Article

Current Biology

Genome Sequence of *Striga asiatica* Provides Insight into the Evolution of Plant Parasitism

Highlights

- The *Striga* genome reflects a three-phase model of parasitic plant genome evolution
- A family of strigolactone receptors has undergone a striking expansion in *Striga*
- Genes in lateral root development are coordinately induced in a parasitic organ
- Host genes and retrotransposons are horizontally transferred into *Striga*

Authors

Satoko Yoshida, Seungill Kim, Eric K. Wafula, ..., Claude W. dePamphilis, Doil Choi, Ken Shirasu

Correspondence

ken.shirasu@riken.jp

In Brief

Yoshida et al. report the *Striga* genome sequence, providing insights into parasitic plant genome evolution and a key resource for the future development of *Striga* control strategies. The genome also shows evidence for the horizontal transfer of host genes and retrotransposons, indicating gene flow to the parasite from hosts.

Yoshida et al., 2019, Current Biology *29*, 3041–3052
September 23, 2019
https://doi.org/10.1016/j.cub.2019.07.086

CellPress

Current Biology
Article

Genome Sequence of *Striga asiatica* Provides Insight into the Evolution of Plant Parasitism

Satoko Yoshida,[1,2,3] Seungill Kim,[4,5] Eric K. Wafula,[6] Jaakko Tanskanen,[7,8] Yong-Min Kim,[9] Loren Honaas,[6,32] Zhenzhen Yang,[6] Thomas Spallek,[1,10] Caitlin E. Conn,[11] Yasunori Ichihashi,[1,12] Kyeongchae Cheong,[4,5] Songkui Cui,[1,3] Joshua P. Der,[13] Heidrun Gundlach,[14] Yuannian Jiao,[33] Chiaki Hori,[1,15] Juliane K. Ishida,[1] Hiroyuki Kasahara,[1,16] Takatoshi Kiba,[1,17] Myung-Shin Kim,[4,5] Namjin Koo,[9] Anuphon Laohavisit,[1] Yong-Hwan Lee,[4,18]

(Author list continued on next page)

[1]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan
[2]Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
[3]Institute for Research Initiatives, Division for Research Strategy, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
[4]Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul 151-742, Korea
[5]Plant Genomics and Breeding Institute, Seoul National University, Seoul 151-742, Korea
[6]Department of Biology, Pennsylvania State University, University Park, PA 16802, USA
[7]Production Systems, Luke Natural Resources Institute Finland, 00790 Helsinki, Finland
[8]Luke/BI Plant Genomics Laboratory, Institute of Biotechnology and Viikki Plant Science Centre, University of Helsinki, 00014 Helsinki, Finland
[9]Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejon 305-806, Korea
[10]Institute of Plant Physiology and Biochemistry, University of Hohenheim, 70599 Stuttgart, Germany
[11]Department of Genetics, University of Georgia, Athens, GA 30602, USA
[12]RIKEN BioResource Research Center, Tsukuba, Ibaraki 305-0074, Japan
[13]Department of Biological Science, California State University, Fullerton, Fullerton, CA 92831, USA
[14]Plant Genome and Systems Biology (PGSB), Helmholtz Center Munich, Neuherberg 85764, Germany
[15]Research Faculty of Engineering, Hokkaido University, Sapporo 060-8628, Japan
[16]Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Fuchu 183-8509, Japan
[17]Department of Applied Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan
[18]Department of Agricultural Biotechnology, Seoul National University, Seoul, Korea
[19]Department of Cell & Systems Biology, University of Toronto, Toronto, ON M5S-3B2, Canada
[20]Joint BioEnergy Institute, Emeryville, CA 94608, USA
[21]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[22]Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub, 00100 Nairobi, Kenya

(Affiliations continued on next page)

SUMMARY

Parasitic plants in the genus *Striga*, commonly known as witchweeds, cause major crop losses in sub-Saharan Africa and pose a threat to agriculture worldwide. An understanding of *Striga* parasite biology, which could lead to agricultural solutions, has been hampered by the lack of genome information. Here, we report the draft genome sequence of *Striga asiatica* with 34,577 predicted protein-coding genes, which reflects gene family contractions and expansions that are consistent with a three-phase model of parasitic plant genome evolution. *Striga* seeds germinate in response to host-derived strigolactones (SLs) and then develop a specialized penetration structure, the haustorium, to invade the host root. A family of SL receptors has undergone a striking expansion, suggesting a molecular basis for the evolution of broad host range among *Striga* spp. We found that genes involved in lateral root development in non-parasitic model species are coordinately induced during haustorium development in *Striga*, suggesting a pathway that was partly co-opted during the evolution of the haustorium. In addition, we found evidence for horizontal transfer of host genes as well as retrotransposons, indicating gene flow to *S. asiatica* from hosts. Our results provide valuable insights into the evolution of parasitism and a key resource for the future development of *Striga* control strategies.

INTRODUCTION

Striga is a genus of parasitic plants in the Orobanchaceae family that includes major agricultural weeds. *S. asiatica* and *S. hermonthica* infect grain crops such as sorghum, millet, maize, upland rice, and sugarcane, causing $US billions of annual yield losses [1–3]. *Striga* has evolved unique parasitic adaptations that make infestations extremely difficult to eradicate [3]. A single *Striga* plant produces more than 100,000 small (~200 μm) seeds, which can be wind dispersed for a long distance. The seeds can lie dormant for decades, surviving extreme conditions, until they perceive host-derived germination stimulants, such as strigolactones (SLs) [4, 5]. Once germinated,

Shelley Lumba,[19] Peter McCourt,[19] Jenny C. Mortimer,[1,20,21] J. Musembi Mutuku,[1,22] Takahito Nomura,[23] Yuko Sasaki-Sekimoto,[24] Yoshiya Seto,[25,26] Yu Wang,[27] Takanori Wakatake,[1,28] Hitoshi Sakakibara,[1,17] Taku Demura,[1,2] Shinjiro Yamaguchi,[25,29] Koichi Yoneyama,[23] Ri-ichiroh Manabe,[30] David C. Nelson,[11,31] Alan H. Schulman,[7,8] Michael P. Timko,[27] Claude W. dePamphilis,[6] Doil Choi,[4,5] and Ken Shirasu[1,28,34,*]

[23]Center for Bioscience Research and Education, Utsunomiya University, Utsunomiya 321-8505, Japan
[24]School of Life Science and Technology, Tokyo Institute of Technology, 226-8501, Yokohama, Kanagawa, Japan
[25]Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi 980-8577, Japan
[26]Department of Agricultural Chemistry, School of Agriculture, Meiji University, Kawasaki, Kanagawa 214-8571, Japan
[27]Department of Biology, University of Virginia, Charlottesville, VA 22903, USA
[28]Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan
[29]Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan
[30]Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Kanagawa 230-0045, Japan
[31]Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92521, USA
[32]U.S.D.A. ARS, Wenatchee, WA, USA
[33]Institute of Botany, The Chinese Academy of Sciences, Nanxincun, Xiangshan, Beijing, China
[34]Lead Contact
*Correspondence: ken.shirasu@riken.jp
https://doi.org/10.1016/j.cub.2019.07.086

Striga roots grow toward the host and detect compounds derived from the host cell wall, which induce the development of a specialized organ called the haustorium at the tip of the radicle [6, 7]. The haustorium invades the host root and connects its xylem with that of the host to assimilate water and nutrients. In addition, genetic materials from the hosts are also transferred into *Striga*, but the extent and the precise mechanism of horizontal gene transfer (HGT) remain elusive [8–10].

RESULTS AND DISCUSSION

The Structure and Evolution of the *Striga* Genome

The genome of the *S. asiatica* strain that invaded the United States in the 1950s [2] was sequenced and assembled using a combination of Illumina-based whole-genome shotgun technology and Sanger-based BAC library end sequencing. The Kmer-based estimation of the *S. asiatica* genome size is approximately 600 megabase pairs (Mb), and 472 Mb of the genome was assembled with an N50 scaffold size >1.3 Mbp (contig N50 > 16.2 kbp and 393 × read coverage; Data S1A), in which a total of 34,577 genes was predicted (for detail, see Data S2A and S2B).

Global gene family phylogenetic analysis and genome structure and synteny analysis with the closely related nonparasitic plant *Mimulus (Erythranthe) guttatus* (Figure 1) both indicate that the *S. asiatica* genome retains evidence of at least two whole-genome duplication (WGD) events (Figures 2A–2D; Data S2C). We examined the divergence patterns of synonymous substitution rates (K_s) for Lamiales-wide duplicate genes identified by an integrated syntenic and phylogenomic analysis. Comparison of gene trees for 1,440 orthologous single-copy genes showed that the length for the branch leading to *S. asiatica* was longer than that leading to *Mimulus* suggesting that *S. asiatica* has experienced a more rapid molecular evolution than *Mimulus* (Figure 1). We identified two significant duplication components in *S. asiatica* at mean $K_s \approx 0.47$ (younger) and 1.22 (older) as well as one significant component for *Mimulus* at mean $K_s \approx 0.94$ (Figure 2B). The older *Striga* K_s peak and the single peak of the *Mimulus* K_s distribution represent a shared ancestral WGD event for Lamiales (Figure 2C). As expected the *S. asiatica* peak is shifted to the right (a higher K_s value) because of the accelerated rate of evolution for *S. asiatica*. The prominent younger peak in the *Striga* K_s distributions represents a duplication event that occurred after the divergence of lineages leading to *S. asiatica* and *Mimulus.*

Parasitic plant evolution is thought to progress through three phases: phase I, evolutionary gain of a haustorium; phase II, loss of functions that are supplemented by a host resource; and phase III, specialization of the parasitic relationship [11, 12] (Figure 2D). Shifts of gene expression (in scope and/or specificity) and changes in the global functional gene profile presumably accompany innovation during parasite evolution. Thus, we examined shifts of parasite gene expression and function by genome-scale comparative analyses to identify the signatures of each phase. Using the list of *S. hermonthica* "haustorium" orthogroups defined in Yang et al. [13], with a parallel analysis that identifies genes with tissue-specific expression in *Arabidopsis*, we found that haustorial genes are significantly enriched for tissue-specific orthogroups in *S. asiatica* (Data S1B). Concordant with Yang et al. [13], this pattern was strongest for pollen orthogroups. This suggests that haustorium innovation during phase I may have involved co-option of genes with tissue-specific gene expression.

