Report

Current Biology

A Massive Expansion of Effector Genes Underlies Gall-Formation in the Wheat Pest *Mayetiola destructor*

Highlights

- The plant-galling *Mayetiola destructor* genome is replete with putative effector genes
- The SSGP-71 effector gene family is the largest known arthropod gene family
- SSGP-71 genes encode E3-ubiquitin-ligase mimics
- Two SSGP-71 genes elicit effector-triggered immunity in resistant wheat

Authors

Chaoyang Zhao,
Lucio Navarro Escalante, ...,
Jeffery J. Stuart, Stephen Richards

Correspondence

stephenr@bcm.edu

In Brief

Zhao et al. reveal the effects of a genetic arms race on a plant-galling insect genome. One-eighth of the genes encode putative gall effectors, and the largest family encodes 426 secreted E3-ubiquitin-ligase-mimicking proteins necessary for galling and is similar to bacterial plant pathogen effectors, a remarkable example of molecular convergence.

Zhao et al., 2015, Current Biology *25*, 613–620
March 2, 2015
http://dx.doi.org/10.1016/j.cub.2014.12.057

CellPress

Current Biology 25, 613–620, March 2, 2015 http://dx.doi.org/10.1016/j.cub.2014.12.057

Report

A Massive Expansion of Effector Genes Underlies Gall-Formation in the Wheat Pest *Mayetiola destructor*

Chaoyang Zhao,[1,27] Lucio Navarro Escalante,[1] Hang Chen,[2,3] Thiago R. Benatti,[1,28] Jiaxin Qu,[4] Sanjay Chellapilla,[5] Robert M. Waterhouse,[6,7,8,9] David Wheeler,[10] Martin N. Andersson,[11] Riyue Bao,[12,13] Matthew Batterton,[4] Susanta K. Behura,[14] Kerstin P. Blankenburg,[4] Doina Caragea,[5,15] James C. Carolan,[16] Marcus Coyle,[4] Mustapha El-Bouhssini,[17] Liezl Francisco,[4] Markus Friedrich,[13] Navdeep Gill,[18] Tony Grace,[5] Cornelis J.P. Grimmelikhuijzen,[19] Yi Han,[4] Frank Hauser,[19] Nicolae Herndon,[5] Michael Holder,[4] Panagiotis Ioannidis,[6,7] LaRonda Jackson,[4] Mehwish Javaid,[4] Shalini N. Jhangiani,[4] Alisha J. Johnson,[20] Divya Kalra,[5] Viktoriya Korchina,[4] Christie L. Kovar,[4] Fremiet Lara,[4] Sandra L. Lee,[4] Xuming Liu,[21] Christer Löfstedt,[11] Robert Mata,[4] Tittu Mathew,[4] Donna M. Muzny,[4] Swapnil Nagar,[5] Lynne V. Nazareth,[4] Geoffrey Okwuonu,[4] Fiona Ongeri,[4] Lora Perales,[4] Brittany F. Peterson,[1] Ling-Ling Pu,[4] Hugh M. Robertson,[22] Brandon J. Schemerhorn,[20] Steven E. Scherer,[4] Jacob T. Shreve,[1,29] DeNard Simmons,[4] Subhashree Subramanyam,[23] Rebecca L. Thornton,[4] Kun Xue,[24] George M. Weissenberger,[4] Christie E. Williams,[25] Kim C. Worley,[4] Dianhui Zhu,[4] Yiming Zhu,[4] Marion O. Harris,[26] Richard H. Shukle,[20] John H. Werren,[10] Evgeny M. Zdobnov,[6,7] Ming-Shun Chen,[21] Susan J. Brown,[5] Jeffery J. Stuart,[1] and Stephen Richards[4,*]

