673 **Bromham, L., Cowman, P. F., and Lanfear, R.** (2013). Parasitic plants have increased
674 rates of molecular evolution across all three genomes. *BMC Evolutionary Biology*
675 **13**:126.
676 **Cai, L., Arnold, B. J., Xi, Z., Khost, D. E., Patel, N., Hartmann, C. B., Manickam, S.,**
677 **Sasirat, S., Nikolov, L. A., Mathews, S., et al.** (2021). Deeply altered genome
678 architecture in the endoparasitic flowering plant *Sapria himalayana* Griff.
679 (Rafflesiaceae). *Current Biology* **31**:1002-1011.e9.
680 **Conn, C. E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B.,**
681 **Westwood, J. H., Shirasu, K., Bond, C. S., Dyer, K. A., and Nelson, D. C.**
682 (2015). Convergent evolution of strigolactone perception enabled host detection in
683 parasitic plants. *Science* **349**:540–543.
684 **Cook, C. E., Whichard, L. P., Turner, B., Wall, M. E., and Egley, G. H.** (1966).
685 Germination of witchweed (*Striga lutea* Lour.): Isolation and properties of a potent
686 stimulant. *Science* **154**:1189–1190.
687 **Cook, C. E., Whichard, L. P., Wall, M., Egley, G. H., Coggon, P., Luhan, P. A., and**
688 **McPhail, A. T.** (1972). Germination stimulants. II. Structure of strigol, a potent
689 seed germination stimulant for witchweed (*Striga lutea*). *J. Am. Chem. Soc.*
690 **94**:6198–6199.
691 **Cui, S., Kubota, T., Nishiyama, T., Ishida, J. K., Shigenobu, S., Shibata, T. F.,**
692 **Toyoda, A., Hasebe, M., Shirasu, K., and Yoshida, S.** (2020). Ethylene signaling

693 mediates host invasion by parasitic plants. *Science Advances* Advance Access
694 published October 2020, doi:10.1126/sciadv.abc2385.
695 **Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O.,**
696 **Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., et al.** (2021). Twelve
697 years of SAMtools and BCFtools. *GigaScience* **10**:giab008.
698 **Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M.,**
699 **Zheng, C., Alberti, A., Anthony, F., Aprea, G., et al.** (2014). The coffee genome
700 provides insight into the convergent evolution of caffeine biosynthesis. *Science*
701 **345**:1181–1184.
702 **Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C.,**
703 **Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., et al.** (2017). De novo
704 assembly of the Aedes aegypti genome using Hi-C yields chromosome-length
705 scaffolds. *Science* **356**:92–95.
706 **Ellinghaus, D., Kurtz, S., and Willhoeft, U.** (2008). LTRharvest, an efficient and
707 flexible software for de novo detection of LTR retrotransposons. *BMC*
708 *Bioinformatics* **9**:18.
709 **Eriksson, O., and Kainulainen, K.** (2011). The evolutionary ecology of dust seeds.
710 *Perspectives in Plant Ecology, Evolution and Systematics* **13**:73–87.
711 **Fawcett, J. A., Maere, S., and Peer, Y. V. de** (2009). Plants with double genomes might
712 have had a better chance to survive the Cretaceous–Tertiary extinction event. *PNAS*
713 **106**:5737–5742.
714 **Fu, W., Liu, X., Zhang, N., Song, Z., Zhang, W., Yang, J., and Wang, Y.** (2017).
715 Testing the hypothesis of multiple origins of holoparasitism in Orobanchaceae:
716 Phylogenetic evidence from the last two unplaced holoparasitic genera, Gleadovia
717 and Phacellanthus. *Frontiers in Plant Science* **8**:1380.
718 **Fujioka, H., Samejima, H., Suzuki, H., Mizutani, M., Okamoto, M., and Sugimoto,**
719 **Y.** (2019). Aberrant protein phosphatase 2C leads to abscisic acid insensitivity and
720 high transpiration in parasitic *Striga*. *Nat. Plants* **5**:258–262.
721 **Hu, J., Fan, J., Sun, Z., and Liu, S.** (2020). NextPolish: a fast and efficient genome
722 polishing tool for long-read assembly. *Bioinformatics* **36**:2253–2255.

723 **Jaeger, K. E., Pullen, N., Lamzin, S., Morris, R. J., and Wigge, P. A.** (2013).
724 Interlocking Feedback Loops Govern the Dynamic Behavior of the Floral
725 Transition in *Arabidopsis*. *The Plant Cell* **25**:820–833.
726 **Jin, J., and Eaton, D. A. R.** (2022).Pedicularis cranolopha genome reference project
727 (https://www.ncbi.nlm.nih.gov/bioproject/805569).
728 **Joel, D. M., Gressel, J., and Musselman, L. J. eds.** (2013). *Parasitic Orobanchaceae*.
729 Berlin, Heidelberg: Springer Berlin Heidelberg.
730 **Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L.** (2019). Graph-based
731 genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat*
732 *Biotechnol* **37**:907–915.
733 **Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A.**
734 **M.** (2017). Canu: scalable and accurate long-read assembly via adaptive *k* -mer
735 weighting and repeat separation. *Genome Res.* **27**:722–736.
736 **Lemaire, B., Huysmans, S., Smets, E., and Merckx, V.** (2011). Rate accelerations in
737 nuclear 18S rDNA of mycoheterotrophic and parasitic angiosperms. *J Plant Res*
738 **124**:561–576.
739 **Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,**
740 **Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup**
741 (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
742 **25**:2078–2079.
743 **Li, X., Feng, T., Randle, C., and Schneeweiss, G. M.** (2019). Phylogenetic
744 relationships in Orobanchaceae inferred from low-copy cuclear genes:
745 Consolidation of major clades and identification of a novel position of the non-
746 photosynthetic orobanche clade sister to all other parasitic Orobanchaceae. *Front.*
747 *Plant Sci.* **10**:902.
748 **Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., and**
749 **Fan, W.** (2020). Estimation of genomic characteristics by analyzing k-mer
750 frequency in de novo genome projects. *arXiv* Advance Access published February
751 26, 2020.

