The Parasitic Plant Genome Project: New Tools for Understanding the Biology of 
Orobanche and Striga

Author(s): James H. Westwood, Claude W. dePamphilis, Malay Das, Mónica Fernández-Aparicio, Loren 
A. Honaas, Michael P. Timko, Eric K. Wafula, Norman J. Wickett, and John I. Yoder


Published By: Weed Science Society of America

DOI: http://dx.doi.org/10.1614/WS-D-11-00113.1

URL: http://www.bioone.org/doi/full/10.1614/WS-D-11-00113.1
Parasitic plants of the Orobancheaceae are among the most devastating agricultural weeds. Witchweeds (Striga spp.) and broomrapes (Orobanche and Phelipanche spp.) are especially destructive and affect large areas of the globe, including many developing nations, where they have potential to greatly decrease yield and quality of host crops. At present, over 50 million ha of the arable farmland under cultivation with cereals and legumes in sub-Saharan Africa are infested with one or more Striga species, resulting in annual yield losses that are estimated to exceed US$10 billion (Ejeta 2007; Scholes and Press 2008). The impact of Orobanche parasitism on yields typically ranges from 20 to 100% crop loss, depending on the infestation (Barker et al. 1996; Bernhard et al. 1998; Parker and Riches 1993; Sauerborn 1991). Although accurate numbers on infested acreage and yield losses are difficult to obtain, it is evident that despite efforts aimed at their control, both witchweed and broomrape persist and are expanding their negative impact on crop productivity (Parker 2009). The seeds may lie dormant in the soil for years, only germinating in the presence of potential host species that exude specific chemical signals. The parasite seedling must then attach to the host and form vascular connections to access water and other resources required for growth. Each of these steps involves exquisite communication between parasite and host that represent both fascinating biological adaptations and potential points of weakness that can be targeted for parasite control. However, understanding of parasite development and mechanisms of host interaction has been constrained by a lack of parasite gene sequence information that has lagged behind that of model and crop species (Torres et al. 2005; Westwood 2001). To fill this gap, the Parasitic Plant Genome Project (PPGP) was initiated with the goal of discovering as much of the transcribed gene sets as possible from three parasitic species of the Orobancheaceae, including one representative each from the witchweed and broomrape groups. In this paper, the design of the PPGP, its major outputs, and the potential of the emerging transcriptomic data to contribute to control of parasitic weeds is discussed.

Rationale and Approach of the PPGP

Selection of the Focal Species. An important consideration at the start of the PPGP was selecting the specific species for

DOI: 10.1614/WS-D-11-00113.1

* First, third, and fourth authors: Associate Professor and Postdoctoral Scholars, Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, VA 24061; second, fifth and seventh authors, Professor, Graduate Student, and Postdoctoral Scholar, Department of Biology, Institute of Molecular Evolutionary Genetics, and Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802; sixth author: Professor, Department of Biology, University of Virginia, Charlottesville, VA 22904; eighth author: Professor, Department of Plant Sciences, University of California, Davis, CA 95616. Current address of third author: Institute of Biochemical Plant Pathology, Helmholtz Zentrum München, 85764 Neuherberg, Germany. Additional address for fourth author: Institute for Sustainable Agriculture, IAS-CSIC, Dept. of Plant Breeding, Córdoba, 14080, Spain. Corresponding author’s E-mail: westwood@vt.edu

The Parasitic Plant Genome Project: New Tools for Understanding the Biology of Orobanche and Striga

James H. Westwood, Claude W. dePamphilis, Malay Das, Mónica Fernández-Aparicio, Loren A. Honaas, Michael P. Timko, Eric K. Wafula, Norman J. Wickett, and John I. Yoder*
study. In addition to the strong economic justification for studying witchweeds and broomrapes, parasites of the Orobanchaceae are excellent subjects for study because this family is unique among parasitic plant lineages in containing genera that span the full spectrum of host dependence from free-living autotrophs to obligate holoparasites (Westwood et al. 2010). This creates a comparative framework for evolutionary analysis that provides insight into mechanisms underlying the unique features of parasitism. In addition, the species chosen also have been relatively well studied and methods for their specific manipulation under controlled laboratory conditions have been developed.

The broomrapes represent the evolutionary extreme of heterotrophy, being holoparasites that lack expanded leaves and photosynthetic capacity and therefore are completely host dependent. For the broomrapes, important weedy species considered were Orobanche crenata (Forsk.), O. cumana (Wallr.), O. minor (Sm.), O. aegyptiaca, and O. ramosa (L.), which attack many important dicotyledonous crop plants including food legumes in North Africa, Europe, the Middle East and West Asia. Orobanche aegyptiaca was ultimately selected because of its broad host range that includes the model species tobacco (Nicotiana tabacum L.), tomato (Solanum lycopersicum L.), and Arabidopsis thaliana (L.) Heynh. (Goldwasser et al. 2002; Westwood 2000), and a large body of scientific literature. Parasite species that can be cultivated on host plants with sequenced genomes and transformation systems will speed experimental studies of host–parasite interactions, and have been the focus of our work.