[1]Department of Entomology, Purdue University, West Lafayette, IN 47097, USA
[2]Department of Entomology, Kansas State University, Manhattan, KS 66056, USA
[3]The Research Institute of Resource Insects, Chinese Academy of Forestry, Bailongsi, Kunming 650224, People's Republic of China
[4]Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
[5]KSU Bioinformatics Center, Division of Biology, Kansas State University, Manhattan, KS 66506, USA
[6]Department of Genetic Medicine and Development, University of Geneva Medical School, Rue Michel-Servet 1, 1211 Geneva, Switzerland
[7]Swiss Institute of Bioinformatics, Rue Michel-Servet 1, 1211 Geneva, Switzerland
[8]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA
[9]The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
[10]Department of Biology, University of Rochester, Rochester, NY 14627, USA
[11]Department of Biology, Lund University, 223 62 Lund, Sweden
[12]The University of Chicago Bioinformatics Core, Biological Sciences Division, Center for Research Informatics, The University of Chicago, Chicago, IL 60637, USA
[13]Department of Biological Sciences, Wayne State University, Detroit, MI 48202, USA
[14]Department of Biological Sciences, University of Notre Dame, IN 46556, USA
[15]Department of Computer and Information Science, Kansas State University, Manhattan, KS 66506, USA
[16]Department of Biology, National University of Ireland Maynooth, Maynooth, Ireland
[17]International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat BP 6299, Morocco
[18]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[19]Center for Functional and Comparative Insect Genomics, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
[20]USDA-ARS, Department of Entomology, Purdue University, West Lafayette, IN 47907, USA
[21]USDA-ARS, Department of Entomology, Kansas State University, Manhattan, KS 66056, USA
[22]Department of Entomology, University of Illinois at Urbana-Champaign, 320 Morrill Hall, 505 South Goodwin Avenue, Urbana, IL 61801, USA
[23]Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA
[24]College of Life and Environment Sciences, Minzu University, Beijing 100081, China
[25]USDA-ARS, Department of Agronomy, Purdue University, West Lafayette, IN 47907 USA
[26]Department of Entomology, North Dakota State University, Fargo, ND 58108, USA

[27]Present address: Department of Entomology, The Ohio State University, Wooster, OH 44691, USA
[28]Present address: Centro de Tecnologia Canavieira, São Paulo 13400-970, Brazil
[29]Present address: Department of Biochemistry and Molecular Biology and the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA
*Correspondence: stephenr@bcm.edu

Summary

Gall-forming arthropods are highly specialized herbivores that, in combination with their hosts, produce extended phenotypes with unique morphologies [1]. Many are economically important, and others have improved our understanding of ecology and adaptive radiation [2]. However, the mechanisms that these arthropods use to induce plant galls are poorly understood. We sequenced the genome of the Hessian fly (*Mayetiola destructor*; Diptera: Cecidomyiidae), a plant parasitic gall midge and a pest of wheat (*Triticum* spp.), with the aim of identifying genic modifications that contribute to its plant-parasitic lifestyle. Among several adaptive modifications, we discovered an expansive reservoir of potential effector proteins. Nearly 5% of the 20,163 predicted gene models matched putative effector gene transcripts present in the *M. destructor* larval salivary gland. Another 466 putative effectors were discovered among the genes that have no sequence similarities in other organisms. The largest known arthropod gene family (family SSGP-71) was also discovered within the effector reservoir. SSGP-71

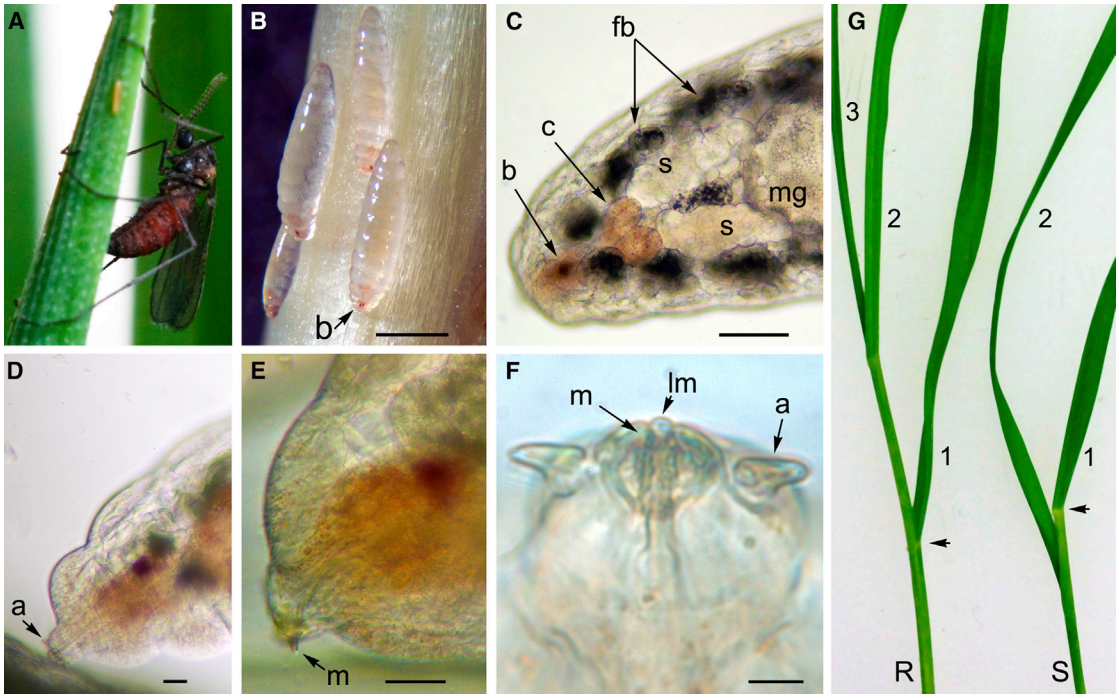

Figure 1. *M. destructor* and Its Damage to Wheat Seedlings

(A) Adult female depositing eggs on a wheat leaf. Females use chemo- and mechanoreceptors to identify and orient themselves on host plants for oviposition.