Next, we identified functions associated with shifts in gene content by reconstructing each orthogroup (approximate gene family) in a common ancestor of *Striga* and *Mimulus*, as well as successively earlier common ancestors (Data S1C and S2C). Among the 10,248 orthogroups, approximately ~23% showed changes in gene numbers inferred for the *Striga* lineage (647 contractions, 1,742 expansions, 456 losses, and 152 gains; Data S1D, S1E, and S3). The relative age of genes in contracted orthogroups was significantly older (two-tailed Mann-Whitney U test, $p < 2.2e\text{-}16$) than genes in expanded families (Figure 2D; Data S1E). In addition, the expanded gene families show higher non-synonymous/synonymous substation (Kn/Ks) ratios compared to the contracted gene families (Student's t test, $p < 4.7e{-}10$; Figure S1), suggesting that the expanded gene families are under more relaxed selection pressure. The relatively younger expanded gene families, apparently gained largely as a result of the *Striga* WGD (Figures 3B and 3E), potentially

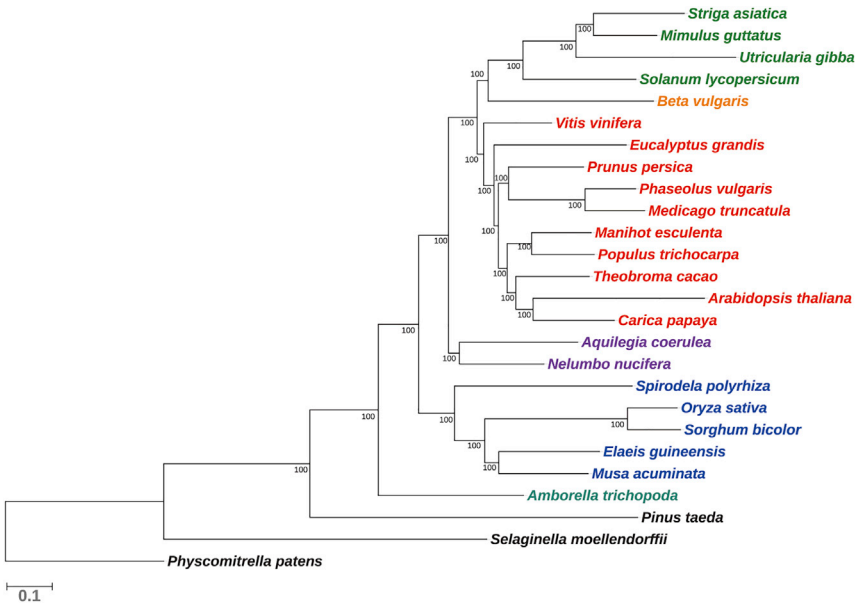


Figure 1. The Maximum Likelihood Species Tree

Phylogenetic tree of 26 representative plant species (Data S1C) was estimated from the concatenated data matrix for 1,440 single-copy orthogroup genes obtained from the BUSCO classification. Bootstrap values were 100% for each node. The scale indicates number of substitution per site.

provided a source of genes to encode specialized traits in the parasite.

Significant (Benjamini corrected $p < 0.05$) signatures of gene family contractions were detected in two photosynthesis-related KEGG pathways (Data S1F and S1G). Additionally, an analysis of Gene Ontology (GO) terms among contracted lineages showed several photosynthesis-related cellular compartment (CC) terms and biological process (BP) terms were significantly over-represented (Data S1H and S2C; Figure 2D). These contractions are consistent with *Striga*'s high reliance on host carbon [14, 15]. Furthermore, significantly enriched GO BP terms associated with leaf anatomy and function were detected among contracted lineages, consistent with the anatomical and functional reductions in *Striga* leaves. In addition to the well-documented gene losses in parasitic plant plastomes [12, 16], these changes indicate a complementary reduction in reliance on photosynthesis-related gene function [17] representing phase II.

Perhaps the clearest support for Searcy's phase II are substantial contractions in gene families annotated with GO BP terms that relate to abiotic and biotic stimulus response including virtually all plant hormones (Data S2C, S1H, and S1I; Figure 2D). This includes one in four significant GO BP terms that are seven times more numerous in contracted lineages than expanded ones. This pattern of loss points to an increasingly insensitive parasite sensing apparatus that is likely supplemented by the host. Concordant with this evolutionary signature, empirical evidence suggests that *Striga* lost abscisic acid sensitivity to regulate water loss machinery and maintains constitutively open stomata even under drought conditions [18, 19] contributing to a net carbon loss in the host leaves [20].

The transition from phase II to phase III may in some cases be blurred from a functional standpoint because, for instance, the host plant could complement water stress response pathways, while decreasing water potential in the parasite could be adaptive [9]. Indeed, significantly enriched water relations terms can be found among both expanded and contracted lineages, yet orthogroup contractions dominate water relation signatures indicating that altered water relations may largely, but not exclusively, represent older phase II losses. In GO CC profiles, contractions are biased toward structural and photosynthesis related genes families—consistent with phase II complementation. However, the newer and expanded gene families are significantly biased toward endocytosis and intracellular transport, suggesting that phase III innovations contribute to host resource acquisition processes. The expansions in cellular transport machinery may help explain how *Striga* obtains photosynthate-derived host resources even though direct phloem connections are lacking [15, 20].

Host Recognition—Evolution of SL Receptors

As an obligate pathogen, *Striga* requires nutrients from a host within a few days after germination. One unique aspect of the specialized relationship with the host (phase III) in the *Striga* parasitic lifestyle is the ability to germinate after sensing SLs, which indicate presence of a host [5]. In *Arabidopsis*, *D14* and *KAI2*/*D14L* are ancient paralogs that encode receptors for SLs and the karrikins (smoke-derived compounds that stimulate germination of many nonparasitic plants), respectively [21, 22]. *KAI2*, which controls seed germination in *Arabidopsis*, has undergone higher than normal gene duplication in several parasite genomes in the Orobanchaceae [23–25]. A divergent subclade of *KAI2* paralogs (*KAI2d*) has evolved SL perception, which facilitates host detection in seeds. The super-orthogroup that contains the *KAI2* genes was expanded strikingly in *S. asiatica* (Data S1J and S2D). We found that the *S. asiatica* genome encodes 21 *KAI2* paralogs and that 17 of these are in the *KAI2d* class (Figure 3A). Most of the *KAI2d* genes in *S. asiatica* are highly expressed in the seed as well as in seedling stages (Figure 3B). Two other paralogs, *KAI2c1* and *KAI2c2*, cluster with highly conserved *Arabidopsis* (*AtKAI2*) and *Mimulus* proteins (*MgKAI2c*). The intermediate group contains two *KAI2i* paralogs, which are sister to the expanded *KAI2d* clade. *Mimulus KAI2i* (*MgKAI2i*) is branched from the ancestral node of the *Striga KAI2d* and *KAI2i*, suggesting that *Striga KAI2d* genes evolved out of the intermediate group. In addition, seven *KAI2* pseudogenes are also found in the genome, providing further evidence for highly dynamic evolution of the *KAI2* gene family (Figure 3C). *KAI2* paralogs and pseudogenes are often found on the same scaffold (Figures 3C and 3D). All *KAI2* genes retain a single intron

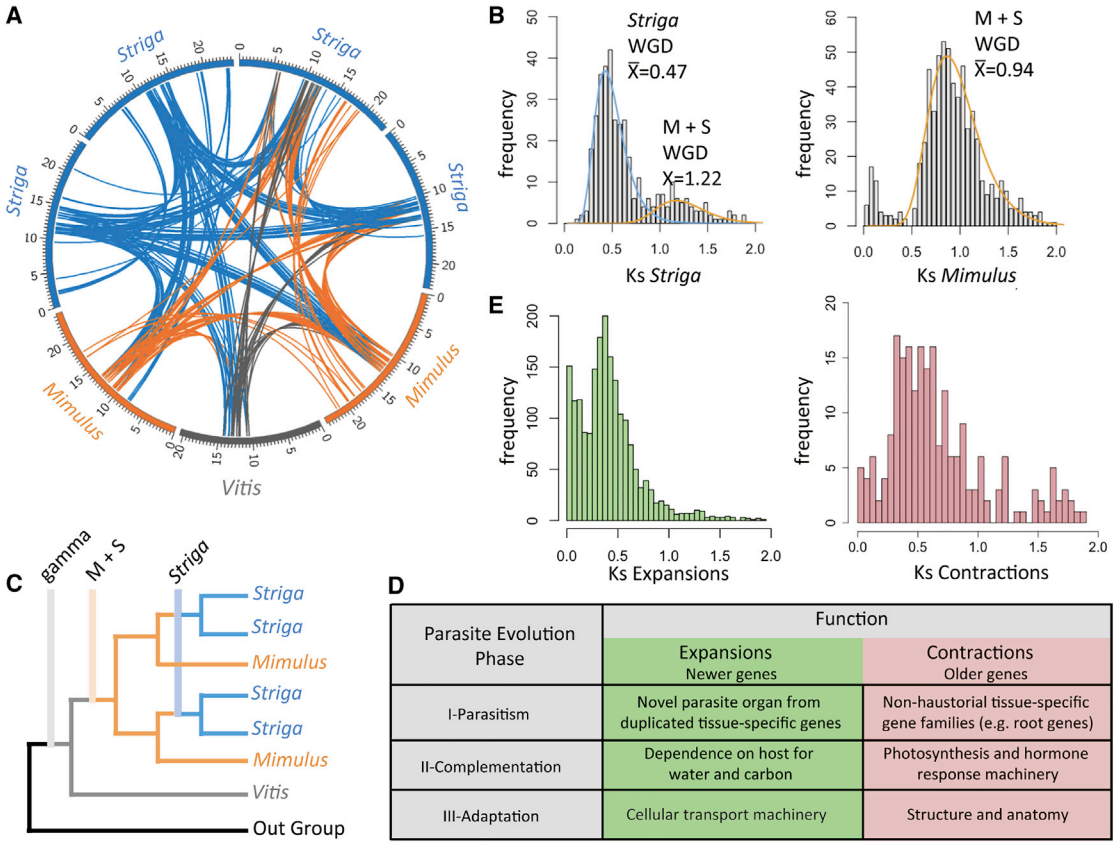


Figure 2. The *Striga asiatica* Genome

(A) Syntenic scaffolds of *Striga* (blue), *Mimulus* (orange), and *Vitis* (gray).

(B) Ks plots of *Striga* and *Mimulus* duplicate genes. Orange and blue colors represent an older and a recent polyploidy event, respectively.

(C) Schematic phylogenetic tree presenting whole-genome duplication events that occurred during the evolution of the lineage leading to *Striga*. Gamma is the genome triplication shared by core eudicots, *Striga* and *Mimulus* share a WGD (M+S), and *Striga* has experienced an independent WGD.

(D) Three-phase model of parasite evolution, showing gene categories with expression shifts, expanded and contracted orthogroups in the *Striga* genome relative to a reconstructed ancestor of *Striga* and *Mimulus*. See Data S2 for details.