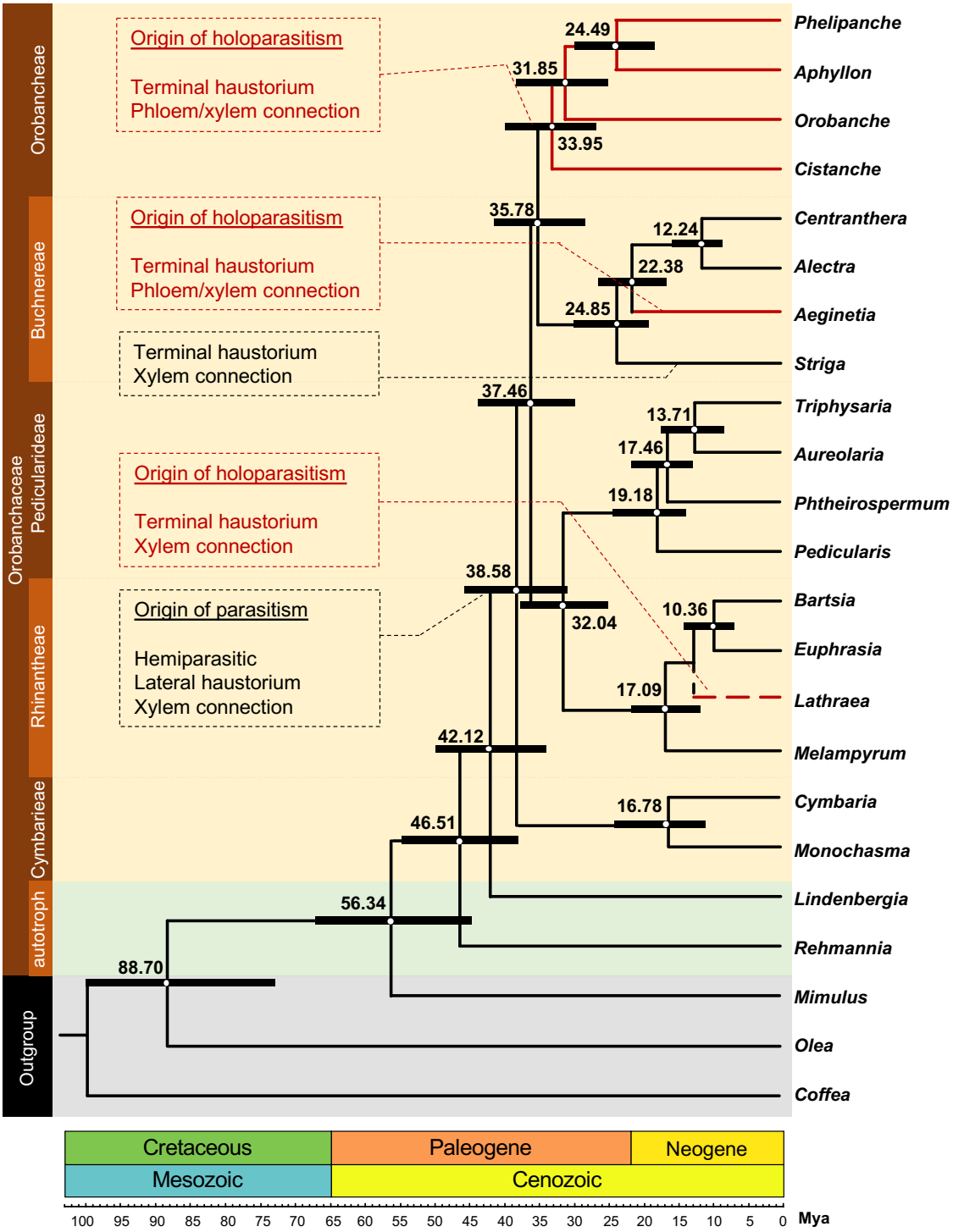
752 **Lyko, P., and Wicke, S.** (2021). Genomic reconfiguration in parasitic plants involves
753 considerable gene losses alongside global genome size inflation and gene births.
754 *Plant Physiology* **186**:1412–1423.
755 **Ma, L., Dong, C., Song, C., Wang, X., Zheng, X., Niu, Y., Chen, S., and Feng, W.**
756 (2021). De novo genome assembly of the potent medicinal plant *Rehmannia*
757 *glutinosa* using nanopore technology. *Computational and Structural Biotechnology*
758 *Journal* **19**:3954–3963.
759 **Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and**
760 **Van de Peer, Y.** (2005). Modeling gene and genome duplications in eukaryotes.
761 *Proceedings of the National Academy of Sciences* **102**:5454–5459.
762 **McNeal, J. R., Bennett, J. R., Wolfe, A. D., and Mathews, S.** (2013). Phylogeny and
763 origins of holoparasitism in Orobanchaceae. *American Journal of Botany* **100**:971–
764 983.
765 **Nickrent, D. L.** (2020). Parasitic angiosperms: How often and how many? *TAXON* **69**:5–
766 27.
767 **Ou, S., and Jiang, N.** (2018). LTR_retriever: a highly accurate and sensitive program for
768 identification of Long Terminal Repeat retrotransposons. *Plant Physiol* **176**:1410–
769 1422.
770 **Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B.,**
771 **Elliott, T. A., Ware, D., Peterson, T., et al.** (2019). Benchmarking transposable
772 element annotation methods for creation of a streamlined, comprehensive pipeline.
773 *Genome Biol* **20**:275.
774 **Piednoël, M., Aberer, A. J., Schneeweiss, G. M., Macas, J., Novak, P., Gundlach, H.,**
775 **Temsch, E. M., and Renner, S. S.** (2012). Next-generation sequencing reveals the
776 impact of repetitive DNA across phylogenetically closely related genomes of
777 Orobanchaceae. *Molecular Biology and Evolution* **29**:3601–3611.
778 **Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D.,**
779 **Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., et al.**
780 (2014). A 3D map of the human genome at kilobase resolution reveals principles of
781 chromatin looping. *Cell* **159**:1665–1680.