The witchweeds are similar to broomrapes in being host dependent at the early stages of parasite development. Once germinated and attached to a suitable host, the developing seedlings rapidly emerge from belowground, accumulate chlorophylls, and produce leaves capable of photosynthesis. Witchweeds (S. hermonthica, S. aspera (Wildl.) Benth., and S. asiatica (L.) Kunze) occur mainly in sub-Saharan Africa, India, and Southeast Asia where they affect primarily cereal grass crops maize (Zea mays L.), sorghum (Sorghum bicolor (L.) Moench ssp. bicolor), rice (Oryza sativa L.), and millets (Pennisetum spp.). Another important species, S. genierioides (Willd.) Vatke, attacks primarily leguminous crops (Mohamed et al. 2001). Although autogamous species are attractive for genomic studies because the use of inbred lines facilitates gene and genome assembly, we selected the outcrossing species S. hermonthica because of its greater economic impact in world agriculture. Host species used were maize and sorghum, both of which have sequenced nuclear genomes (Paterson et al. 2009; Schnable et al. 2009).

Triphysaria is a genus of five hemiparasitic species common in grasslands throughout the Pacific Coast (Chuang and Heckard 1991; Hickman 1993). Triphysaria was included because it is a facultative hemiparasite, able to live without a host if necessary, but possessing the specialized machinery that enables it to opportunistically parasitize roots of neighboring plants. Like some other hemiparasites, Triphysaria is a generalist with a very broad host range including monocots and dicots, facilitating comparisons between this plant and Striga and Orobanche growing on the same hosts. Triphysaria versicolor, an outcrossing species, was used in this project.

For comparison we included two nonparasitic relatives as outgroups. One is the basal Orobanchaceae species Lindenbergia philippensis, the other is Mimulus guttatus (Monkeyflower). Mimulus is a genus of traditional Scrophulariaceae (now Phrymaceae) that includes about 160 species with worldwide distribution (Olmstead et al. 2001). Mimulus is a popular genus for ecological investigations since it has a wide distribution and numerous genetic resources already available (Hall and Willis 2005), including a sequenced genome (Wu et al. 2008).

**Tissue Collection Strategy.** The sequencing strategy was based on next-generation sequencing of tissue-specific and whole plant transcriptomes to discover the largest set of expressed sequences for important protein coding genes. This is the most cost-effective approach for discovery and characterization of most expressed genes in plants (Wall et al. 2009). The relatively large size of the parasite genomes (Westwood et al. 2010) makes full genomic DNA sequence, assembly, and annotation a challenging task (Imelfort and Edwards 2009) but recent advances in genome sequencing and de novo assembly are making this an attainable goal. For transcriptome sequencing, messenger RNA (mRNA) was isolated and converted to complementary DNA for sequencing as described in Wickett et al. (2011).

To capture the broadest possible set of transcripts that represent the full range of gene expression in the focal species, it was necessary to include many different tissues and developmental stages. The project strategy was to concentrate most effort on collecting stages involving haustorium development and function because the haustorium is the unique organ that defines parasitic plants, and because the haustorium is a likely target for approaches to interfere with the parasite–host interaction. However, tissues from additional life stages were also included to complete the coverage of gene expression, providing valuable points of comparison to haustorial stages and because parasitic plants exhibit specialized metabolism, signaling, and development at multiple points in their life cycle (Westwood et al. 2010; Yoder 2001). Life cycles of Triphysaria, Striga, and Orobanche were divided into stages on the basis of the biology of parasitism that were comparable among the three species (Figure 1). Although these stages may appear to be quite specific, it is important to recognize that each stage was in turn comprised of pools of cells from different tissues at different developmental substages, again with the rationale of sequencing the maximum number of expressed genes.

The unique goal and strategy of the PPGP led to the development of a novel system for categorizing stages (Figure 1), and understanding stage composition is essential for interpreting results from the project. Stage 0 includes seeds that have been imbibed for various lengths of time, covering the period of conditioning as well as seeds that have been stimulated to germinate by treatment with a germination stimulant, GR-24 (Magnnus et al. 1992) for up to 6 h. For S. hermonthica and O. aegyptiaca, this is a key step in development because it is tied to perception of a stimulant exuded by a host root (Joel et al. 1995; Yoneyama et al. 2008). For T. versicolor, which does not require a germination stimulant, seeds were imbibed but not germinated. Therefore, stage 0 contains expressed sequence tags important in preparing for—and responding to—germination triggers.