(E) Ks plots of expanded and contracted *Striga* genes. Age of contracted genes categorizes significantly older than expanded genes categories. See also Figure S1.

at a conserved position. Tandem *KAI2* paralogs typically share the same orientation, consistent with localized *KAI2* duplication by unequal recombination. Interestingly, *KAI2i,* which is ancestral to *KAI2d* genes, is located next to *Striga*-specific *KAI2d7* and *KAI2d8* (Scaffold 62; Figures 3C and 3D), suggesting that the *Striga*-specific *KAI2d* clade originally may have been derived by the tandem duplication of *KAI2i*. If different *KAI2d* paralogs have specificity for distinct types of SLs, then the rapid evolution of the *KAI2d* clade likely enabled *Striga* seeds to recognize a wide range of hosts [23–25]. We noted that the high level of expression of many *KAI2d* homologs have a high level of expression at the seedling stage, suggesting that the host-derived SL may influence other functions beyond germination.

Development of the Invading Organ, the Haustorium

Immediately after germination, *Striga* grows toward the host and detects cell wall-derived compounds [6]. This initiates a drastic developmental reprogramming, resulting in the formation of a haustorium that invades the host root (Figure 4A). To investigate gene expression dynamics during haustorium development, RNA sequencing (RNA-seq) analysis was performed with the most devastating *Striga* species, *S. hermonthica* (Data S1K–S1M and S2E). Principal component analysis (PCA) and self-organizing map (SOM) clustering were used to classify the transcripts into twelve clusters, each with a distinct expression pattern specific to one or more developmental stages (Figures 4A and 4B). The GO enrichment analysis of these clusters (Benjamini and Hochberg corrected $p < 0.05$; Figure 3C; Data S1N) projected a similar sequence of molecular events during *Striga* parasitism. Clusters 2, 3, and 6 showed expression patterns specific to the seed; transcripts in these clusters are enriched for GO terms related to post-embryonic development and to embryonic development toward the end of seed dormancy (Benjamini and Hochberg corrected $p < 0.05$; Figure 4C). The seedling-specific cluster 12 showed enrichment in defense responses as well as in transcriptional regulatory activity (Benjamini and Hochberg corrected $p < 0.05$; Figure 4C). This suggests that the seedling has already started to change its transcriptional profile to enable parasitization of host plants; i.e., the primary haustorium formation may be coupled with seed germination in *S. hermonthica*.

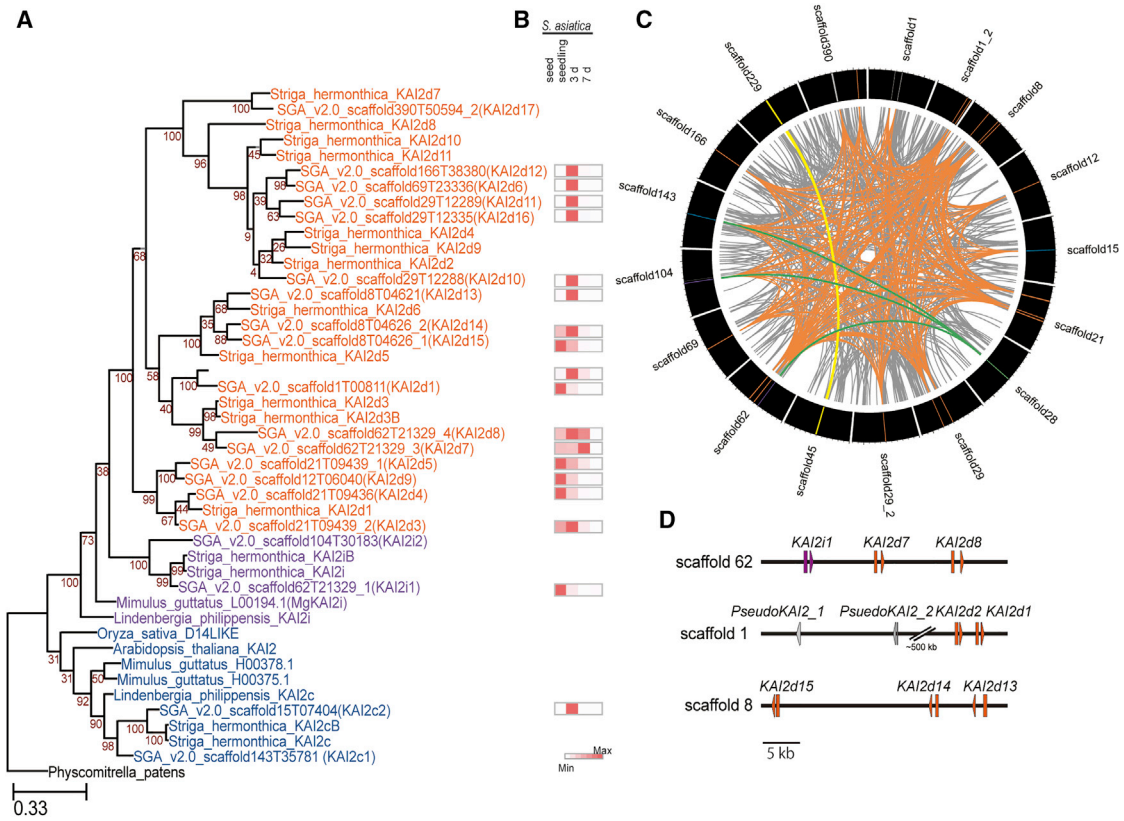


Figure 3. The Evolution of Strigolactone (SL) Receptor Genes in *S. asiatica*

(A) Maximum-likelihood phylogeny of predicted amino acid sequences of *KAI2*/*D14-LIKE* homologs in *S. asiatica* and *S. hermonthica* together with other non-parasitic species. The tree was generated based on the JTT-matrix-based model. Bootstrap values are shown at the bases of branches. The scale shows inferred number of evolutionary changes per amino acid. Conserved, intermediate, and divergent clades are shown in blue, purple, and orange, respectively.

(B) Scaled expression levels of *S. asiatica KAI2* genes at indicated stages.

(C) Local similarities detected between the genomic regions containing *KAI2*/*D14-LIKE* (blue for *KAI2c*, purple for *KAI2i*, and orange for *KAI2d*), *D14* (green), *DLK2* (yellow) homologs, and/or their pseudogenes (gray). Locally aligned genomic regions among scaffolds (blastZ score >15,000) are connected with solid lines. Orange and yellow lines represent regions containing *KAI2* or psuedo-*KAI2* and *DLK2* homologs, respectively. Gray lines connect locally similar regions outside *KAI2*/*D14*/*DLK2* genes. Nucleotide numbers in the scaffold are written beside the scaffold.

(D) Schematic representation of tandemly duplicated *KAI2* homologs in the genome. See Data S2 for details.

Our SOM analysis allowed us to capture a subsequent peak of gene expression from seedling to 7 days, represented by clusters 9, 1, 5, 4, 8, 7, 11, 10, in that order (Figure 4C). The temporal expression patterns of several selected genes were confirmed by qRT-PCR upon host and nonhost interactions (Figure S2; Data S2E). While the early gene expression was induced by DMBQ treatments as well as host and nonhost interactions, the expression of middle- and late-stage genes was not seen in the interaction with nonhost *Lotus japonicus* (Figure 4D; Data S2E). Because *S. hermonthica* is able to penetrate tissues of nonhost *Arabidopsis* and *L. japonicus*, but not establish xylem connections with *L. japonicus* [27], the early genes are likely to be important for haustorium formation and host penetration, while the genes involved in the middle to late stages of haustorial development may associate with xylem connection formation and/or host materials acquisition. *In situ* hybridization analysis highlights the tissue-specific expression of such genes. An early-stage gene, encoding the peroxidase, is exclusively expressed at the intrusive cells that are aligned at host-parasite interface (Figures 4E and 4F), whereas various 7-day-specific genes are highly expressed in the hyaline body (Figures 4G–4J), a specific parenchymatic tissue whose characteristics include dense cytoplasm, organelle-rich structure, and high metabolic activity [28]. The hyaline body is proposed to function as a sink for host materials, and the high expression of catabolic enzymes such as proteases within this tissue may contribute to such a function. The middle and late genes include the recruitment of catalytic activity-related genes (especially hydrolases) during host penetration, transport-related genes during host nutrient acquisition, and signal transduction-related genes during resource allocation. In fact, among the identified 1,292 CAZyme (carbohydrate-active enzyme)-categorized genes [29], 252 are differentially expressed during invasion stages (Figure 5; Data S1O, S1P, and S2E). Specifically, enzymes targeting primary cell wall components, such as those degrading pectin, are highly upregulated (Figures 5C and 5D). In addition, many proteases are upregulated at late stages of infection.

Comparative studies of development in an evolutionary context have been routinely employed to understand developmental mechanisms and to deduce how the regulatory changes

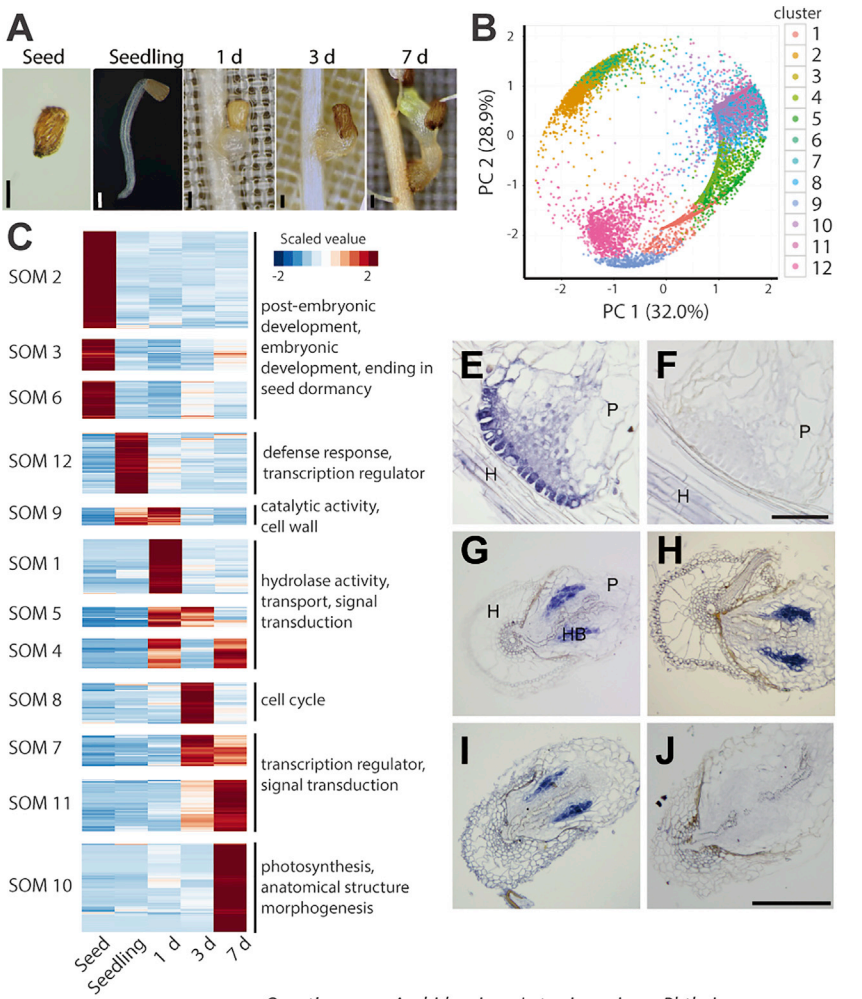


Figure 4. Transcriptional Reprogramming in Haustorium Development

(A) Developmental stages used for the transcriptome analysis of *S. hermonthica*. Seeds, preconditioned seeds; seedlings, 48 h after 10 nM strigol [26] treatment; 1 d, whole *S. hermonthica* seedlings 1 day after rice infection; 3 d and 7 d, *S. hermonthica* haustoria attached to rice tissues at 3 and 7 days after rice infection. Scale bar, 100 μm.