782 **Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H., and Qi, J.** (2018).
783 Widespread Whole Genome Duplications Contribute to Genome Complexity and
784 Species Diversity in Angiosperms. *Molecular Plant* **11**:414–428.
785 **Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., Kaneko,**
786 **T., Nakamura, Y., Shibata, D., Aoki, K., et al.** (2012). The tomato genome
787 sequence provides insights into fleshy fruit evolution. *Nature* **485**:635–641.
788 **Schneeweiss, G. M.** (2013). Phylogenetic relationships and evolutionary trends in
789 Orobanchaceae. In *Parasitic Orobanchaceae* (ed. Joel, D. M.), Gressel, J.), and
790 Musselman, L. J.), pp. 243–265. Berlin, Heidelberg: Springer Berlin Heidelberg.
791 **Schulte, P., Alegret, L., Arenillas, I., Arz, J. A., Barton, P. J., Bown, P. R., Bralower,**
792 **T. J., Christeson, G. L., Claeys, P., Cockell, C. S., et al.** (2010). The Chicxulub
793 asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science*
794 **327**:1214–1218.
795 **Searcy, D. G., and MacInnis, A. J.** (1970). Measurements by DNA renaturation of the
796 genetic basis of parasitic reduction. *Evolution* **24**:796–806.
797 **Shen, G., Liu, N., Zhang, J., Xu, Y., Baldwin, I. T., and Wu, J.** (2020). *Cuscuta*
798 *australis* (dodder) parasite eavesdrops on the host plants' FT signals to flower. *Proc*
799 *Natl Acad Sci USA* **117**:23125–23130.
800 **Shi, J., and Liang, C.** (2019). Generic Repeat Finder: a high-sensitivity tool for genome-
801 wide de novo repeat detection. *Plant Physiol* **180**:1803–1815.
802 **Simillion, C., Vandepoele, K., Montagu, M. C. E. V., Zabeau, M., and Peer, Y. V. de**
803 (2002). The hidden duplication past of *Arabidopsis thaliana*. *PNAS* **99**:13627–
804 13632.
805 **Su, W., Gu, X., and Peterson, T.** (2019). TIR-Learner, a new ensemble method for TIR
806 transposable element annotation, provides evidence for abundant new transposable
807 elements in the maize genome. *Mol Plant* **12**:447–460.
808 **Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., Shen, G., Qi, J., Qin,**
809 **Y., Li, J., et al.** (2018). Large-scale gene losses underlie the genome evolution of
810 parasitic plant *Cuscuta australis*. *Nat Commun* **9**:2683.

811 **Toh, S., Holbrook-Smith, D., Stogios, P. J., Onopriyenko, O., Lumba, S., Tsuchiya,**
812 **Y., Savchenko, A., and McCourt, P.** (2015). Structure-function analysis identifies
813 highly sensitive strigolactone receptors in *Striga*. *Science* **350**:203–207.
814 **Torsvik, T. H., and Cocks, L. R. M.** (2016). Paleogene. In *Earth History and*
815 *Palaeogeography*, pp. 240–255. Cambridge: Cambridge University Press.
816 **Tsuchiya, Y., Yoshimura, M., Sato, Y., Kuwata, K., Toh, S., Holbrook-Smith, D.,**
817 **Zhang, H., McCourt, P., Itami, K., Kinoshita, T., et al.** (2015). PARASITIC
818 PLANTS. Probing strigolactone receptors in *Striga hermonthica* with fluorescence.
819 *Science* **349**:864–868.
820 **Twyford, A. D.** (2018). Parasitic plants. *Current Biology* **28**:R857–R859.
821 **Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S.,**
822 **Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., et al.** (2014). Pilon: an
823 integrated tool for comprehensive microbial variant detection and genome assembly
824 improvement. *PLoS One* **9**:e112963.
825 **Waters, M. T., Gutjahr, C., Bennett, T., and Nelson, D. C.** (2017). Strigolactone
826 signaling and evolution. *Annu. Rev. Plant Biol.* **68**:291–322.
827 **Westwood, J. H., Yoder, J. I., Timko, M. P., and dePamphilis, C. W.** (2010). The
828 evolution of parasitism in plants. *Trends in Plant Science* **15**:227–235.
829 **Wicke, S., Müller, K. F., dePamphilis, C. W., Quandt, D., Bellot, S., and**
830 **Schneeweiss, G. M.** (2016). Mechanistic model of evolutionary rate variation en
831 route to a nonphotosynthetic lifestyle in plants. *PNAS* **113**:9045–9050.
832 **Wickett, N. J., Honaas, L. A., Wafula, E. K., Das, M., Huang, K., Wu, B., Landherr,**
833 **L., Timko, M. P., Yoder, J., Westwood, J. H., et al.** (2011). Transcriptomes of
834 the parasitic plant family Orobanchaceae reveal surprising conservation of
835 chlorophyll synthesis. *Curr Biol* **21**:2098–2104.
836 **Wolfe, A. D., Randle, C. P., Liu, L., and Steiner, K. E.** (2005). Phylogeny and
837 biogeography of Orobanchaceae. *Folia Geobot* **40**:115–134.
838 **Wu, S., Han, B., and Jiao, Y.** (2020). Genetic contribution of paleopolyploidy to
839 adaptive evolution in angiosperms. *Molecular Plant* **13**:59–71.
840 **Xu, Z., and Wang, H.** (2007). LTR_FINDER: an efficient tool for the prediction of full-
841 length LTR retrotransposons. *Nucleic Acids Res* **35**:W265-268.