(B–F) Light-microscopic images of living 2-day-old first-instar larvae attacking individual epidermal cells using syringe-like mandibles and salivary secretions [14, 15]. a, antenna; b, supraesophageal ganglia; c, suboesophageal ganglia; fb, fat bodies; lm, labrum; m, mandible; mg, midgut; s, salivary gland.

(B) After hatching, first-instar larvae move down the leaf and under the older leaves to attack the base of the seedling (bar, 0.5 mm).

(C) View through the transparent dorsal cuticle of the larval thorax (bar, 0.1 mm).

(D) A larva with its mouthparts attached to the leaf surface (bar, 10 μm).

(E) Larva with its mandibles extended (bar, 5 μm).

(F) The head of a larva (bar, 5 μm).

(G) Near-isogenic wheat seedlings 10 days after *M. destructor* infestation. The resistant seedling (R) has three leaves (1–3) and shows no visible symptoms of *M. destructor* feeding. The susceptible seedling (S) has only two leaves and will grow no further. Note the position of the ligule of the first leaf on both plants (arrows) and how the second leaf blade of the S plant fails to grow past this point.

proteins lack sequence homologies to other proteins, but their structures resemble both ubiquitin E3 ligases in plants and E3-ligase-mimicking effectors in plant pathogenic bacteria. SSGP-71 proteins and wheat Skp proteins interact in vivo. Mutations in different SSGP-71 genes avoid the effector-triggered immunity that is directed by the wheat resistance genes *H6* and *H9*. Results point to effectors as the agents responsible for arthropod-induced plant gall formation.

Results and Discussion

Recent investigations suggest that "effector" proteins, structurally undefined proteins that suppress host defense and modulate host cell activities [3–6], are involved in the formation of insect-induced plant galls [7–10]. Other investigations have demonstrated that genomic sequencing can reveal the effector repertoires of important plant parasites [11–13]. We therefore sequenced the genome of the wheat-galling midge *Mayetiola destructor* (Figure 1), to search for effectors and other modifications that permit the insect to live as a gall-forming plant parasite. An inbred strain selected from the avirulent Great Plains (GP) biotype was used for sequencing. Approximately 26 million reads (34-fold genome coverage) were assembled with an existing fingerprinted contig (FPC)-based contig map, consisting of end-sequenced *M. destructor* bacterial artificial chromosomes (BACs) [16]. This allowed us to position 59% of the genome sequence on the four *M. destructor* polytene chromosomes. The final genome assembly (NCBI BioProject PRJNA45867) contains 153 Mb of assembled contig sequence with a 14 kb contig N50 length, across an extent of 186 Mb (including gaps between contigs) and a 756 kb scaffold N50 length.

To assess the completeness of the assembly, we examined the *M. destructor* genome using benchmarking sets of universal single-copy orthologs (BUSCOs). When genome assembly is complete, BUSCOs from the appropriate phylogenetic clade are found as single copies in the newly sequenced genome [17]. Using the monarch butterfly, *Danaus plexippus*, as the outgroup reference species, we compared the *M. destructor* genome assembly to two species of *Drosophila* and two mosquitoes. *M. destructor* has slightly more missing or duplicated BUSCOs than *Drosophila melanogaster* and *Anopheles gambiae*, but fewer than *Culex quinquefasciatus*. In addition, *M. destructor* BUSCO length recovery was similar to that of the other dipteran genomes. We therefore concluded that the *M. destructor* genome assembly was sufficiently robust for further analysis.