(B) The expression profile of each transcript is represented in PCA space with SOM node memberships indicated by different colors. A total of twelve clusters showing expression patterns specific to one or more stages were defined. The percentage shown along the x or y axis represents the percentage of variance explained by each component.

(C) Heatmap of normalized gene expression of each transcript separated by SOM clustering with selected enriched GO terms ($p < 0.05$).

(D) Expression heatmap of stage-specific *S. hermonthica* genes in interaction with host (*O. sativa*) and nonhosts (*Arabidopsis*, *Lotus japonicus*, and *Phtheirospermum*) interactions.

(E–J) *In situ* hybridization on haustorial sections of *S. hermonthica* at 1 day (E and F) and 7 days (G–J) after rice infection. The hybridized signal (blue) represents the localization of the transcript of an early-expressing gene encoding peroxidase (E) and late-expressing genes encoding subtilase 1 (G), LRR kinase (H), or cytokinin oxidase/dehydrogenase (I). The sense probe of peroxidase (F) and *subtilase1* (J) was used as a negative control. H, host plant; P, parasite. Scale bar, 200 μm. See also Figure S2.

in gene expression contribute to morphological diversity [30]. Since our genome analysis indicated potential sub-functionalization and/or co-option of existing genes from tissue-specific gene families (phase I), we hypothesized that parasitic plants may have employed a pre-existing developmental program to produce the haustorium. One such program is lateral root formation, as this also creates new xylem connections in roots. Out of the known 18 lateral root development (LRD) genes in *Arabidopsis* [31], we identified, respectively, 18 and 17 LRD orthologs in the *S. asiatica* genome and the *S. hermonthica* transcriptome (Data S1Q and S2E). Among these genes, *SLR(IAA14)*, *ARF19*, and *LAX3* orthologs are specifically expressed during the early stage of haustorium development (Figures 6A and S3). *SLR(IAA14)* and *ARF19* function as a module to regulate the expression of the auxin influx carrier *LAX3*, which localizes auxin accumulation during LRD [32] (Figure 6B). Thus, the *SLR(IAA14)-ARF19-LAX3* component might be utilized to initiate auxin accumulation during *Striga* haustoria formation. We also detected another putative target of the *SLR(IAA14)-ARF19* module, the *LBD18* ortholog, which is highly expressed in the early stage (Figure 6A). *Arabidopsis LBD18* activates cell proliferation in the lateral root primordia [33]. Correspondingly, cell proliferation is highly active in haustoria [34], suggesting that the *LBD18* ortholog might have a conserved function to coordinate the spatial pattern of cell proliferation during haustorium formation. In the later stages of haustoria formation, such as 3 days and 7 days, we observed the upregulation of *ARF5* and of *ARF8* homologs (Figure 6A). *ARF5* follows *SLR(IAA14)-ARF19* expression to control lateral root organogenesis [35], whereas *ARF8* activates lateral root meristem in response to nitrogen availability [36]. Therefore, these genes might be involved in the later stages of haustorium formation when host penetration occurs and vasculature connections are formed. Note that no upregulation of two other LRD-related genes, *ABERRANT LATERAL ROOT FORMATION 4* (*ALF4*) and *ARABIDOPSIS*

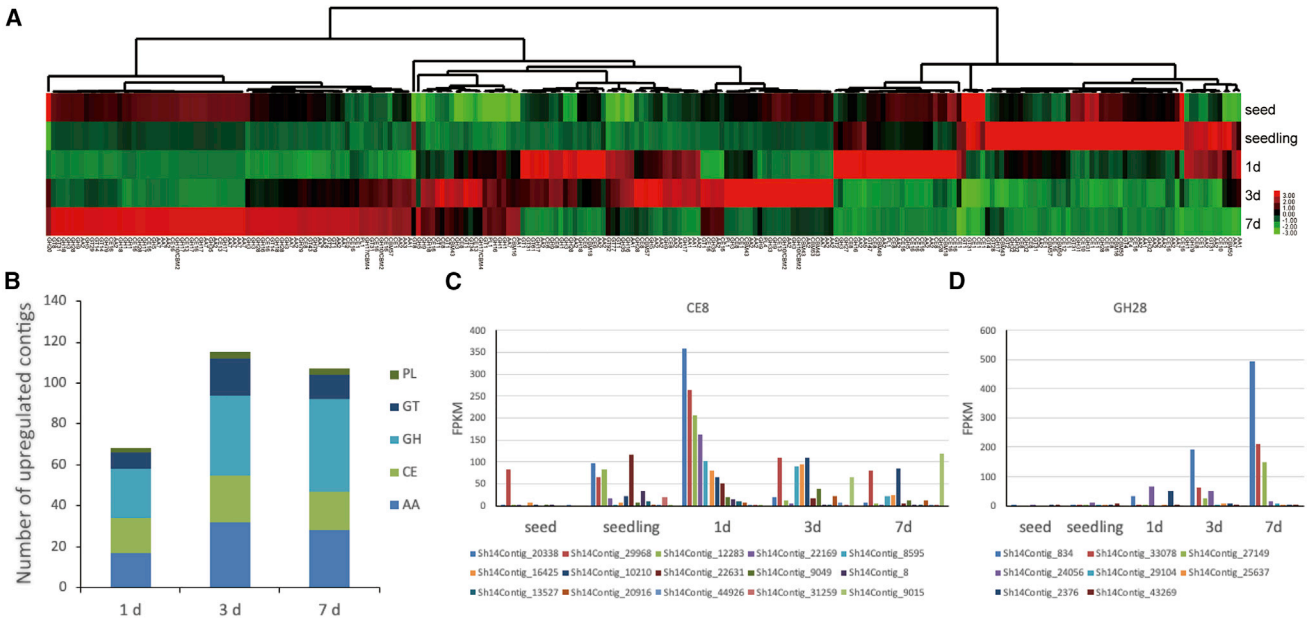


Figure 5. CAZyme Classification of the *S. hermonthica* Transcriptome

(A) Clustering and heatmap of the differentially expressed genes containing CAZyme motifs.

(B) Number of significantly upregulated contigs containing each class of CAZyme motifs. Contigs carrying AA and GH motifs are highly upregulated at 3 and 7 days after host interaction.

(C and D) Expression patterns of CE8 family containing pectin methyl esterases (C) and GH28 family containing polygalacturonases (D).

CRINKLY 4 (*ACR4*), were detected in *S. hermonthica* haustoria, but, surprisingly, their orthologs (*ALF4*: LOC_Os08 g19320; *ACR4*: LOC_Os03 g43670) were upregulated in host plants 1 day after infection (Figure 6A). As *ACR4* expression is dependent on *SLR(IAA14)-ARF19* to specify LRD cell identity in *Arabidopsis* [37] and *ALF4* functions in maintaining the mitotically competent state of the pericycle cells in LRD [38], *ACR4* and *ALF4* might link the interaction between *S. hermonthica* and its host. Taken together, certain LRD genes in *S. asiatica* and *S. hermonthica* are activated during haustorium formation, and, interestingly, the expression orders follow developmental time frames similar to those during LRD in *Arabidopsis*