842 **Xu, Y., Lei, Y., Su, Z., Zhao, M., Zhang, J., Shen, G., Wang, L., Li, J., Qi, J., and**
843 **Wu, J.** (2021). A chromosome-scale *Gastrodia elata* genome and large-scale
844 comparative genomic analysis indicate convergent evolution by gene loss in
845 mycoheterotrophic and parasitic plants. *The Plant Journal* **108**:1609–1623.
846 **Yang, Z., Wafula, E. K., Honaas, L. A., Zhang, H., Das, M., Fernandez-Aparicio,**
847 **M., Huang, K., Bandaranayake, P. C. G., Wu, B., Der, J. P., et al.** (2015).
848 Comparative Transcriptome Analyses Reveal Core Parasitism Genes and Suggest
849 Gene Duplication and Repurposing as Sources of Structural Novelty. *Molecular*
850 *Biology and Evolution* **32**:767–790.
851 **Yokota, T., Sakai, H., Okuno, K., Yoneyama, K., and Takeuchi, Y.** (1998). Alectrol
852 and orobanchol, germination stimulants for *Orobanche minor*, from its host red
853 clover. *Phytochemistry* **49**:1967–1973.
854 **Yoshida, S., Cui, S., Ichihashi, Y., and Shirasu, K.** (2016). The haustorium, a
855 specialized invasive organ in parasitic plants. *Annu. Rev. Plant Biol.* **67**:643–667.
856 **Yoshida, S., Kim, S., Wafula, E. K., Tanskanen, J., Kim, Y.-M., Honaas, L., Yang,**
857 **Z., Spallek, T., Conn, C. E., Ichihashi, Y., et al.** (2019). Genome sequence of
858 *Striga asiatica* provides insight into the evolution of plant parasitism. *Curr Biol*
859 **29**:3041-3052.e4.
860 **Young, N. D., Steiner, K. E., and dePamphilis, C. W.** (1999). The evolution of
861 parasitism in Scrophulariaceae/Orobanchaceae: Plastid gene sequences refute an
862 evolutionary transition series. *Annals of the Missouri Botanical Garden* **86**:876.
863
864
865

866 **Figures**



867

868

869 **Figure 1.** Phylogenetic and molecular clock dating analysis of the major clades in the

870 Orobanchaceae.

The phylogenetic tree was constructed using low-copy OGs from 19 orobanchaceous species and three outgroup plants. A gray background indicates autotrophic outgroups; a light grayish-green background represents orobanchaceous autotrophic plant clades; a pale-yellow background represents orobanchaceous parasitic plants; red branches indicate holoparasites. The red dashed line for *Lathraea* highlights that the position of this plant is inferred from the work of Li et al. (2019). The numbers near the internodes represent divergence times (Mya) based on molecular clock dating, and black bars at the internodes represent 95% HPD for divergence times. All internodes are 100% supported. Important evolutionary events are shown in the dashed boxes. The vertical bar on the left indicates five monophyletic clades (Cymbarieae, Orobancheae, Rhinantheae, Buchnereae, and Pedicularideae) and two autotrophic clades (Rehmannieae and Lindenbergieae, which are shown as "autotroph"). The bar on the bottom indicates the geologic time scale.

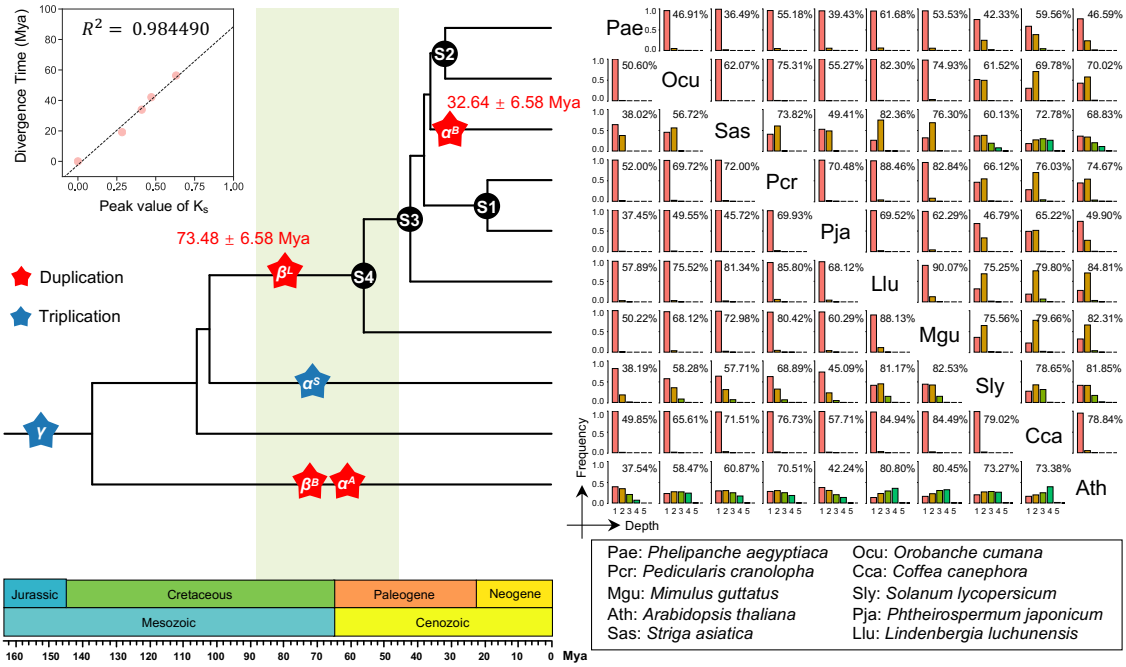


Figure 2. Genome evolution in the Orobanchaceae.