Stages 1 and 2 represent radicle growth before and after the addition of a haustorial inducing factor and so encompasses the transition from nonparasitic root growth to haustorium formation. Stage 1 consists of noninduced roots for T. versicolor and germinated seeds with radicles for S. hermonthica and O. aegyptiaca. Haustorial growth was induced in
vitro by addition of 2,6-dimethoxybenzoquinone (DMBQ) for *T. versicolor* and *S. hermonthica*, as these species are well known to respond to this chemical by initiating formation of a morphologically distinctive haustorium (Yoder 2001). *O. aegyptiaca*, on the other hand, does not develop the swollen, hairy haustorium typical of the other species, and the sensitivity of *Orobanche* to DMBQ is unclear. To ensure inclusion of stages encompassing haustorium formation, *O. aegyptiaca* seedlings (stage 1) were placed for 6, 12, and 24 h on roots of aseptically grown *Arabidopsis*. The resulting

<table>
<thead>
<tr>
<th>Stage</th>
<th>Developmental feature</th>
<th>Triphysaria versicolor</th>
<th>Striga hermonthica</th>
<th>Orobanche aegyptiaca</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Seeds conditioned and simulated; pre-germination</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>1</td>
<td>Root before haustorial initiation</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td>Root after haustorial initiation (exposed to haustorial induction factors or host roots)</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td>Haustorial penetration stages, pre vascular connection</td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td>Post vascular connection</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td>4.1</td>
<td>Early development</td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td>4.2</td>
<td>Late development</td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
</tr>
<tr>
<td>5</td>
<td>Pre-emergent tissues</td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td>5.1</td>
<td>Shoots</td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
</tr>
<tr>
<td>5.2</td>
<td>Roots</td>
<td><img src="image28.png" alt="Image" /></td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
</tr>
<tr>
<td>6</td>
<td>Post-emergent tissues</td>
<td><img src="image31.png" alt="Image" /></td>
<td><img src="image32.png" alt="Image" /></td>
<td><img src="image33.png" alt="Image" /></td>
</tr>
<tr>
<td>6.1</td>
<td>Vegetative structures leaves/stems</td>
<td><img src="image34.png" alt="Image" /></td>
<td><img src="image35.png" alt="Image" /></td>
<td><img src="image36.png" alt="Image" /></td>
</tr>
<tr>
<td>6.2</td>
<td>Reproductive structures: floral buds (up to anthesis)</td>
<td><img src="image37.png" alt="Image" /></td>
<td><img src="image38.png" alt="Image" /></td>
<td><img src="image39.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 1. Comparative illustration of the key life stages targeted for sequencing in the three parasitic species, *Triphysaria versicolor*, *Striga hermonthica*, and *Orobanche aegyptiaca*.
haustorial-induced *O. aegyptiaca* seedlings were collected by rinsing them from the *Arabidopsis* roots with water before they had firmly adhered to or penetrated the host roots.

Stages 3 and 4 cover the period of development related to haustorial penetration and maturation. Stage 3 parasites had firmly adhered to the host root and were in various stages of invasion, but had not formed a vascular connection to the host. In *S. hermonthica*, the hallmark of a successful connection is the unfolding of the cotyledons (Hood et al. 1998), and this was used as a marker for stage 4. For *O. aegyptiaca*, the indicator of vascular connection was the initiation of swelling of the tubercle, and this occurred approximately 8 to 10 d after addition of GR-24. Stage 4 in *O. aegyptiaca* was subdivided to capture the complexity of development in this parasite, with stage 4.1 representing early growth and typified by expansion of the tubercle before development of secondary roots. The transition to stage 4.2 occurs at about 12 d after the addition of GR-24 and is marked by the development of roots that lead this stage to be termed the “spider” stage (Fig. 1, *O. aegyptiaca* stage 4.2). This subdivision of stage 4 distinguishes gene expression in the tubercle body itself from that of root growth that involves processes aimed at forming new host attachments via secondary roots (Westwood et al. 2010).

Stage 5 represents underground growth and thus the continued fully heterotrophic dependence of the parasite on the host. For *S. hermonthica*, this phase includes shoot growth before emergence from the soil, and although *S. hermonthica* at this stage consists mostly of shoot, some roots develop adventitiously. Shoots and roots were collected separately and are represented by the subdivision of this stage into stage 5.1 (shoots) and 5.2 (roots). For *O. aegyptiaca* stage 5 is characterized by a mass of secondary roots arising from the tubercle (5.2) and a floral shoot (5.1). No equivalent to stage 5 exists for *T. versicolor* because once the haustorial connection is established, the root tip resumes normal growth and the process of detecting a new host root, developing a haustorium, and connecting to the host is repeated. The parasite in this case never experiences a belowground phase of complete host dependence.

