Transcriptomes of the Parasitic Plant Family Orobancheaceae Reveal Surprising Conservation of Chlorophyll Synthesis

Norman J. Wickett, 1, 5, 6 Loren A. Honas, 1 Eric K. Wafula, 1 Malay Das, 2, 6 Ken Huang, 3 Biao Wu, 4 Lena Landherr, 1 Michael P. Timko, 2 John Yoder, 3 James H. Westwood, 2 and Claude W. dePamphilis 1, 6

1 Department of Biology, Institute of Molecular Evolutionary Genetics, and Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA
2 Department of Plant Pathology, Physiology, and Weed Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
3 Department of Biology, University of Virginia, Charlottesville, VA 22904, USA
4 Department of Plant Sciences, University of California, Davis, CA 95616, USA
5 Present address: Chicago Botanic Garden, Glencoe, IL 60022, USA
6 Present address: Institute of Biochemical Plant Pathology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Correspondence: nwickett@chicagobotanic.org (N.J.W.), cwd3@psu.edu (C.W.d.)

Summary

Parasitism in flowering plants has evolved at least 11 times [1]. Only one family, Orobancheaceae, comprises all major nutritional types of parasites: facultative, hemiparasitic (partially photosynthetic), and holoparasitic (nonphotosynthetic) [2]. Additionally, the family includes Lindenbergia, a nonparasitic genus sister to all parasitic Orobancheaceae [3–6]. Parasitic Orobancheaceae include species with severe economic impacts: Striga (witchweed), for example, affects over 50 million hectares of crops in sub-Saharan Africa, causing more than $3 billion in damage annually [7]. Although gene losses and increased substitution rates have been characterized for parasitic plant plastid genomes [5, 8–11], the nuclear genome and transcriptome remain largely unexplored. The Parasitic Plant Genome Project (PPGP; http://ppgp.huck.psu.edu) [2] is leveraging the natural variation in Orobancheaceae to explore the evolution and genomic consequences of parasitism in plants through a massive transcriptome and gene discovery project involving Triphysaria versicolor (facultative hemiparasite), Striga hermonthica (obligate hemiparasite), and Phelipanche aegyptiaca (Orobanchaceae [12]; holoparasite). Here we present the first set of large-scale genomic resources for parasitic plant comparative biology. Transcriptomes of above-ground tissues reveal that, in addition to the predictable loss of photosynthesis-related gene expression in P. aegyptiaca, the nonphotosynthetic parasite retains an intact, expressed, and selectively constrained chlorophyll synthesis pathway.

Results and Discussion

The transcriptomes of three parasitic plants representing three nutritional strategies in Orobancheaceae were sequenced:

Triphysaria versicolor, a facultative hemiparasite (here grown autotrophically); Striga hermonthica, an obligate hemiparasite; and Phelipanche aegyptiaca, a nonphotosynthetic holoparasite. Although previous studies have reported cDNA sequences and expressed sequence tag (EST) analyses of single parasitic plant species [13, 14], the results here represent the first comparative transcriptome analysis of parasitic species that differ in their photosynthetic capacity and significantly advance the genomic resources available to study parasitic plant biology.

EST Assembly and Annotation

In total, 1,278,054 454 FLX reads and 102,521,594 Illumina GAIIx paired reads were generated from above-ground, reproductive, and vegetative shoots of T. versicolor, S. hermonthica, and P. aegyptiaca (see Table S1 available online). Combined assemblies for each species resulted in 51,479 to 67,794 unigenes (>200 bp) with N50 values ranging from 663 to 952 (Table S2) and are available from the Parasitic Plant Genome Project (PPGP) database (http://ppgp.huck.psu.edu). Unigenes were sorted into an objective classification scheme (http://fgp.bio.psu.edu/tribedb/10_genomes) [15] that assigns each unigene to a SuperTribe, Tribe, and Orthogroup along with associated annotations. Membership in Tribes (global approximations of gene families) and Orthogroups (potentially narrower lineages stemming from a single ancestral gene in the reference species set) were used to investigate the overall classification of the transcriptomes.