The palaeopolyploidization in the Orobanchaceae and outgroup lineages. The right histogram matrix indicates the coverage depths of orthologous colinear fragments (OCFs) from pairwise species comparisons. The OCFs from a given species pair were mapped respectively to each species to create two histograms indicating the coverage depths of OCFs (X-axis: coverage depths of OCFs, Y-axis: relative frequency). The max depths of OCFs mapped to species B can be used to infer the WGD events in species A after speciation of species A and B, and vice versa (see Supplemental Methods). In the matrix, each row and column of the histogram indicate the two species used to create the OCFs, and the species of the column is the one that was the mapping-target species. The percentage in the upper right corner of each histogram gives the proportion of genes covered by OCF in the target species to the total genes. The WGD events inferred from the histogram matrix are marked with stars in the species tree on the left (blue: triplication, red: duplication), and the positions of the stars represent the times of the WGD events, which were inferred from the peak value of K_s. The black circles on the species tree represent the four speciation events which were used to calculate the

903 correlation between K_s values and divergence times (see Supplemental Methods). The
904 correlation between the peak value of K_s and divergence times is shown in the upper left
905 corner.
906
907

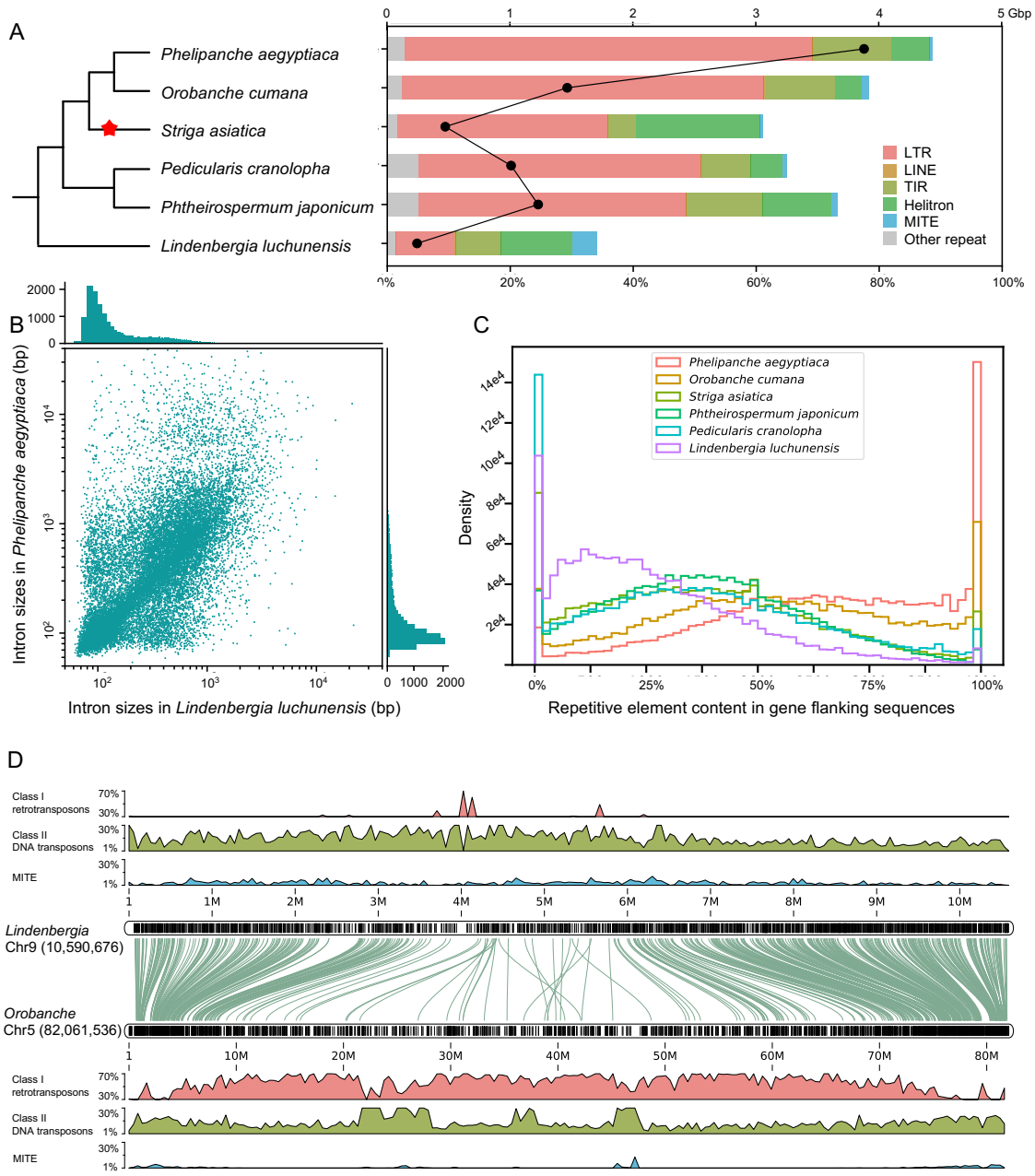


Figure 3. Expansion of repetitive elements in the genomes of species in the Orobanchaceae.

(A) Genome sizes and proportions of repetitive elements in six species of Orobanchaceae. The black dots and upper X-axis represent the genome sizes, and the bars and lower X-axis represent the percentages of different types of repetitive elements in the genomes. The WGD event is marked with a red star in the species tree.