Stage 6 encompasses all aboveground growth. Although emerged tissues lack the unique morphology of the haustorium, they provide a valuable perspective on parasite evolution through comparisons of photosynthesis-related processes that differ dramatically among the three parasite species. Specifically, whereas *T. versicolor* and *S. hermonthica* have apparently normal leaves, those of *O. aegyptiaca* exist as reduced scale structures. These vegetative stems and leaves were designated substage 6.1 (Figure 1). The one feature in which the parasite species would not be expected to show great differences is in genes specific to floral biology, so we sampled reproductive tissues represented by buds from the time of first appearance through anthesis (stage 6.2).

All of the stages described above were sequenced from nonnormalized libraries so that it would be possible to obtain quantitative information on mRNA abundance through read counts. In addition, aliquots of each developmentally specific stage were pooled and normalized libraries were made for each parasite species. A similar normalized library of *L. philippensis* contained RNA from seeds, roots, stems, leaves, and flowers. Normalized libraries are an efficient method to obtain low-copy-number transcripts (Wall et al. 2009; Zhulidov et al. 2004) and help to provide more consistent de novo assembly across the entire transcriptome.

Transcriptome sequencing was conducted using a combination of 454 and Illumina technologies (Zhang et al. 2011). Sequencing methods have been undergoing very rapid technological evolution and the PPGP was a beneficiary of these advances. Initial plans to use traditional Sanger sequencing were replaced by 454 FLX and later by Illumina. The sequence yield (Table 1) reflects the relative efficiencies of the methods. Initial 454 FLX half-plate runs produced about 50 MB of sequence data, and subsequent use of 454 titanium increased this output to about 150 MB per run. Illumina sequencing was the most cost-effective method and generated 2,500 to 4,000 MB in a single lane. Because of these advances and associated cost savings, the PPGP was able to conduct replicate sequencing runs on most tissues and expand the number of tissues sequenced beyond those included in the original plan.

**Transcriptome Assembly.** The greatest challenge created by the shift to Illumina sequencing was assembling contigs efficiently and accurately for organisms that lack reference genomes. We used two approaches to assemble the ultrahigh-throughput Illumina RNA seq reads collected for this project, both described in Wickett et al. (2011). Briefly, the first approach used the CLC assembly cell (CLC bio, 10 Rogers St # 101, Cambridge, MA 02142) to perform assemblies of a single lane of Illumina sequence data, as well as to assemble multiple libraries in a single step. A second approach used the NextGENe (SoftGenetics, LLC., 100 Oakwood Ave, Suite 350, State College, PA 16803) platform, where multiple iterations of the consolidation algorithm, followed by a maximum overlap assembly, assembled the most highly represented genes in the library. The raw reads were then mapped back to the assembled contigs, followed by the removal of the matched reads from the read pool. The consolidation step was run again using the unmatched reads, assembling the next most highly represented genes in the read pool. This cycle of skimming off classes of highly represented genes from the read pool was repeated until the consolidation step produced no contigs above a specific length threshold (200 base pairs).

We have used, among others, a measure of assembly quality based on the capture and coverage of a set of putatively conserved nuclear single copy genes (Duarte et al. 2010). From the PlantTribes 2.0 10-genome scaffold (http://fgp.bio.psu.edu/tribedb/10_genomes/index.pl, see also Wall et al. 2008), we identified 285 orthogroups (orthogroups approximate orthologous gene clusters) that contain one member from each of seven sequenced angiosperm genomes (*Arabidopsis, Carica papaya* (L.), *Medicago truncatula* (Gaertn.), *Oryza sativa* (L.), *Populus trichocarpa* (Torr. & Gray), *Sorghum bicolor*, and *Vitis vinifera* (L.)). Each set of assembled unigenes was sorted into the PlantTribes scaffold and those unigenes that were sorted into each of the 970 putative single-copy genes were identified. For each orthogroup in each assembly, the identified unigenes were translated against a reference protein (here, *V. vinifera*) using GeneWise. (GeneWise, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, U.K.). The translated unigenes were then aligned to the reference gene, a scaffold sequence was created, and the coverage of the reference gene
Applying Genomic Information to Parasitic Weed Control

The PPGP has produced a large amount of data that have value from both basic and applied research perspectives. Easily collected aboveground tissues (stages 5 and 6) were the first into the analysis pipeline and provide insights into parasite (number of positions filled/gene length) was reported. Figure 2 shows the coverage of the single copy gene set by the assembled aboveground libraries using the NextGENe consolidation method. Although more detailed analyses are in progress, these results indicate that the sequencing strategy has succeeded in identifying a large percentage of the parasite genes and providing good coverage of most genes.