The functional classification of genes in these three parasites revealed two primary patterns: first, a general, global similarity of functional categories across all three species (Figure 1; Figure S1), and second, numerous fine-scale differences among them (Figure 1; Figure 2). A large degree of similarity is expected in the transcriptomes of different plant cells regardless of function or photosynthetic ability; whether the large number of Tribes and Orthogroups shared between the two hemiparasites reflects their nutritional strategies requires a more careful dissection, as we have done here, of the functional annotations assigned to overlapping sets of unigenes. The complete summary of annotated unigenes with GO terms is provided in Table S3A, Table S3B, and Table S3C.

Read Mapping, Evolutionary Constraints, and Chlorophyll Synthesis

For each set of overlapping and uniquely expressed Orthogroups (Figure 1A), the proportion of the total number of Orthogroups assigned to each term for each Gene Ontology (GO) was plotted in order to scan for GO terms that are over- or underrepresented in each overlapping set (Figures 1B–1D). If the set of Orthogroups expressed in all species (sector A in Figure 1A) is taken as the “background” distribution of Orthogroups across the GO Slim terms, then extreme deviations from this can be easily visualized. The most obvious of these is the Chloroplast term of the Cellular Component ontology, for which the Orthogroups expressed in both the autotrophically grown hemiparasite and the obligate hemiparasite (S. hermonthica and T. versicolor; sector D in Figure 1A)
account for three times the proportion of total unigenes as in the background.

For each species, Illumina reads were mapped to the combined assembly in order to approximate expression levels. Using the mapping information for each unigene, a normalized number of reads (counts per million total mapped, cpm) per Orthogroup was calculated (Table S4). Because the three plants sequenced represent three points along a photosynthetic spectrum, Orthogroups corresponding to structural components of the photosynthetic apparatus were extracted from the read mapping results. Of the Photosystem I and II reaction center proteins, structural

Figure 1. Proportion of Orthogroups Assigned to Gene Ontology Terms for Unigenes Expressed in One, Two, or Three of the Parasitic Plants *Phelipanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*

(A) Proportion of PlantTribe’s Orthogroups assigned to GO Slim terms for assembled cDNA sequences (unigenes) expressed either uniquely in one of three parasitic plants (sectors E, F, and G; *Phelipanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*), expressed in two of three plants (sectors B, C, and D), or expressed in all three plants (sector A).

(B–D) For each section of the Venn diagram in (A), the proportion of Orthogroups assigned to a specific GO Slim term (per total number of Orthogroups in that section) was plotted for the Cellular Component (B), Biological Process (C), and Molecular Function (D) ontologies. See also Table S3A, Table S3B, and Table S3C.
components of the cytochrome b6/f complex and photosynthetic electron transport, and F-type ATPases identified from the Arabidopsis reference pathway (http://www.genome.jp/kegg/), all Orthogroups showed reduced representation in S. hermonthica compared to T. versicolor, with no expression detected in P. aegyptiaca (Figure 2B). The absence of detectable transcripts in Phelipanche may be due either to the absence of expression or to large-scale gene losses and pseudogenization as has been observed in parasitic plant plastid genomes [8–10, 16, 17].

In contrast to what was observed for structural components of the photosynthetic reaction centers and electron transport chain, expression of the chlorophyll biosynthesis pathway was detected in all three species. The number of mapped reads is reduced for many genes in the nonphotosynthetic P. aegyptiaca (Figure 2A) but is still represented by measurable read depths of between 5 and 615 cpm. Furthermore, reconstructed transcripts of chlorophyll biosynthesis genes include a full-length reading frame that can be aligned to their closest homologs in the hemiparasites. In order to look for evidence of evolutionary constraint in these protein-coding sequences in each species, the value of dN/dS (ω) was calculated for ten genes encoding enzymes in either the shared heme/chlorophyll synthesis pathway or the branch of this pathway dedicated to chlorophyll synthesis (Figure 2A; Table 1). Using a likelihood ratio test to determine whether the likelihood of the data is significantly better when ω is free to vary relative to nonholoparasites, we see that only coproporphyrinogen oxidase is characterized by a (weakly) significant relaxation of purifying constraint (increase in ω).