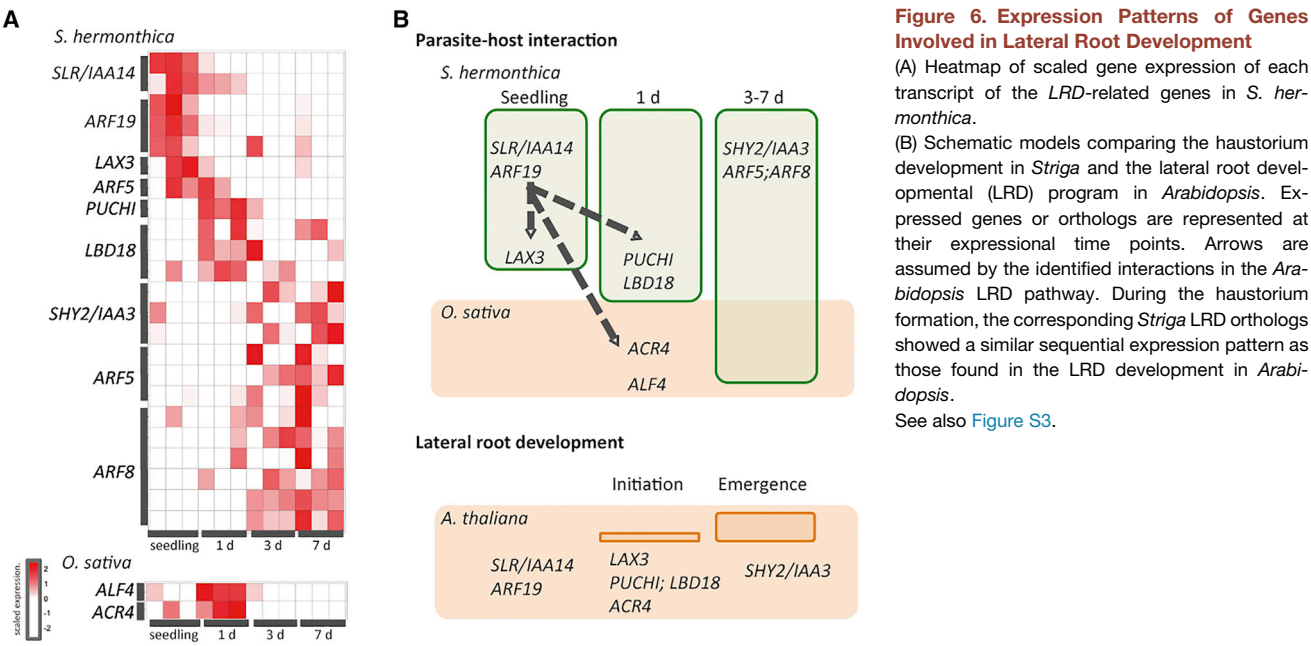
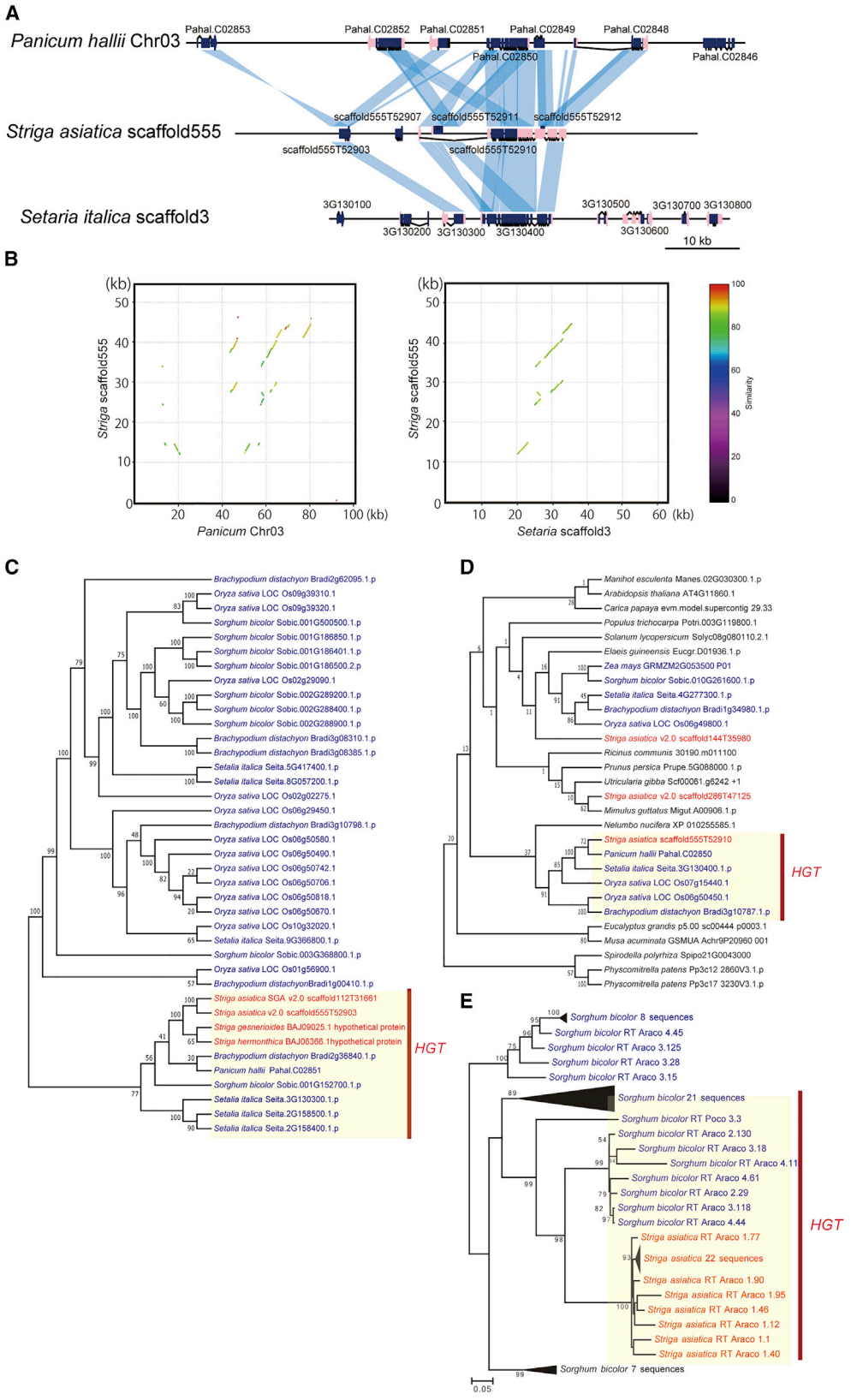


Figure 6. Expression Patterns of Genes Involved in Lateral Root Development

(A) Heatmap of scaled gene expression of each transcript of the *LRD*-related genes in *S. hermonthica*.

(B) Schematic models comparing the haustorium development in *Striga* and the lateral root developmental (LRD) program in *Arabidopsis*. Expressed genes or orthologs are represented at their expressional time points. Arrows are assumed by the identified interactions in the *Arabidopsis* LRD pathway. During the haustorium formation, the corresponding *Striga* LRD orthologs showed a similar sequential expression pattern as those found in the LRD development in *Arabidopsis*.

See also Figure S3.



(legend on next page)

(Figure 6B), suggesting that haustorium formation, which confers parasite function in parasitic plants, might be evolved partly through the recruitment of parasitic plant and host LRD programs.

Horizontal Gene Transfer

Genetic materials such as mRNAs are transferred from hosts to parasitic plants [39]. The transferred material may also be integrated into the germline of the parasites [8, 40]. To understand the extent of such HGT events, the *S. asiatica* genome was compared with other dicot and monocot genomes to find *Striga* genes that clustered with monocot orthologs. We identified 34 potential HGT candidates in the *S. asiatica* genome (Figure 7; Data S1R and S2F). Two of the HGT candidate genes are aligned in tandem in an approximately 30 kbp region in the genome of *S. asiatica.* The orthologs of the two genes, including introns and untranslated regions, are also located in tandem in the genomes of two Poaceae, *Panicum hallii* and *Setaria italica*, suggesting transfer of a large (~100 kb in *P. hallii*) genomic segment from host to parasite (Figures 7A and 7B). Phylogenetic analyses showed that the two *S. asiatica* genes clustered only with Poaceae genes, supporting HGT from host to parasite (Figures 7C and 7D). Interestingly, a few other genomic regions contain multiple HGT genes in close proximity (Data S2F), although the syntenic regions are not found in the Poaceae genomes, possibly due to rearrangement of the host genome after the gene transfer. These data suggest that the inter-species transfer of large genomic fragments may have occurred multiple times.

Because transposable elements were previously reported as HGT targets [10], we conducted phylogenetic analyses for all the reverse transcriptase (*rt*) domains in *S. asiatica* and for representative *rt* sequences from both eudicots and monocots (Data S2F). Our analyses included 35,690 from *Copia* and 54,973 from *Gypsy* elements in the publicly available plant genome sequences. Clusters containing both *S. asiatica rt* sequences and monocot sequences were analyzed further. Three putative HGT events were identified. One of these, comprising ~80 total *rt* sequences, includes 29 *S. asiatica rt* in a cluster with 48 diverse *Sorghum bicolor rt*, suggesting a direct horizontal transfer from *S. bicolor*, a natural host of *Striga* (Figure 7E), and subsequent amplification of *rt* sequences in the *Striga* genome. Two other trees, in which *S. asiatica rt* sequences are found nested within an exclusively Poaceae clade, having their closest orthologs, respectively, in *Oryza* and *Z. mays* or in *Oryza* and *S. bicolor*, suggest additional transfers from Poaceae hosts to *Striga* (Figure S4). These results indicate that *Striga* acquired genetic materials from its hosts with higher frequency compared to the autotrophic angiosperms, which may have influenced the parasite's evolution and adaption.

Outlook

Striga remains the greatest biological constraint to food production in its endemic areas in Africa, and thus its genomic and transcriptomic sequences are important tools for understanding its parasitic strategies and for developing efficient, knowledge-based management programs. In addition, the genome information provides a basis for understanding the origin of parasitism during the course of evolution. Similar to recently published stem parasites dodder (*Cuscuta* spp) genomes [40, 42], *Striga* evolved rapidly compared to autotrophic species, acquired genes from their hosts via HGT, and recruited root developmental programs for haustorial formation. Both parasites have lost genes related to environmental sensing, leaf developmental processes, and photosynthesis, as predicted for the degratory phase of parasite evolution, but *Striga* frequently retains portions of reduced gene families, reflecting its status as a leafy hemiparasite that is photosynthetically competent while being highly dependent on host-derived carbon. Detailed comparisons of nuclear genomes from fully heterotrophic Orobanchaceae and other parasitic plants with different levels of host dependency will deliver further insights into the evolution of parasitism.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Whole genome shotgun sequencing, assembly and annotation of S. asiatica
  - RNA sequencing
  - Genome comparative analysis
  - RT-qPCR
  - *In situ* hybridization
  - Identification of horizontally transferred genes and retrotransposons
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.07.086.

Figure 7. Horizontal Gene Transfers between Host and *Striga*

(A) Comparison of genomic regions between *P. hallii*, *S. asiatica*, and *S. italica*. The regions that show high similarity (LastZ score >5,000) are connected with sky-blue lines. Coding sequences are shown as dark-blue boxes, and untranslated regions are shown as pink boxes.

(B) A dot plot comparing an approx. 60 kb region in *S. asiatica* scaffold555 and either 100 kb region of *P. hallii* chromosome 3 (left) or 60 kb region of *S. italica* scaffold 3 (right) visualized by nucmer program in nummer [41] (default option). Similarity percentages are shown as rainbow color scale.

(C) Phylogenetic tree of a hypothetical protein (555T52903) that previously was found as horizontally transferred gene in *Striga hermonthica* ESTs [8].

(D) Phylogenetic tree of an Arginin-tRNA synthetase-like protein (555T52910).

(E) Phylogenetic trees of nucleotide sequences for reverse transcriptase in horizontally transferred retrotransposons from a host (*Sorghum*) to *S. asiatica*. The trees were unrooted and based on the maximum-likelihood method. Local support values are shown for branches. *Striga* genes are shown in red, and genes from grass species are shown in blue. HGT events are highlighted with yellow. See also Figure S4.