917 (B) Size distribution of introns that are conserved in Llu and Pae. Each dot represents an
918 intron, and the histograms on the top and to the right show the size distribution of introns
919 in Llu and Pae, respectively.
920 (C) Repetitive element content in gene flanking regions in six species of Orobanchaceae.
921 (D) Comparison of chromosomal repetitive elements and gene positions in chromosome
922 9 of *Lindenbergia luchunensis* and chromosome 5 of *Orobanche cumana*. Black lines in
923 the boxes indicate the positions of coding genes on the chromosomes; area graphs
924 indicate the proportion of different types of repeat sequences in the corresponding
925 chromosomal regions; curved lines connect orthologous gene pairs. The length of the bin
926 used for plotting is 1/1000 of the total chromosome length.

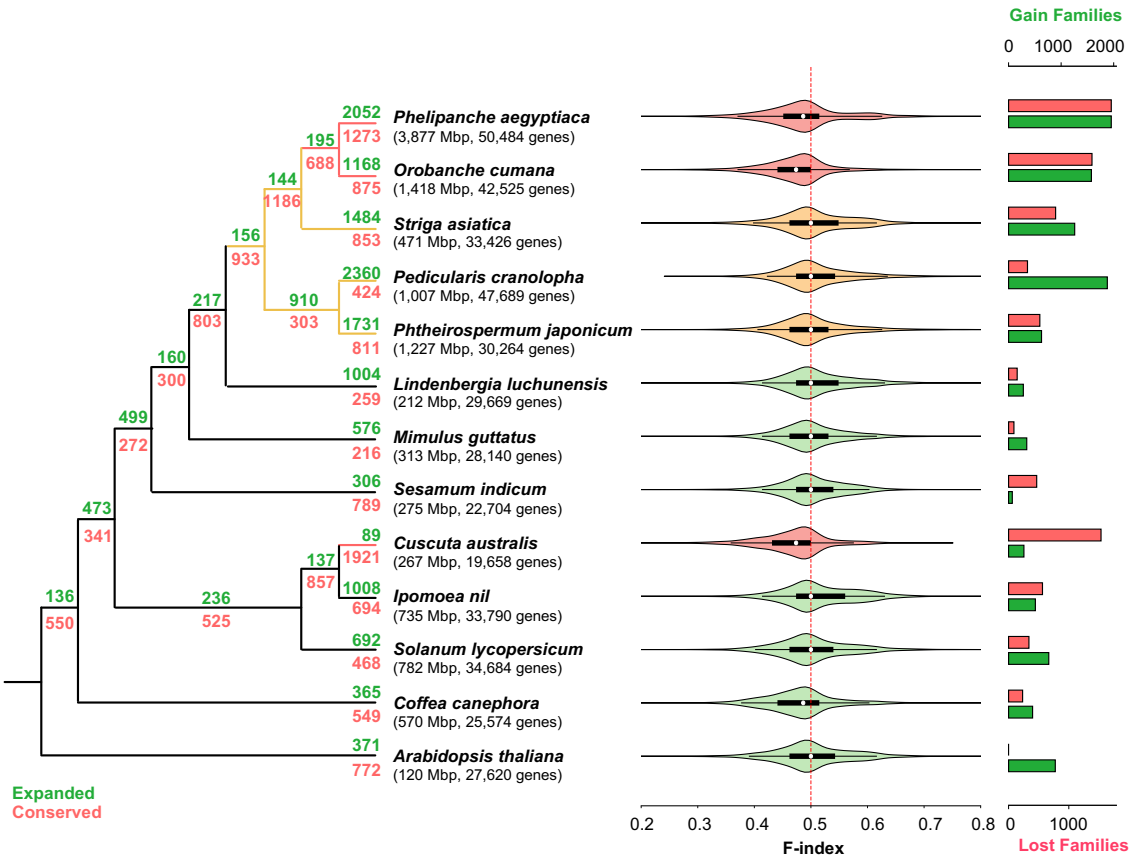


Figure 4. Gene family expansion and contraction in the Orobanchaceae.

On each branch of the species tree, the green number indicates the numbers of expanded gene families, and the red number indicates the number of conserved gene families. Each density curve in the violin plot shows the F-index (see Supplemental Methods) distribution of conserved gene families in each of the species indicated to the left. In the density curves, soft red, soft orange, and grayish green indicate holoparasitic, hemiparasitic, and autotrophic plant species, respectively. The bar plot on the right shows the numbers of conserved gene families which were completely lost in each species (red bars with lower X-axis) and the numbers of species-specific gene families (green bars with upper X-axis).