Our data demonstrate that the major chlorophyll binding proteins of the photosynthetic apparatus (photosystem reaction centers, light-harvesting antenna complexes; Figure 2B) are likely not expressed in the above-ground tissues of P. aegyptiaca. If chlorophyll synthesis is occurring, the lack of photosynthetic components necessary to stabilize it could be the basis for its exceptionally low abundance. Thus, our findings are consistent with previous studies reporting that some nonphotosynthetic Orobanchaceae produce an extremely low but detectable amount of chlorophyll a [18]. Additionally, we show that the enzyme responsible for chlorophyll b synthesis, chlorophyllide a oxygenase [19], is expressed in all three parasites, although at a low level in P. aegyptiaca. Detection of chlorophyll b has not been demonstrated by high-performance liquid chromatography in Orobanchaceae [18]. In the absence of producing structural components for light harvesting (e.g., reaction center light-harvesting complex [LHC] proteins), selective pressure may be imposed against expressing the chlorophyll biosynthesis pathway enzymes and expending energy to make chlorophylls.
that cannot be utilized. If chlorophylls a and b are produced in *P. aegyptiaca* and are not participating in photosynthesis, one might wonder whether chlorophylls are serving an alternative role in the plant.

In addition to giving rise to chlorophylls, the porphyrin biosynthetic pathway produces hemes and their derivatives, which function in photosynthesis and as cofactors in enzymatic reactions and components in nonphotosynthetic electron transport. The biosynthetic steps from aminolevulinic acid (ALA) formation through protoporphyrin IX (ProtoIX) are in common for both heme and chlorophyll biosynthesis, with the Fe and Mg chelation steps marking the branchpoint between the two [20]. Six enzymatic reactions are required to convert ProtoIX, the last common intermediate of heme and chlorophyll formation, into chlorophyll a (Chla). Transcripts encoding all six of these enzymes are expressed in all three parasites. Although this is not surprising in *Striga* and *Tryphysaria*, given that both are photosynthetic upon emergence, the detection of transcripts encoding all of the enzymes involved in chlorophyll formation in the holoparasite *P. aegyptiaca*, generally considered nonphotosynthetic, is unexpected. Moreover, our analysis of transcript assemblies is consistent with purifying selection, strongly suggesting that the *P. aegyptiaca* transcripts could be translated into functional enzymes for chlorophyll synthesis, raising several intriguing questions.

It is well documented that the buildup of porphyrins and chlorophyll intermediates, such as protochlorophyllide (Pchlide), can be deleterious as a result of their strong photosensitizing properties [21]. In dark-grown plants, the ratio of bound Pchlide (photoconvertible pigment-protein complexes formed with the protochlorophyllide oxidoreductase [POR] isozymes) to free Pchlide (F631) is important because the bound form is photoprotective and suppresses photooxidative damage. If *P. aegyptiaca* accumulated even small amounts of Pchlide during its growth in the dark, when exposed to light on emergence this would be deleterious to growth. The capacity to avoid energy transfer from Pchlide to triplet oxygen is through the transformation of Pchlide into Chlide as part of Chlide-POR-NADPH ternary complexes, which are strong energy quenchers, leading to avoidance of reactive oxygen species (ROS) formation. Thus, having expression of all steps up to POR formation and photoconversion would be important.