ACKNOWLEDGMENTS

This work is partially supported by MEXT KAKENHI (no. 24228008, 15H05959, 17H06172 to K.S.; no. 18H02464 and 18H04838 to S. Yoshida; no. 15K18589 to Y.I.; 17H06474 to S. Yamaguchi; 17H06473 to H.S.; and 25891029 and 17K15142 to S.C.), JSPS Postdoctoral Fellowship (to J.M.M. and C.H.), JSPS Research Fellowship for Young Scientist (to T.W.), the RIKEN Special Postdoctoral Researchers Program (to Y.I., J.C.M., and T.S.), NSF Postdoctoral Research Fellowship (no. 1711545 to C.C.), Academy of Finland Project 266430 (to A.H.S.), NSF IOS-1737153 and IOS-1740560 (to D.C.N.), NSF IOS-1238057 (to C.W.D. and M.P.T.), MEXT Cell Innovation program (to R.M. in CLST RIKEN), the International Atomic Energy Agency Research (no. 20645 to S. Yoshida and no. 20634 to J.M.M.), and NSERC (to S.L.). This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1A6A3A04004014) to S.K., by a grant from the Agricultural Genome Center of the Next Generation Biogreen 21 Program of RDA (project no. PJ013153) to D.C., and RIKEN president fund to K.S. HiSeq2000 sequencing was performed by Genome Network Analysis Support facility in RIKEN CLST. Computational analysis was partially performed using the super computer system of National Institute of Genetics, Research Organization of Information and Systems and with the Pouta platform of CSC - IT Centre for Science (Project hy7047), who are thanked for generous computational resources. Strigol and *S. hermonthica* seeds were kindly provided by Drs. K. Mori and A.G.E. Babikar, respectively. M.P.T. and Y.W. were supported in part by grants from the NSF (IOS-1238057 and IOS-1213059) and the Kirkhouse Trust. J.C.M.'s work was funded as part of the DOE Joint BioEnergy Institute supported by the U.S. Department of Energy (DE-AC02-05CH11231).

AUTHOR CONTRIBUTIONS

K.S. conceived the project, designed the content, and organized the manuscript. M.P.T. provided plant materials. S. Yoshida, T.S., and R.M. performed data generation and sequencing analysis. S.K., Y.-M.K., K.C., M.-S.K., Y.-H.L., and D.C. performed *de novo* genome assembly; E.W. and C.W.D. performed genome-scale annotation and duplication analysis. S. Yoshida, T.S., Y.I., J.M.M., A.L., J.K.I., T.W., H.K., T.K., H.S., T.N., Y.S., S. Yamaguchi, K.Y., Y.S.-S., C.E.C., D.C.N., S.L., P.M., C.H., J.C.M., and T.D. performed gene annotation and individual gene family analysis. Y.W. and M.P.T. analyzed transcriptional factors. E.W., L.H., Z.Y., J.D., and C.W.D. performed comparative genome analysis, phases of parasite evolution, and whole-genome duplication analysis. J.T., H.G., and A.H.S. performed TE annotation and searched for horizontally transferred TEs. S.C. performed *in situ* hybridization. Y.I. and S. Yoshida analyzed the transcriptome data. S. Yoshida, Y.I., T.S., S.C., S.K., Y.-M.K., D.C., S.L., P.M., E.K.W., L.H., C.W.D., M.P.T., A.H.S., and K.S. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 1, 2019
Revised: July 23, 2019
Accepted: July 30, 2019
Published: September 12, 2019

REFERENCES

1. Ejeta, G. (2007). The Striga scourge in Africa: a growing pandemic. In Integrating New Technologies for Striga Control, towards Ending the Witch-Hunt, G. Ejeta, and J. Gressel, eds. (World Scientific Publishing), pp. 3–16.
2. Iverson, R.D., Westbrooks, R.G., Eplee, R.E., and Tasker, A.V. (2011). Overview and status of the witchweed (Striga asiatica) eradication program in the Carolinas. In Invasive Plant Management Issues and Challenges in the United States: 2011 Overview, A.R. Leslie, and R.G. Westbrooks, eds. (American Chemical Society), pp. 51–68.
3. Spallek, T., Mutuku, M., and Shirasu, K. (2013). The genus *Striga*: a witch profile. Mol. Plant Pathol. *14*, 861–869.
4. Cook, C.E., Whichard, L.P., Turner, B., Wall, M.E., and Egley, G.H. (1966). Germination of witchweed (*Striga Lutea* Lour)—isolation and properties of a potent stimulant. Science *154*, 1189–1190.
5. Al-Babili, S., and Bouwmeester, H.J. (2015). Strigolactones, a novel carotenoid-derived plant hormone. Annu. Rev. Plant Biol. *66*, 161–186.
6. Joel, D.M. (2013). The haustorium and the life cycles of parasitic Orobanchaceae. In Parasitic Orobanchaceae, D.M. Joel, J. Gressel, and L.J. Musselman, eds. (Springer), pp. 21–24.
7. Chang, M., and Lynn, D.G. (1986). The haustorium and the chemistry of host recognition in parasitic angiosperms. J. Chem. Ecol. *12*, 561–579.
8. Yoshida, S., Maruyama, S., Nozaki, H., and Shirasu, K. (2010). Horizontal gene transfer by the parasitic plant *Striga hermonthica*. Science *328*, 1128.
9. Westwood, J.H., Yoder, J.I., Timko, M.P., and dePamphilis, C.W. (2010). The evolution of parasitism in plants. Trends Plant Sci. *15*, 227–235.
10. Yang, Z., Zhang, Y., Wafula, E.K., Honaas, L.A., Ralph, P.E., Jones, S., Clarke, C.R., Liu, S., Su, C., Zhang, H., et al. (2016). Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. Proc. Natl. Acad. Sci. USA *113*, E7010–E7019.
11. Searcy, D.G., and MacInnis, A.J. (1970). Measurements by DNA renaturation of the genetic basis of parasitic reduction. Evolution *24*, 796–806.
12. dePamphilis, C.W. (1995). Genes and genomes. In Parasitic Plants, M. Press, and J. Graves, eds. (Chapman and Hall).
13. Yang, Z., Wafula, E.K., Honaas, L.A., Zhang, H., Das, M., Fernandez-Aparicio, M., Huang, K., Bandaranayake, P.C., Wu, B., Der, J.P., et al. (2015). Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol. Biol. Evol. *32*, 767–790.
14. Westwood, J. (2013). The physiology of the established parasite-host association. In Parasitic Orobanchaceae, D.M. Joel, J. Gressel, and L.J. Musselman, eds. (Springer Berlin Heidelberg).
15. Rogers, W.E., and Nelson, R.R. (1962). Penetration and nutrition of *Striga asiatica*. Phytopathology *52*, 1064–1070.
16. Wicke, S., Müller, K.F., de Pamphilis, C.W., Quandt, D., Wickett, N.J., Zhang, Y., Renner, S.S., and Schneeweiss, G.M. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell *25*, 3711–3725.
17. Wickett, N.J., Honaas, L.A., Wafula, E.K., Das, M., Huang, K., Wu, B., Landherr, L., Timko, M.P., Yoder, J., Westwood, J.H., and dePamphilis, C.W. (2011). Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. Curr. Biol. *21*, 2098–2104.
18. Press, M.C., Tuohy, J.M., and Stewart, G.R. (1987). Gas exchange characteristics of the sorghum-striga host-parasite association. Plant Physiol. *84*, 814–819.
19. Fujioka, H., Samejima, H., Suzuki, H., Mizutani, M., Okamoto, M., and Sugimoto, Y. (2019). Aberrant protein phosphatase 2C leads to abscisic acid insensitivity and high transpiration in parasitic Striga. Nat. Plants *5*, 258–262.
20. Press, M.C., Smith, S., and Stewart, G.R. (1991). Carbon acquisition and assimilation in parasitic plants. Funct. Ecol. *5*, 278–283.
21. Waters, M.T., Nelson, D.C., Scaffidi, A., Flematti, G.R., Sun, Y.K., Dixon, K.W., and Smith, S.M. (2012). Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in Arabidopsis. Development *139*, 1285–1295.
22. Scaffidi, A., Waters, M.T., Sun, Y.K., Skelton, B.W., Dixon, K.W., Ghisalberti, E.L., Flematti, G.R., and Smith, S.M. (2014). Strigolactone hormones and their stereoisomers signal through two related receptor proteins to induce different physiological responses in Arabidopsis. Plant Physiol. *165*, 1221–1232.
23. Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., and Nelson,

D.C. (2015). PLANT EVOLUTION. Convergent evolution of strigolactone perception enabled host detection in parasitic plants. Science *349*, 540–543.