Westwood et al.: Parasitic Plant Genome Project  •  299
NextGENe assemblies were analyzed against 970 genes. Interestingly, the nonphotosynthetic O. aegyptiaca appears to retain a set of genes for chlorophyll synthesis, despite the absence of expression of other key photosynthetic genes. It is also interesting that categorization of genes on the basis of their predicted function, cellular component, or metabolic process produced generally similar profiles in parasite species (and even compared with Mimulus), despite striking morphological differences among these species. We hypothesize that gene regulation, as mediated by cis and trans elements, will prove to be of equal or greater importance than the actual gene coding sequences in explaining unique features of parasite development. Identification and characterization of regulatory elements should be an objective for future research projects.

In addition to understanding parasite evolution, the goal of the PPGP is to contribute to the control of parasitic weeds. Considering that the growth and development of parasitic weeds are so closely intertwined with that of their hosts, advances in understanding the biology of parasitism should be particularly powerful for identifying parasite weaknesses and potential control points. The many levels of the parasite interaction with their hosts suggest that genomics tools could be applied in several different ways.

**Parasite Races and Crop Breeding.** Parasitic weed populations are intriguing because they exhibit specificity toward hosts and some parasite species have evolved races with virulence specific to host cultivars. The transcriptome data from the PPGP provide sequence information that can be used for the development of molecular markers to assist in characterizing parasite populations. Knowledge of parasite population diversity and host specialization will facilitate breeding of parasite-resistant crops.

Race-specific resistance is known to exist in both Striga–host and Orobanche–host interactions. Lane et al. (1994) were the first to report the existence of multiple races of Striga gesnerioides parasitic on cowpea (Vigna unguiculata L. Walp.), leading to the suggestion that a gene-for-gene mechanism may be operating in these Striga–host interactions. Support for this conclusion and the role of effector-triggered immunity comes from the recent cloning and characterization of a canonical coiled-coil–nucleotide-bridging site (NBS)–leucine-rich repeat (LRR) containing R protein required for race-specific resistance of cowpea to S. gesnerioides (Li and Timko 2009). Similarly, Vrâncaneanu et al. (1980) were the first to suggest that resistance responses in O. cumana–sunflower interactions were governed by a gene-for-gene mechanism and subsequently multiple pathogenic races of the parasite were identified that were recognized by the products of single dominant resistance genes (Letousy et al. 2007; Molinero-Ruiz et al. 2006), or dominant genes influenced by modifying genes that produce reversal of the expected dominance (Férez-Vich et al. 2004; Velasco et al. 2007). At present, the identity and nature of the O. cumana resistance genes in sunflower are unknown, but mapping studies have identified three NBS-LRR containing R-protein homologues that map close to one of the resistance genes, leading to the suggestion that a cluster of R genes may be involved in resistance to O. cumana (Fernández-Martínez et al. 2009; Radwan et al. 2008).

Apart from the O. cumana–sunflower and S. gesnerioides–cowpea interactions, resistance to other Orobanche and Striga spp. appears to be polygenic and not race specific and to be influenced by environmental factors (Díaz-Ruíz et al. 2009; Scholes and Press 2008). The polygenic resistance manifested in hosts may be reflective of a greater number and diversity of avirulence (Avr) factors and effectors present in the parasites, host specificity, and breeding habits of the parasite. Other factors on the host side may include the host genetic variability, breeding status, and time frame of coevolution with parasites. Regarding Orobanche, several levels of host-driven differentiation have been described in O. minor growing on Trifolium pretense L. and Daucus carota ssp. gumminfer (Thorogood et al. 2008, 2009), O. foetida on Cicer arietinum L. and Vicia faba L. (Román et al. 2007), and O. ramosa on Brassica napus L., Cannabis sativa L., and N. tabacum (Benharrat et al. 2005; Brault et al. 2007). It is possible that races have not evolved in these species because of the lack of selection pressure by the host due to the absence of highly resistant cultivars to any of these species.

The practical impact of parasite races on crop breeding is to frustrate efforts to develop sustainable resistance. In addition to the successive breakdown in host resistance by races of O.
cumana or S. gesnerioides described above, a recurring problem in breeding resistance to parasitic weeds is that the crop resistance developed through breeding at one location may not hold up when the crop is moved to new regions with different parasite populations. Overcoming this problem will require detailed characterization of parasite populations and the genes—or associated genetic markers—that contribute to host and race specificity.

**Host–Parasite Interactions.** The complex interactions between parasitic Orobanchaceae and their hosts present a daunting hurdle for research aimed at understanding the functions of single genes or small groups of genes in the parasitic interaction. Genome-scale studies are needed to dissect the interaction and focus on core processes involved in parasitism. The rich transcriptome resource developed in the PPGP will enable this type of research to be conducted on parasitic plants.