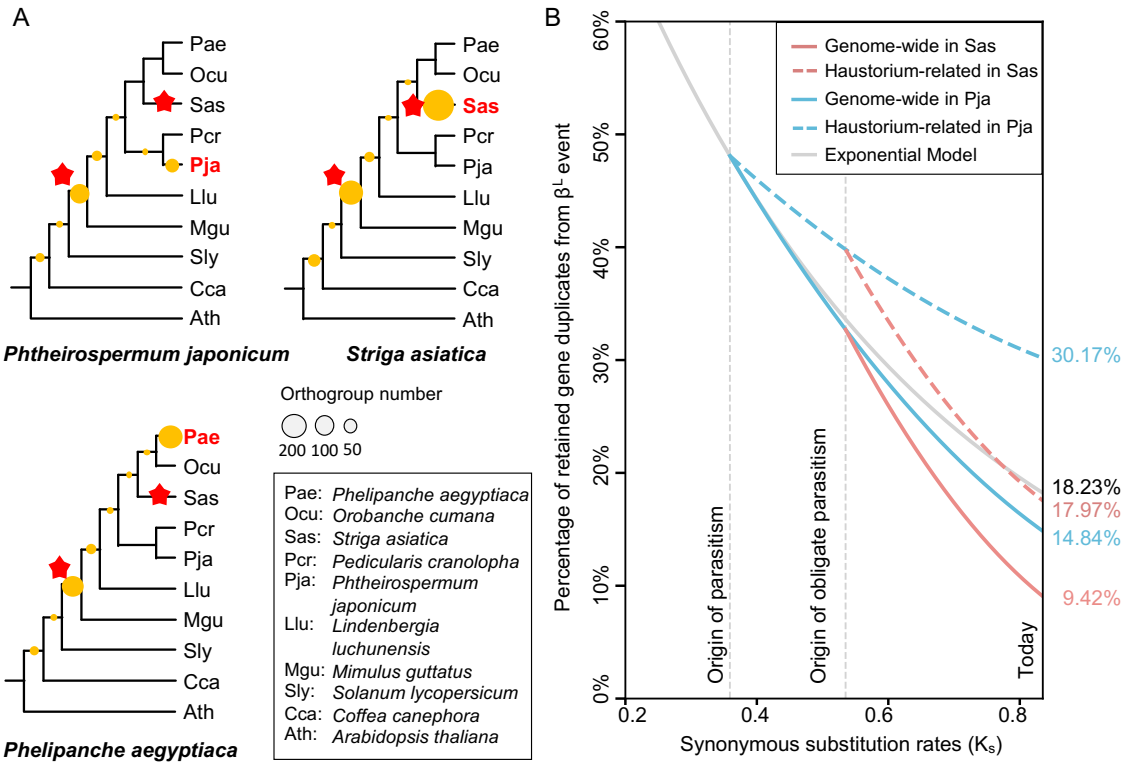


Figure 5. The role of the ancient WGD event β^L in the emergence and evolution of parasitism in the Orobanchaceae.

(A) Phylogenetic placement of gene duplications detected in the highly expressed genes in the haustorium. The highly expressed genes in the haustorium of each species were identified. Gene trees were constructed for all the gene families comprising these highly expressed genes. The duplication events were inferred from the topologies of the gene trees. The size of the yellow circle on each tree node represents the number of gene duplication events occurred during that time. The WGD events are marked with red stars in the species tree.

(B) Loss of gene duplicates after the β^L WGD event in *Striga asiatica* and *Phtheirospermum japonicum*. The X-axis (K_s) represents time after the β^L WGD event (see Supplemental Figure 5 for details of calculation). The gray curve represents the predicted values based on an exponential model under neutral selection (Supplemental Methods). Red and blue curves were inferred from the predicted initial values and the

958 observed end values respectively in *Striga asiatica* and *Phtheirospermum japonicum*
959 (Supplemental Table 10).
960
961

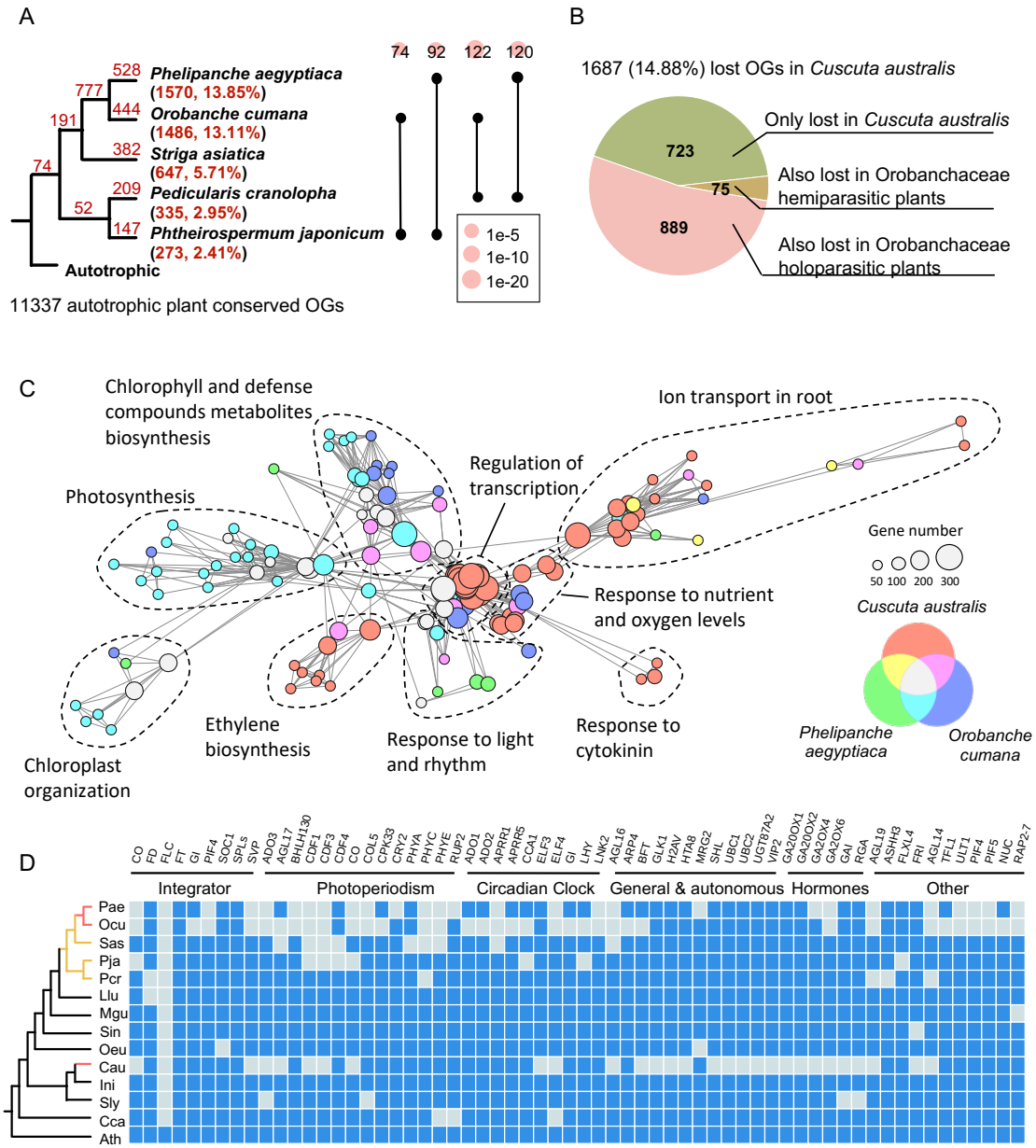


Figure 6. Gene loss in the orobanchaceous parasitic plants and *Cuscuta australis*.