24. Tsuchiya, Y., Yoshimura, M., Sato, Y., and Kuwata, K. (2015). Probing strigolactone receptors in Striga hermonthica with fluorescence. Science *349*, 864–868.
25. Toh, S., Holbrook-Smith, D., Stogios, P.J., Onopriyenko, O., Lumba, S., Tsuchiya, Y., Savchenko, A., and McCourt, P. (2015). Structure-function analysis identifies highly sensitive strigolactone receptors in Striga. Science *350*, 203–207.
26. Hirayama, K., and Mori, K. (1999). Synthesis of (+)-Strigol and (+)-Orobanchol, the germination stimulants, and their stereoisomers by employing lipase-catalyzed asymmetric acetylationas the key step. Eur. J. Org. Chem. *1999*, 2211–2217.
27. Yoshida, S., and Shirasu, K. (2009). Multiple layers of incompatibility to the parasitic witchweed, *Striga hermonthica*. New Phytol. *183*, 180–189.
28. Visser, J.H., Inge, D., and Kollmann, R. (1984). The "hyaline body" of the root parasite *Alectra orobanchoides* benth. (Scrophulariaceae)—Its anatomy, ultrastructure and histochemistry. Protoplasma *121*, 146–156.
29. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*, D490–D495.
30. Peter, I., and Davidson, E. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970–985.
31. Lavenus, J., Goh, T., Roberts, I., Guyomarc'h, S., Lucas, M., De Smet, I., Fukaki, H., Beeckman, T., Bennett, M., and Laplaze, L. (2013). Lateral root development in *Arabidopsis*: fifty shades of auxin. Trends Plant Sci. *18*, 450–458.
32. Swarup, K., Benková, E., Swarup, R., Casimiro, I., Péret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., et al. (2008). The auxin influx carrier LAX3 promotes lateral root emergence. Nat. Cell Biol. *10*, 946–954.
33. Berckmans, B., Vassileva, V., Schmid, S.P.C., Maes, S., Parizot, B., Naramoto, S., Magyar, Z., Alvim Kamei, C.L., Koncz, C., Bögre, L., et al. (2011). Auxin-dependent cell cycle reactivation through transcriptional regulation of *Arabidopsis* E2Fa by lateral organ boundary proteins. Plant Cell *23*, 3671–3683.
34. Ishida, J.K., Yoshida, S., Ito, M., Namba, S., and Shirasu, K. (2011). *Agrobacterium rhizogenes*-mediated transformation of the parasitic plant Phtheirospermum japonicum. PLoS ONE *6*, e25802.
35. De Smet, I., White, P.J., Bengough, A.G., Dupuy, L., Parizot, B., Casimiro, I., Heidstra, R., Laskowski, M., Lepetit, M., Hochholdinger, F., et al. (2012). Analyzing lateral root development: how to move forward. Plant Cell *24*, 15–20.
36. Gifford, M.L., Dean, A., Gutierrez, R.A., Coruzzi, G.M., and Birnbaum, K.D. (2008). Cell-specific nitrogen responses mediate developmental plasticity. Proc. Natl. Acad. Sci. USA *105*, 803–808.
37. De Smet, I., Vassileva, V., De Rybel, B., Levesque, M.P., Grunewald, W., Van Damme, D., Van Noorden, G., Naudts, M., Van Isterdael, G., De Clercq, R., et al. (2008). iReceptor-like kinase ACR4 restricts formative cell divisions in the Arabidopsis root Ive. Science *322*, 594–597.
38. DiDonato, R.J., Arbuckle, E., Buker, S., Sheets, J., Tobar, J., Totong, R., Grisafi, P., Fink, G.R., and Celenza, J.L. (2004). Arabidopsis ALF4 encodes a nuclear-localized protein required for lateral root formation. Plant J. *37*, 340–353.
39. Kim, G., LeBlanc, M.L., Wafula, E.K., dePamphilis, C.W., and Westwood, J.H. (2014). Plant science. Genomic-scale exchange of mRNA between a parasitic plant and its hosts. Science *345*, 808–811.
40. Vogel, A., Schwacke, R., Denton, A.K., Usadel, B., Hollmann, J., Fischer, K., Bolger, A., Schmidt, M.H.W., Bolger, M.E., Gundlach, H., et al. (2018). Footprints of parasitism in the genome of the parasitic flowering plant Cuscuta campestris. Nat. Commun. *9*, 2515.
41. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12.
42. Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., Shen, G., Qi, J., Qin, Y., Li, J., et al. (2018). Large-scale gene losses underlie the genome evolution of parasitic plant Cuscuta australis. Nat. Commun. *9*, 2683.
43. Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578–579.
44. Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C.J., et al. (2014). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol. *164*, 513–524.
45. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.
46. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
47. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.
48. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.
49. Mirarab, S., Nguyen, N., Guo, S., Wang, L.-S., Kim, J., and Warnow, T. (2015). PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J. Comput. Biol. *22*, 377–386.
50. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.
51. Haas, B.J., Delcher, A.L., Wortman, J.R., and Salzberg, S.L. (2004). DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics *20*, 3643–3646.
52. Ames, R.M., Money, D., Ghatge, V.P., Whelan, S., and Lovell, S.C. (2012). Determining the evolutionary history of gene families. Bioinformatics *28*, 48–55.
53. Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics *9*, 18.
54. Steinbiss, S., Willhoeft, U., Gremme, G., and Kurtz, S. (2009). Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. *37*, 7002–7013.
55. Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. *11*, R14.
56. Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.
57. Wall, P.K., Leebens-Mack, J., Chanderbali, A.S., Barakat, A., Wolcott, E., Liang, H., Landherr, L., Tomsho, L.P., Hu, Y., Carlson, J.E., et al. (2009). Comparison of next generation sequencing technologies for transcriptome characterization. BMC Genomics *10*, 347.
58. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. *33*, 1635–1638.
59. Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97–100.
60. Tang, H., Lyons, E., Pedersen, B., Schnable, J.C., Paterson, A.H., and Freeling, M. (2011). Screening synteny blocks in pairwise genome comparisons through integer programming. BMC Bioinformatics *12*, 102.

61. Tang, H., and Lyons, E. (2012). Unleashing the genome of brassica rapa. Front. Plant Sci. *3*, 172.

62. Zimmermann, P., Bleuler, S., Laule, O., Martin, F., Ivanov, N.V., Campanoni, P., Oishi, K., Lugon-Moulin, N., Wyss, M., Hruz, T., and Gruissem, W. (2014). ExpressionData—a public resource of high quality curated datasets representing gene expression across anatomy, development and experimental conditions. BioData Min. *7*, 18.

63. Hejátko, J., Blilou, I., Brewer, P.B., Friml, J., Scheres, B., and Benková, E. (2006). In situ hybridization technique for mRNA detection in whole mount Arabidopsis samples. Nat. Protoc. *1*, 1939–1946.

64. Gladyshev, E.A., Meselson, M., and Arkhipova, I.R. (2008). Massive horizontal gene transfer in bdelloid rotifers. Science *320*, 1210–1214.

65. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.

66. Llorens, C., Futami, R., Covelli, L., Domínguez-Escribá, L., Viu, J.M., Tamarit, D., Aguilar-Rodríguez, J., Vicente-Ripolles, M., Fuster, G., Bernet, G.P., et al. (2011). The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. *39*, D70–D74.

67. Miele, V., Penel, S., and Duret, L. (2011). Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics *12*, 116.

68. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

69. Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads, *24*, pp. 1384–1395.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
(+)-Strigol	Prof. Kenji Mori (Tokyo Univ.)	N/A
MS salts	Wako Chemical, Kyoto, Japan	# 392-00591
Phytagel	Sigma & Aldrich	# P8169
Critical Commercial Assays		
Nucleon Phytopure DNA extraction kit	GE healthcare	#RPN8510
TruSeq DNA Sample Prep kit	Illumina	#15026486
Mate-Pair library Prep kit	Illumina	# PE-930-1003
RNAeasy Plant Kit	QIAGEN	# 74904
TruSeq RNA Sample Prep kit	Illumina	# RS-122-2001
ReverTra Ace qPCR RT Kit	Toyobo	# FSQ-201
THUNDERBIRD SYBR qPCR kit	Toyobo	# QPS-201
Deposited Data		
S. asiatica genome sequence raw data	DDBJ: DRA008823	N/A
S. asiatica transcriptome sequence raw data	DDBJ: DRA0083088	N/A
S. hermonthica transcriptome sequence raw data	DDBJ: DRA008615, DDBJ: DRA003608	N/A
S. hermonthica genome sequence raw data	genbank: PRJNA551337	N/A
S. gesnerioides genome sequence raw data	genbank: PRJNA551339	N/A
S. asiatica genome assembly and annotation	DDBJ: BKCP01000001-BKCP01013846	N/A
S. hermonthica transcriptome assembly	DDBJ: ICPL01000001-ICPL01081559	N/A
S. asiatica BAC-end sequences	DDBJ: LB427106-LB478049	N/A
Experimental Models: Organisms/Strains		
Striga asiatica	Provided from Prof. Mike Timko (U. Virginia, VA, USA)	UVA1
Striga hermonthica	Provided from Prof. Abdel G. E. Babiker (Environment and Natural Resources and Desertification Research Institute, Sudan)	N/A
Oryza sativa (japonica, c.v. Koshihikari)	Rice Genome Resource Center (RGRC), Tsukuba, Japan	N/A
Arabidopsis thaliana (ecotype: Col-0)	*Arabidopsis* biological resource center (ABRC)	Col-0
Lotus japonicus (ecotype: MG-20)	Legume base (https://www.legumebase.brc.miyazaki-u.ac.jp/lotus/)	N/A
Oligonucleotides		
See Data S1S	N/A	N/A
Software and Algorithms		
FLASH	https://ccb.jhu.edu/software/FLASH/	N/A
Platanus	[69]	N/A
SSPACE	[43]	N/A
MAKER-P	[44]	N/A
CLC assembly cell	https://filgen.jp/Product/BioScience21-software/CLC/index11-g.htm	N/A
CLC genomic workbench	https://filgen.jp/Product/BioScience21-software/CLC/index11-g.htm	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	N/A
RSEM	http://deweylab.github.io/RSEM/	N/A
Orthofinder	[45]	N/A
RaxML	[46]	N/A
FastTree	[47]	N/A
trimAl	[48]	N/A
PASTA	[49]	N/A
CODEML	[50]	N/A
CoGE	https://genomevolution.org	N/A
DAGChainer	[51]	N/A
DupliPHY	[52]	N/A
RaxML	[49]	N/A
LtrHarvest	[53]	N/A
LtrDigest	[54]	N/A
Other		
Striga asiatica genome and predicted genes and protein in multifasta format, annotation in gff3 file format.	https://datadryad.org/	https://doi.org/10.5061/dryad.53t3574
Striga hermonthica transcriptome assembly and predicted protein sequences in multifasta format, and functional annotation and GO information	https://datadryad.org/	https://doi.org/10.5061/dryad.53t3574
Retrotransposon sequences and phylogenetic trees appeared in Figures 7E and S4	https://datadryad.org/	https://doi.org/10.5061/dryad.53t3574

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for materials, resources and reagents, including mosquito lines, should be directed to and will be fulfilled by the Lead Contact, Ken Shirasu (ken.shirasu@riken.jp)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Seeds of the *S. asiatica* US strain were originally obtained from the USDA Methods Development Center (Whiteville, N.C.) and the seeds from a single plant after six rounds of self-fertilization were used as starting materials. The seeds were surface sterilized with 5% commercial bleach solution (containing final sodium hypochlorite concentration at approx. 0.3%) for 5 min and washed with excess amount of sterile water at least 5 times. The sterile seeds were preconditioned on GM media (full strength of MS salts, 0.01% Myo-inositol, 1% Sucrose, 0.5% Phytagel (Sigma)) for 10 days and the germination was induced by adding 10 nM strigol [26]. The germinated *S. asiatica* seedlings were transferred to new GM media and grown *in vitro* in a 26°C chamber at a long-day (16-h light/8-h dark) condition. For *S. asiatica* shoot propagation, the shoots were cut and transferred to new GM media every month. When *S. asiatica* shoots were transferred into the new GM media, multiple shoots were induced.

S. asiatica and *S. hermonthica* infection to rice (*Oryza sativa*, c.v. Koshihikari) was performed in the rhizotron system as previously published [27]. *S. hermonthica* seed and seedling samples were collected after preconditioning on glass-fiber filter paper (Watman GF/A) for 10 days, and before and after 10 nM strigol treatment for 2 days, respectively. *S. hermonthica* samples for 1-day post infection were carefully removed from rice roots using forceps. For the 3- and 7-day post infection samples, haustorial parts (include host tissues) were carefully excised using razor blades. For the control, rice roots without *S. hermonthica* infection were also harvested at the same day as 7-d samples. All samples were collected in triplicates of independent experiments. *S. asiatica* haustorium samples were harvested by excising the infected parts with a razor blade together with rice roots. For shoot and root samples, the sterile *S. asiatica* seeds were germinated on MS media containing sucrose and grown *in vitro* for one month.