One approach to elucidating global gene expression profiles of crop hosts interacting with weedy Orobanchaceae has been the microarray, which relies on knowledge of gene sequences for probe design. To date microarrays have only been applied to understanding responses of hosts (with full genome sequences) to parasitism. For example, Swarbrick et al. (2008) profiled resistant and susceptible rice cultivars challenged with *Striga*, and Dita et al. (2009) used microarrays to analyze gene expression in two accessions of *Medicago* that display different modes of resistance to *O. crenata*. Smaller-scale approaches like suppression–subtractive hybridization have been used in nonsequenced hosts challenged with weedy Orobanchaceae to identify differentially expressed host genes (Die et al. 2007; Hiraoka and Sugimoto 2008; Hiraoka et al. 2009). However, all of this work has focused on the host response to parasitism, driven by the availability of well-developed genomic resources for these plants.

Data from the PPGP begins to fill the sequence information gap on the side of the parasite and provides a missing piece of the puzzle. Indeed, the groundwork for powerful comparative frameworks in which to study parasitism has been laid by years of carefully cataloguing and describing interactions of crop hosts and the weedy Orobanchaceae, with special attention to modes and degrees of host resistance (see Cardoso et al. 2011; Parker 1991, 2009; Perez-de-Luque et al. 2008; Rispail et al. 2007; Rubiales et al. 2006, 2009; Thorogood and Hiscock 2010; Yoder and Scholes 2010). Combining this existing catalogue of data on host–parasite interactions with new tools for high-throughput sequencing and de novo assembly and analysis holds the key to developing long-term control strategies. One application of PPGP sequence data could be the design of parasite-specific microarrays to make detailed studies of transcriptome changes during key developmental stages. Recent advances in sequencing technology (Marguerat and Bähler 2010; Ohtsu et al. 2007) have greatly reduced the cost of high-throughput sequencing, making it possible to accurately measure gene expression on a genome-wide scale by sequencing replicate RNA samples. Either approach could have broad utility in identifying and confirming which parasite genes are central to parasite growth and virulence.

**Pathogen Effectors from Parasitic Plants.** The vast majority of plant pathogens studied (including bacteria, fungi, oomycetes, and nematodes) secrete proteins, referred to as effectors, into plant cells to facilitate their ingress and establishment on their hosts (Block et al. 2008; Davis et al. 2008; Ellis et al. 2009; Tyler 2009; Zhou and Chai 2008). At present, the nature of pathogen effectors, their diversity, and their host targets are all areas of active investigation. Presuming that parasitic plant–host interactions are governed by the same rules as other plant pathogen–host interactions, whereby hosts use a combination of pathogen-/effector-triggered immunity in their defense response, one would like to know the nature of the Avr gene products or other effector molecules present in species such as *Triphysaria*, *Striga*, and *Orobanche*. That such factors indeed exist and evolve under host selective pressures is demonstrated by the observations that parasite races have emerged to overcome existing R-mediated detection in cowpea and sunflower (Alonso et al. 1996; Lane et al. 1994; Timko et al. 2007).

Comparative transcriptomics has been successfully used to identify candidate genes encoding Avr factors and effectors in a variety of other plant pathogens (Baxter et al. 2010; Bowen et al. 2009; Cramer et al. 2006; Raffaele et al. 2010; Roze et al. 2008; Schirawski et al. 2010; Spanu et al. 2010) and the application of high-throughput sequencing and computational approaches will likely provide significant novel information when applied to parasitic weeds. It is difficult to predict what Avr factor or effector genes from a parasitic plant will look like and how their encoded products might functionally annotate. There are many examples of the identification of effectors on the basis of their similarity to known proteins or from the presence of known domains or motifs. This includes similarities to effectors (usually Avr proteins) from other pathogens, the presence of eukaryote-like domains and motifs, or homology to gene products encoding anti-apoptotic proteins and inhibitors of programmed cell death. Equally likely is that they are unique proteins that contain previously uncharacterized domains. Alfano (2009) has recently reviewed various proven strategies for the identification and characterization of Avr factors and other effectors present in or released from the invading pathogens; it is likely that these strategies could be easily adapted to the identification of molecules of haustorial origin that facilitate (or attempt to facilitate) parasite establishment on a suitable host.

One method to identify potential parasite effectors is to take advantage of the observation that, in some cases, pathogens appear to temporally regulate the formation of their effectors during infection. The PPGP data sets lend themselves readily to analysis of temporal changes in gene expression during different critical stages of parasite development (i.e., haustorial formation, attachment, and postattachment growth and host penetration), thus allowing the identification of potential effector candidates. Further experiments could refine the time courses through appropriate use of approaches such as microarrays, subtractive hybridization, or gene expression through deep sequencing (RNA seq). In these experiments, examination of the changing dynamics of gene expression at the haustorium–host root interface could provide significant information for both potential parasite effectors as well as host targets. Understanding the complexity and diversity of virulence molecules (Avr proteins and effectors) present in *Striga* and *Orobanche* species could have an immediate impact on how breeders think about constructing strategies for pyramiding resistance.
Disrupting Specific Parasite Genes or Processes. Genetic Engineering and Gene-Silencing Strategies. In addition to supporting traditional breeding efforts with molecular markers, data from the PPGP project will provide resources for exploring novel genetic strategies to make crop plants resistant to parasitic weeds. One proven strategy for engineering pest resistance in crops has been to transform the plants with genes encoding pest-specific toxins; the now-classic example is transgenic plants expressing Bacillus thuringiensis crystal endotoxin that is specifically lethal to certain insect herbivores (van Frankenhuyzen 2009). A similar strategy was investigated for parasitic weeds using sarcotoxin IA from the flesh fly Sarcophaga pregrina as a pest-specific toxin. Whereas the sarcotoxin gene had little or no effect on transgenic tomato plants, there was a significant reduction in the number of O. aegyptiaca plants emerging from the pots containing transgenic compared with nontransgenic tomatoes (Hammouche et al. 2005).