(A) History and convergent evolution of gene loss in orobanchaceous parasitic plants. Left: the red numbers on the branches of the species tree represent the numbers of lost OGs estimated based on the principle of maximum parsimony; the number and percentage in brackets under the species name indicate the total number of lost OGs and the proportion of total number of lost OGs to the number of conserved OGs in

971 autotrophic plants, respectively. Right: the numbers on top indicate the numbers of the
972 same lost OGs between the two species indicated by the dumbbells after their divergence
973 from the MRCA; the size of the soft red circle on each number indicates the *p*-value of
974 hypergeometric test.
975 (B) The numbers of commonly and specifically lost OGs in *Cuscuta australis* and
976 orobanchaceous hemiparasitic and holoparasitic plants.
977 (C) Gene ontology (GO) enrichment analysis of lost OGs in *Cuscuta australis*,
978 *Phelipanche aegyptiaca,* and *Orobanche cumana*. Each node in the networks plot
979 represents the GO term of a biological process that was enriched (p-value < 0.02), and its
980 size indicates the number of OGs associated with the GO term. Nodes were assigned
981 colors depending on whether the gene loss occurred specifically or in common with other
982 species; each edge represents the similarity between two GO terms based on the number
983 of common OGs.
984 (D) Loss of genes related to flowering time regulation in orobanchaceous parasitic plants
985 and *Cuscuta australis*. Grey and blue boxes represent absence and presence of the
986 specific gene, respectively. Pathway information was retrieved from FIOR-ID
987 (http://www.phytosystems.ulg.ac.be/florid/).
988

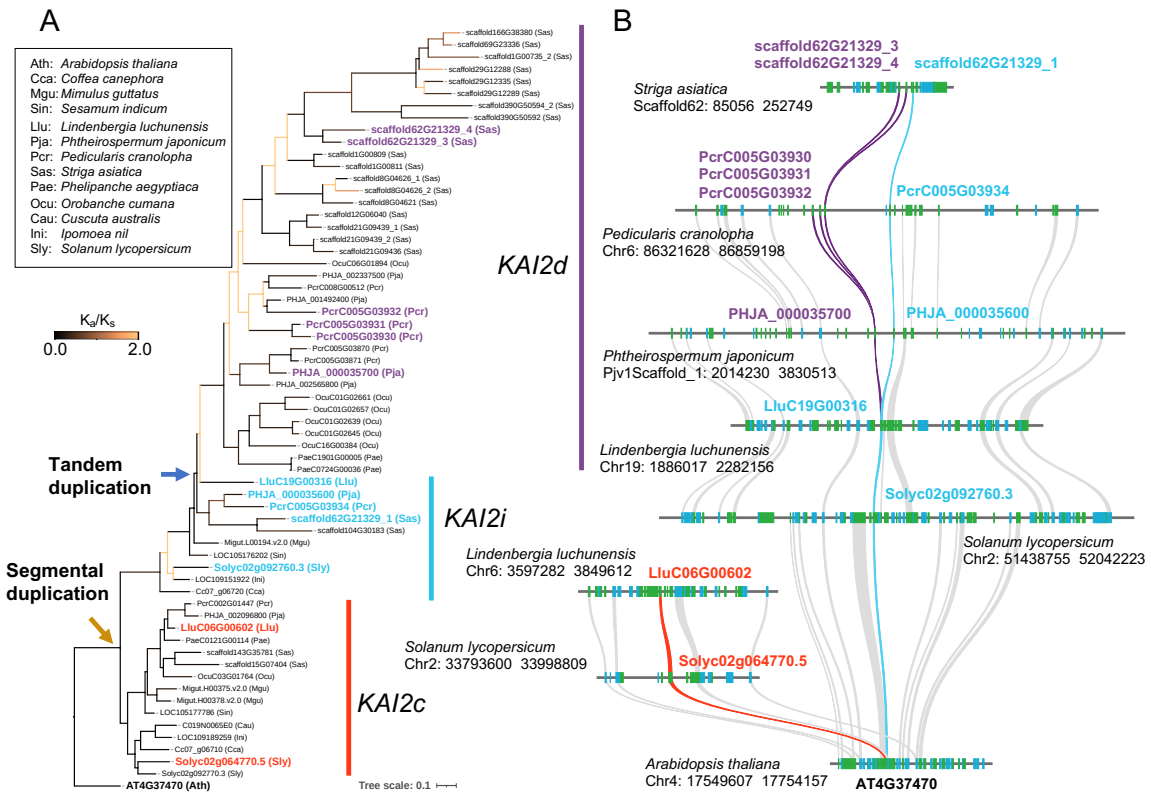


Figure 7. Molecular evolution of the *KAI2* gene family.

(A) The phylogenetic tree of the *KAI2* family in 13 species. Blue and orange arrows indicate duplication events, and the colors of branches represent the values of K_a/K_s. (B) The collinearity relationships of chromosomal fragments containing the KAI2 gene in parasitic plants of the Orobanchaceae and closely related autotrophic species. Syntenic gene pairs are connected by curves. Curves connecting orthologs of *KAI2c*, *KAI2i*, and *KAI2d* genes are highlighted in different colors.