The tetrapyrrole pathway is regulated by metabolic intermediates at the transcriptional and posttranslational levels [22], and tetrapyrrole synthesis must be regulated to prevent damage to the photosynthetic machinery. When photooxidative stress accelerates the degradation of pigment-protein complexes, the synthesis of chlorophyll must slow down to compensate. It may be that because all of the steps are present, but the end-product chlorophyll is not stabilized, the pathway runs at the bare minimum. But why maintain it at all and risk photooxidative damage if it is not being used for photosynthetic activity? One possibility is that expression of the pathway is essential for functions other than photosynthesis. Experimental evidence has accumulated for the specific involvement of plastid-derived signals including tetrapyroles, redox status, and ROS controlling nuclear transcriptional activities (retrograde signaling) for plastid development as well as antiretrograde responses, where the nucleus is controlling plastid genes [23]. If porphyrin intermediates were important for signaling, then the plant would have to deal with the presence of these molecules in the absence of their use in photosynthesis.

Although the detection of an expressed chlorophyll synthesis pathway is surprising, it is notable that some (not all) holoparasitic Orobanchaceae have also been shown to retain and express conserved plastid *rbcL* genes [24–26]. Although functions for *rbcL* that are alternative to its primary role of CO₂ fixation have yet to be uncovered, the pattern of evolutionary conservation [24, 25] and evolutionary constraint following the loss of photosynthesis [27] presents a compelling case that some genes in holoparasites appear to have unexpected or additional functions beyond those normally associated with their primary role in photosynthetic plants. In this case, we do not detect any expression of *rbcS* in *P. aegyptiaca*, whereas tens of thousands of *rbcS* reads are detected in *T. versicolor* and *S. hermonthica*. Additionally, transcripts of *rbcL* are found in our data from both *T. versicolor* and *S. hermonthica*, but not from *P. aegyptiaca*.

### Evolution of Parasitism

The species in our study represent three distinct points along a range of parasitic ability, from facultative (*Tryphysaria*) to obligate hemiparasite (*Striga*) to holoparasite (*Phelipanche*). However, the phylogenetic relationships between these
species have not been clearly resolved [28]. To this end, we used a concatenated alignment of 339 putative single-copy nuclear genes, comprising 510,875 characters, to reconstruct the relationships between our study species, a preliminary assembly of the nonparasitic Lindenbergia philippensis (data stored and available at http://ppgp.huck.psu.edu), representing the basal lineage of Orobanchaceae, and the sequenced genome of the closely related nonparasitic species Mimulus guttatus (http://www.phytozome.net). Although supermatrix approaches to phylogenetic reconstruction such as this can suffer from missing data, the percent coverage of these species in the supermatrix (between 84% and 96%) is excellent and likely mitigates the effect of missing sequence data [29, 30].

Here, the resolution of Lindenbergia sister to parasitic Orobanchaceae is consistent with other analyses [3, 6, 31, 32], as is the inability to resolve, with strong support, the relationship between Triphysaria, Striga, and Phelipanche [28]. This unresolved topology may reflect a rapid radiation of ancestral Orobanchaceae following the shift to parasitism. Geographic shifts, rather than morphological innovations, may disproportionately increase diversification rates [33]; for parasites, this may be true if one considers the range of potential hosts as unfilled geographical niches. Sampling cDNAs from a more exhaustive set of Orobanchaceae may allow us to more accurately resolve the timing of the origin and diversification of parasitism in Orobanchaceae, and the potential role and timing of genome duplications.