METHOD DETAILS

Whole genome shotgun sequencing, assembly and annotation of S. asiatica

The genomic DNA for Illumina library preparation was obtained from *S. asiatica* shoots derived from a single plant. The genomic DNA for BAC library was prepared from the siblings of the plant. The genomic DNA was extracted by using Phytopure DNA extraction kit

(GE healthcare) according to manufacturer's instructions. The Illumina paired-end (PE) and mate-pair (MP) libraries were prepared using the TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA) and Mate-Pair Library Prep Kit (Illumina, San Diego, CA) from according to the manufacturer's instructions. A bacterial artificial chromosome (BAC) library with an average length of 120 kbp was prepared with CopyControl pCC1BAC vector by Amplicon Express Ltd (Washington, USA) and the BAC-end sequencing was performed in the Kazusa DNA Research Institute (Kisarazu, Japan). Whole genome shotgun (WGS) sequencing and BAC-end sequencing were done through Illumina HiSeq 2000 and Sanger ABI3730x1 platforms. Raw sequence data were filtered for bacterial genome contamination, PCR-duplicated reads and low quality reads were error-corrected. Paired-end Illumina reads were merged by FLASH to make longer single reads and the genome assembly and scaffolding were performed by Platanus [69] and by SSPACE [43]. The gene model predictions were performed using MAKER pipeline [44] using *S. asiatica* RNA sequencing described below. Details of read processing, assembly and annotation are described in Data S2A and S2B.

RNA sequencing

Total RNA was extracted from shoots and roots using the RNAeasy Plant Kit (QIAGEN). Illumina PE libraries were constructed using the TruSeq RNA Sample Prep Kit (Illumina) and sequenced by an Illumina HiSeq2000 for 101 cycles per run. The obtained *S. asiatica* RNA sequences were quality-filtered and then used for the gene annotation pipeline and validation of the assembly. *S. hermonthica* sequences were quality trimmed with the fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) using the fastq_quality_trimmer with option –l 60 and –t 30 and assembled by CLC genomics workbench (ver. 5) after removing host gene contamination (for details, see Data S2E). The sequence reads were mapped on *S. hermonthica de novo* assembled contigs concatenated with rice cDNAs by bowtie2. The contigs that are mapped with rice control reads were excluded from the subsequent analysis to avoid contamination of rice sequences. The normalized FPKM values were calculated by RSEM program (for details, see Data S2E). After selecting genes in the upper 75% and 50% quartile of coefficient of variation for the expression across samples, scaled expression values within tissues were used to cluster these genes for a multilevel 3 × 4 hexagonal self-organizing map (SOM). The outcome of SOM clustering was visualized in PCA space where PC values were calculated based on gene expression across samples (R stats package, prcomp function). GO enrichment analysis of contigs detected in SOM was performed using the GOSeq Bioconductor package [55] with Benhamini and Hochberg multiple hypothesis testing correction.

Genome comparative analysis

A maximum likelihood species tree for the 26 representative plant genomes was estimated using a concatenated matrix of trimmed codon alignments for genes from 1,440 BUSCO single copy orthogroups with RAxML [46] (Figure 1). Protein coding genes from 26 plant genomes (Data S1C) including *S. asiatica* were classified into orthogroups using the Orthofinder version 1.1.8 algorithm [45]. We further performed a second iteration of MCL [56] to connect distantly related orthogroups into superorthogroups as described in Wall et al. [57]. Amino acid sequence alignments for each orthogroup were generated with PASTA [49] using a maximum of five iterative refinements. Corresponding DNA codon alignments were trimmed using the heuristic automated method implemented in trimAl version 1.4.rev8 [48]. Approximately-maximum likelihood (ML) analyses were conducted using FastTree version 2.1.10 [47], searching for the best ML tree with the GTR and GAMMA models. The unrooted FastTree phylogenies were traversed and rooted with the most distant taxa the orthogroup using rooting functions implemented in ETE Toolkit, a python phylogenetic framework [58]. The trees were examined for gene duplications in *Striga* and *Mimulus* and the detected duplications were scored using a scoring strategy similar to that described by Jiao et al. [59]. A synonymous mutation (K_s) value for each duplicated sequence pair was calculated using the ML method implemented in CODEML [50] with a minimum alignment length of 300 bp. Structural syntenic analyses were performed with the SynMap tool [60] of the CoGe comparative genomics platform [61]. The genomes of *Mimulus* and *Vitis* were compared to the genome of *Striga* with the chaining algorithm DAGChainer [51] with a maximum distance of 20 genes between gene matches, and a minimum of 5 genes to seed a syntenic region. Scaffolds and contigs of *Striga* were ordered and oriented based on their syntenic path to both *Mimulus* and *Vitis*. Parsimony method in DupliPHY [52] was used for reconstruction of the presence and size of each gene family in the common ancestor of *S. asiatica* and of the closely related non-parasite *Mimulus guttatus* as well as of other successively earlier ancestors. The numbers of evolutionary events were estimated using gene counts in each orthogroup at each node of the 26-genome species tree. The tissue-specific orthogroups were defined using *Arabidopsis* microarray expression data [62]. These data are a curated summary of more than 5,000 microarray experiments conducted using the Agilent ATH1 GeneChip. Further details are described in Supplementary Information Section 3. Comparison of the genomic regions containing *KAI2* paralogs was performed by GEvo tool in CoGe. The 60 kb regions containing each *KAI2*, *D14* or *DLK2* paralog were submitted to GEvo with blastZ threshold score 15000. The data is visualized with Circos plot (http://circos.ca). Duplication origins of these loci were predicted as described in Supplementary Information Section 3.3.1.

RT-qPCR

Total RNAs were extracted as described above. cDNAs were synthesized using ReverTra Ace qPCR RT Kit (Toyobo, Japan) and quantitative PCRs were conducted using THUNDERBIRD SYBR qPCR kit (Toyobo, Japan) in Mx3000P qPCR system (Agilent Technologies). RT-qPCR was performed in three segments. Segment 1 consisted of 1 min at 95°C for one cycle, segment 2 consisted either of 15 s at 95°C and 30 s at 60°C for 40 cycles, or 15 s at 95°C, 30 s at 55°C and 30 s at 72°C for 40 cycles and segment 3 consisted of 1 min at 95°C, 30 s at 55°C, and 30 s at 95°C for one cycle. The primer sequences used are listed in Data S1S.

In situ hybridization

Preparation of DIG labeled RNA probe was performed as described previously [63]. The probe fragments were amplified by PCR from the cDNA library of rice infected with *S. hermonthica* using the primers listed in Data S1S. Sense or antisense probes with the length of 600-900 bp were generated using the T7 or SP6 polymerase (Roche) and DIG-UTP mix (Roche). The haustorial tissues attached with host rice were fixed in the freshly prepared PFA fixation buffer composed of 4% (w/v) paraformaldehyde in 1 × PBS buffer (130 mM NaCl, 7 mM Na_2HPO_4, 3 mM NaH_2PO_4, pH 7.4 adjusted by NaOH). The samples were dehydrated by incubation in 1xPBS for 2.5 h and the concentration of ethanol was gradually increased at 4°C (30% for 1 h, 50% for 1h, 70% for overnight, 85% for 1h, 95% for overnight and 100% for 3 h). Samples were then permeabilised by incubation in gradually increasing concentrations of Histo-Clear in ethanol at room temperature (Histo-Clear and ethanol mixture of 1:3 for 1 h, 1:1 for 1h, 3:1 for 1 h and 100% of Histo Clear for 2 h) and in a 1:1 mixture of Histo-Clear and paraffin for 1 h at 60°C. Paraffin was changed 6 times before being embedded on wooden blocks. We followed the steps of *in situ* hybridization as described previously [63] with minor modifications; a concentration of 10 $\mu g/ml^{-1}$ of the probes was used and the use of levamisole in the detection solution was omitted. The images of *in situ* hybridized samples were taken using the light microscopy BX-51 (Olympus).

Identification of horizontally transferred genes and retrotransposons

To analyze the *S. asiatica* genome for genes horizontally transferred from grass host species, the *S. asiatica* annotation was subjected to a BLASTp search with the threshold e-value 1e-10 against a database of combined predicted proteins from the genome of 28 different plant species, including *Striga* host plants, rice, sorghum, foxtail millet, and maize. *S. asiatica* proteins having at least one hit to grass species in their top 20 hits were selected, and modified Alien Index (AI) values [64] were calculated with the following formula: Modified AI = log((Best E-value for dicots) + 1e-200) - log((Best E-value for grasses) + 1e-200). Genes having modified AI > 30 and genes that did not have a dicot hit were selected for further analysis. Using the RAxML program, maximum-likelihood phylogenetic trees were estimated with BLASTp-hit homolog genes from the 28-species database as well as from the non-redundant (nr) database. Manual investigation of the phylogenetic trees found 34 positive HGT candidate genes, which were assigned into 20 orthogroups by orthoMCL analysis. A few of HGT candidates are near each other in the genome, and therefore the genomic regions were compared using CoGE with the GEvo function.

For identification of horizontally transferred retrotransposons, superfamily *Copia* and *Gypsy* elements were retrieved, using LtrHarvest [53] and LtrDigest [54], from the genome sequences of *S. asiatica* and those of the monocots *Sorghum bicolor*, *Zea mays*, *Oryza sativa* ssp. *japonica* and ssp. *indica*, *O. rufipogon*, and *O. glaberrima* and the eudicots *Glycine max*, *Solanum tuberosum*, and *Vitis vinifera*. The *rt* sequences were clustered and the *S. asiatica rt* sequences that were found in clusters mixed with those of other genomes were treated further. These were characterized by exonerate-search [65] using known *rt* sequences from GypsyDB [66] and clustered by homology search against each other (BLASTn -evalue 1e-20) and subsequently by silix-software [67] (silix -i 0.60 -r 0.70). The resulting clusters were aligned with the clustal-omega [68] and prank-ms [65] multiple aligners and phylogenetic trees were constructed by FastTree (fasttree –nt –gtr –gamma) [47]. The details of HGT analysis are described in Data S2F.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses for GO enrichment was performed with either the chi-square test or Fisher's exact test with Benjamini and Hochberg correction for multiple samples. Other statistical analyses were performed with two-tailed Mann-Whitney U test, Student's t test, or one-way ANOVA combined with the post hoc Tukey-Kramer test as indicated in the text or figure legends. Error bars represent SEM.

DATA AND CODE AVAILABILITY

S. asiatica genome and transcriptome sequence data are deposited in DDBJ as accession number DDBJ: DRA007962, DDBJ: DRA008823 and DDBJ: DRA008308. The *S. hermonthica* RNA-seq data are available as accession numbers DDBJ: DRA008615 and DDBJ: DRA003608 in DDBJ. *S. hermonthica* and *S. gesnerioides* genome sequence raw reads are deposited in GenBank as accession number Genbank: PRJNA551337 and Genbank: PRJNA551339, respectively. *S. asiatica* genome assembly and annotation, *S. hermonthica* transcriptome assembly and annotation, and horizontally transferred retrotransposon sequences are available at Dryad data repository (http://datadryad.org/reource/doi:10.5061/dryad.53t3574).

All bioinformatic analyses were performed with open-source or commercially available software. Perl, Python or R scripts were used for run each software according to software manuals.