The overriding limitation to transforming crop plants with pest-specific toxins is the paucity of molecules that are toxic to the pest but not the host. This is particularly tricky when both the host and pathogen are plants. A frequently discussed alternative is to use double-stranded RNAs as pest-specific, oligonucleotide toxins. The strategy is to transform the host plant with a vector bearing a hairpin RNA (hpRNA) that results in a double-stranded RNA homologous to one or more critical parasite genes. Double-stranded RNA molecules, formed by complementary base pairing of transgenic sequences, are processed by nucleases into short interfering RNAs (siRNA) that direct the degradation of endogenous mRNA transcripts homologous to the hpRNA (Helliwell and Waterhouse 2005; Voinnet 2002). If the hpRNAs are designed specifically against the parasite gene sequences, the transgene should have no deleterious effect on the host. However, when the hpRNA is translocated across the haustorium from the host into the parasite, the targeted parasite gene will be silenced, resulting in death of the invading parasite if the silenced gene is critical for parasite development.

The idea of using viral-specific oligonucleotides to control viral pests was proposed by Sanford and Johnston (1985) a decade before RNA interference (RNAi) mechanisms were discovered (Baulcombe 1996). RNAi-based resistance expressed in a host has since been demonstrated against several classes of pests including root-knot nematodes (Huang et al. 2006), western corn rootworm (Baum et al. 2007), and fungal pathogens (Tinoco et al. 2010; Yin et al. 2011). Silencing genes in a parasitic plant using RNAi from a host was visualized by silencing a β-glucoronidase (GUS) reporter gene in transgenic roots of the hemiparasite Triphysaria versicolor (Tomilov et al. 2008). When GUS-expressing parasite roots were infected onto lettuce or Arabidopsis roots bearing a hpGUS construction, the GUS staining and GUS transcript levels were significantly reduced near the site of the haustorial invasion. Inhibition of GUS activity in a host plant was also detected when the parasite was used as a bridge to a root bearing hpGUS, indicating that the silencing signals can move from host to parasite as well as parasite to host.

Although the reporter-silencing experiments showed that gene expression in parasites can be silenced by hpRNAi constructs in the host, the application of this strategy to control parasitic weeds is another matter. Both successes and failures have been reported. One study reported a reduction in Orobanche viability after infection onto tomato plants bearing an hpRNAi targeted against the mannose 6-phosphate reductase (M6PR) gene. The M6PR gene encodes an enzyme required for the biosynthesis of mannitol, a sugar alcohol proposed to regulate the osmotic drive across the haustorium. Inhibition of GUS activity in a host plant was also detected when the parasite was used as a bridge to a root bearing hpGUS, indicating that the silencing signals can move from host to parasite as well as parasite to host.
Old and New Herbicide Targets. Herbicides have been successfully utilized to control the growth of weedy parasitic species, although unique challenges must be addressed. Root parasites like *Orobanche* and *Siriga* must be controlled at an early stage to avoid crop yield losses, but the parasite is underground and not easily reached by chemical or nonchemical means. An effective control strategy is to use herbicides that translocate systemically from the crops to the parasites after attachment on the hosts, but this requires a herbicide that is both readily translocated and highly selective for the parasite. Transgenic crops with target site resistance primarily for acetolactate synthase (ALS) or 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)-inhibiting herbicides have been used in this way and show great promise (reviewed by Gressel 2009). An innovation related to this strategy is the practice of treating crop seeds with herbicide solutions where the concentration used is not toxic to the crops but is lethal to the newly attached parasite (Ransom et al. 2012).