Figure 3. Phylogenetic Relationships Inferred from 339 Nuclear Genes and Ancient Genome Duplication Events Inferred from the Distribution of $K_s$ for Paralogs

Hypothesis of relationships of the three parasitic plants included in this study (Triphysaria versicolor, Striga hermonthica, and Phelipanche aegyptiaca) and outgroup taxa based on 100 maximum-likelihood bootstrap replicates (left) and distributions of $K_s$ values for duplicate gene pairs (right). All branches are supported at 100% unless noted otherwise. The plots on the right indicate the number of gene pairs (potential paralogs) assigned to each $K_s$ (synonymous substitutions per synonymous site) bin. The curves overlaid on the $K_s$ distributions indicate the number of significant components that best describe the distribution (minimize the Bayesian information criterion). For all three plots, a shared fourth component describing a steeper background plot was removed to better visualize the component curves. The red curve describes the "background" distribution of paralogs due to normal gene births and deaths in a genome. Components that describe a peak in gene pairs of $K_s$ other than the background curve (orange and blue curves) may indicate a potential ancient whole-genome duplication event.
Genome duplication has been implicated in the evolution of several major groups of plants [34–38]. Here, analyses of the rate of synonymous substitutions ($K_s$) for paralogous gene pairs, calibrated by lineage-specific $K_s$ from 339 single-copy genes, resulted in distributions consistent with the expected gene birth-death process (“background” distribution; red lines in Figure 3) [37, 39] and several older concentrations of gene duplications that might be indicative of ancient polyploidy events. S. hermonthica and P. aegyptiaca exhibit a strongly distinguishable peak with a mean $K_s$ of 0.32 and 0.38, respectively (Figure 3). Additionally, S. hermonthica is characterized by an older, diffuse peak with a mean $K_s$ of 0.87. T. versicolor, the facultative parasite, exhibits only the background distribution, suggesting the absence of ancient polyploidy. Because of the preliminary nature of the available data, it is difficult to reconstruct ancient genome duplication events in the nonparasitic genus of Orobanchaceae, Lindenbergia. However, the suggestion that the obligate parasites here may share an ancient genome duplication event that is not shared with the facultative parasite could reflect Searcy’s hypothesis [40] that proposed a marked increase in genetic information as new, autotrophic functions are acquired.

Conclusion
The results of the present study add a new layer of molecular evolutionary and gene expression analysis to emerging evidence pointing to a complicated set of constraints on the photosynthetic apparatus. For example, recent studies of mycoheterotrophs, plants that obtain fixed atmospheric carbon from fungi associated with another, autotrophic plant [41], suggest that the loss of observable chlorophyll may lag behind the loss of photosynthetic capacity as a plant evolves further toward a completely heterotrophic life history [42], or that chlorophyll and photosynthesis can be simultaneously reduced in achlorophyllous forms of an otherwise chlorophyllous species [43]. The evidence presented here for constraints on the chlorophyll synthesis genes in the absence of detectable gene expression of crucial components of the photosynthetic apparatus suggests that the evolution of plants from autotrophic to partially autotrophic to fully heterotrophic is not a simple, linear trajectory. It also highlights the valuable contribution of genomic and transcriptomic data to our understanding of complex physiological processes.

Accession Numbers
The NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) accession numbers for the sequences reported in this paper are SRRX040930, SRRX041138, SRRX005614, and SRRX008135 (Triphysaria versicolor); SRRX040928, SRRX040929, SRRX006132, and SRRX006133 (Striga hermonthica); and SRRX040924, SRRX040925, SRRX006130, and SRRX006131 (Phelipanche aegyptiaca).

Additional Information
Supplemental Information includes one figure, four tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cub.2011.11.011.

Acknowledgments
Support by the National Science Foundation Plant Genome Research Program (IOS #0701748 to J.H.W., C.W.d., M.P.T., and J.Y.) for our work on the Parasitic Plant Genome Project, with additional support from the US Department of Agriculture (Hatch project number 135657 to J.H.W.), is gratefully acknowledged. We thank Danny Joel for Phelipanche seeds, Paula Ralph and Verlyn Stromberg for laboratory assistance, Lynn Tomsho and Stephen Schuster for 454 sequencing, and Yongde Bao for Illumina sequencing. The Mimulus guttatus sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the user community.

Received: February 12, 2011
Revised: August 17, 2011
Accepted: November 4, 2011
Published online: December 8, 2011

References