The nutritional dependence of parasitic species on their hosts raises questions about parasite ability to survive on essential metabolites derived from the host vs. those synthesized using the parasite's own metabolic machinery. To address the question of whether parasitic species contain herbicide target genes, we searched the PPGP database to identify putative orthologs of herbicide target proteins from *Arabidopsis*. Table 2 shows that nearly all characterized herbicide target sites are present in parasitic species, with the exception that *O. aegyptiaca* appears to lack the psbA gene that encodes the D1 protein involved in photosynthetic electron transport. The loss of photosynthetic gene expression is described in detail by Wickett et al. (2011) and indicates the lack of effectiveness of photosynthesis-inhibiting herbicides on *O. aegyptiaca*, but the presence of low levels of gene expression in the chlorophyll biosynthesis pathway surprisingly suggests that herbicides inhibiting protoporphyrinogen oxidase may have an effect on these weeds. We also found relatively low coverage of *O. aegyptiaca* genes for phytoene desaturase (52%) and 4-hydroxyphenylpyruvate dioxygenase (23%), which are targeted by bleaching herbicides. This lower coverage is consistent with loss of photosynthesis in *O. aegyptiaca* in that it may reflect reduced levels of carotenoid synthesis or greater sequence divergence in these genes.

Aside from the genes related to photosynthesis, we obtained excellent coverage of genes for other metabolic processes targeted by herbicides (Table 2). There was no obvious trend toward expansion or contraction of gene families among the species with different levels of parasitism and overall numbers of unigenes were consistent with other weed-sequencing projects (Peng et al. 2010; Riggins et al. 2010). More detailed investigation into evolution of herbicide target genes is needed, but it appears that parasitic species of the Orobanchaceae have retained these genes and likely require them for successful growth even given opportunities to acquire resources from hosts during much of their life cycle. Given the prediction that parasites will likely evolve resistance to herbicides (Gressel et al. 1996), the PPGP sequences provide researchers with baseline information needed to identify and track target site resistance in putatively resistant populations.

Knowledge of parasite gene sequences and expression patterns will also be useful in identifying new targets for parasitic weed control and for understanding the effects of herbicides on the parasites. Just as some conventional targets such as photosynthetic inhibitors may not be applicable for parasitic plants, other processes unique to parasites may be targeted by herbicides. The use of species-specific transcriptome expression patterns, as described above, is just one way to identify essential and unique parasite processes that could lead to development of new mechanisms of herbicide selectivity between crops and parasitic weeds. Furthermore, it is now possible to simultaneously study both crop host and parasite gene expression and determine not only the primary effects of herbicides, but other secondary effects they might have on the parasite–host interaction. Since plant gene expression data are already available for many ALS, EPSPS, and 2,4-D herbicides from *Arabidopsis*, comparing these profiles with those from *Orobanche* or *Siriga* global gene expression data would help to pinpoint the major molecular players that determine the specificities (Das et al. 2010; Manabe et al. 2007; Raghavan et al. 2005).

**Conclusions**

The transcriptome sequencing of three parasitic species of the Orobanchaceae is a significant advance in understanding and ultimately controlling parasitic weeds. The sequences represent parasite gene expression at all life stages, and have been coordinated to maximize ability to compare expression profiles between the species and thereby infer which genes are unique to the evolution of parasitism. Although analysis of the data is still in the early stages, we have already begun to see evidence for gene loss and examples of surprising gene retention associated with the evolution of holoparasitism (Table 2, Wickett et al. 2011).

Despite the great advance represented by the PPGP database, many questions remain to be explored. It will be important to profile gene expression over more fine-scale time courses to understand how expression changes during important developmental transitions such as haustorium initiation and integration into host tissues. Also, more precise dissection of tissues, such as that provided by laser pressure catapult microdissection, will emphasize those genes expressed only in the haustorium cells that are in contact with the host. This is the most likely place to find parasite effector molecules and key parasite transporter genes that are essential for host colonization and acquisition of resources.

Gene expression alone will not reveal the complex workings of the parasite–host interaction. Transcriptomics, proteomics, and metabolomics must be integrated to create a unified picture of parasitism. In addition, full genome sequences of parasites will be needed to maximize interpretation of the transcriptome data and to gain access to regulatory sequences that are most likely to account for the dramatic morphological variation among the parasitic species. We fully expect that advances in sequencing and analytic technology will allow for this type of analysis in the near future, which will further accelerate the pace of discovery. Although parasitic weeds have been a largely intractable problem for decades, current advances in genomics will enable a revolution in the types of research that are possible and the pace at which it proceeds. The ability to effortlessly obtain parasite gene sequences, clone them, and characterize them will accelerate the process of determining how parasites live and how best to control them.
Acknowledgments

This work is supported by National Science Foundation (NSF) Plant Genome award DBI-0701748 to J.H.W., C.W.D., M.P.T., and J.I.Y. Additional support is acknowledged to J.H.W. from the U.S. Department of Agriculture (Hatch project no. 135798), to M.P.T. from the NSF (IBN-0322420) and Kirkhouse Charitable Trust, and to J.I.Y. from NSF (IBN-0236545).

Literature Cited


Wallingford et al.: Parasitic Plant Genome Project


Received July 1, 2011, and approved August 28, 2011.