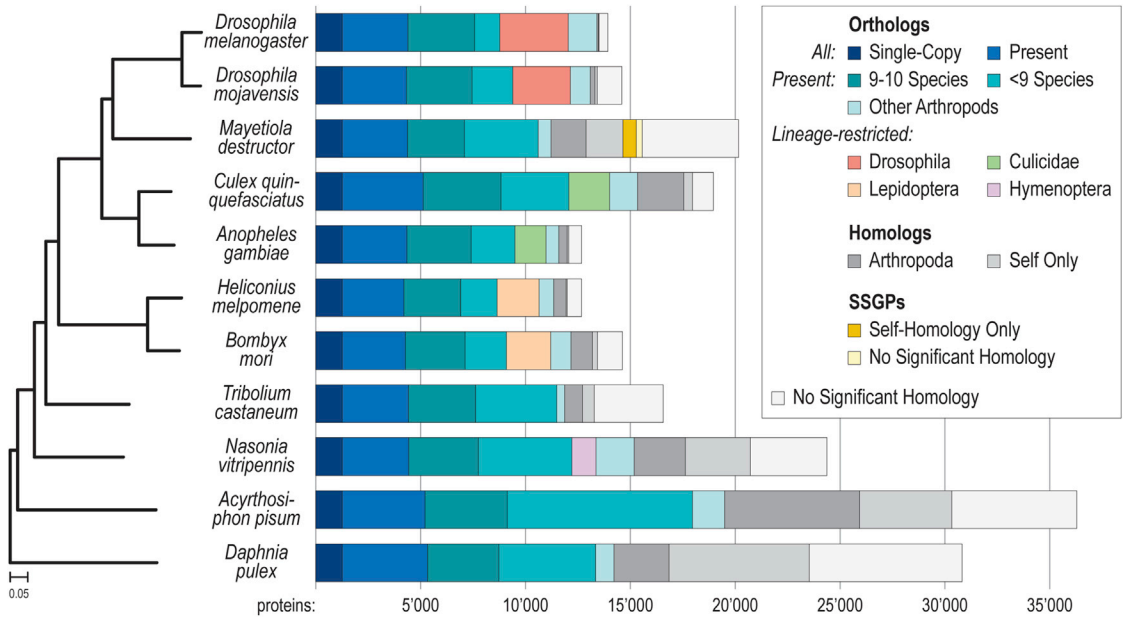

Figure 2. The Molecular Phylogeny and Orthology Homology Assignments of *M. destructor*

Genes were compared to four other dipterans, five species from other insect orders, and the crustacean outgroup, *Daphnia pulex*. The maximum-likelihood species phylogeny computed from the concatenated alignment of single-copy orthologous protein-coding genes places *M. destructor* as a sister group to the brachyceran Drosophilids; branch length units are in substitutions per site, and all nodes have bootstrap support values of 100%. The majority of the *M. destructor* gene repertoire has identifiable orthologs or homologs in other arthropod species, but it also includes relatively large numbers of single-copy genes (white; no-homology fraction) and gene families (light gray; self-homology fraction) that are unique to *M. destructor*. SSGP sequences (yellow and orange fractions) compose 15% of these gene fractions. Intriguingly, of the other insects with large no- and self-homology fractions, both the pea aphid and the parasitic wasp *Nasonia* are obligate parasites. Additional genes in *Daphnia* are suggested to be involved in environmental response. See also Figures S1–S5.

E Chromosomes Are S Chromosome Copies

Like other gall midges, *M. destructor* has germline-limited "E chromosomes" that are eliminated from the presumptive somatic cells during the early cleavage divisions of embryogenesis [8]. To test the hypothesis that E chromosomes are unlike somatic (S) chromosomes in sequence, we sequenced genomic DNA from female embryos collected before E chromosome elimination and compared the sequence with the newly assembled genome (Figure S1). Mapping rates of sequence reads derived before and after chromosome elimination were not significantly different. Thus, the E chromosomes appear to be additional copies of the four *M. destructor* S chromosomes. Results of fluorescence in situ hybridization (FISH) experiments using *M. destructor* BAC clones as probes are consistent with this interpretation (Figure S1). The mechanisms that maintain the structural integrity of the maternally inherited E chromosomes are unknown.

M. destructor Gene Phylogeny Groups Gall Midges with the Brachycera

Automated annotation of the *M. destructor* genome assembly using the MAKER2 pipeline [18], combined with expressed sequence tag (EST) sequence data, generated 20,163 protein-coding gene models. Homologous and orthologous relationships of these genes with those from other sequenced arthropods were assessed using the triangulated best-reciprocal-hit clustering approach of OrthoDB [17]. All analyses were performed using OrthoDB6 (http://www.orthodb.org), which includes 45 arthropod species. We found 1,255 universal single-copy orthologs across ten arthropod species (Figure 2), which were used to compute a maximum-likelihood phylogeny with RAxML [19]. Our analysis placed *M. destructor* as a sister group to the Drosophilids (Figure 2).

Dipteran diversity is traditionally partitioned into two principal suborders: the Nematocera comprises mosquito-like flies with long antennae, and the Brachycera contains stout and faster-moving flies with short antennae [20]. The gall midges, along with marsh flies, gnats, and other midges, make up the nematoceran infraorder, Bibionomorpha. Therefore, the placement of *M. destructor* with the Drosophilids (Brachycera) confirms the Nematocera to be a paraphyletic group, consistent with previous analyses placing the Bibionomorpha as a sister group to the Brachycera [20]. Interestingly, we did observe a fraction of rapidly evolving genes grouped with mosquitoes or basal to both flies and mosquitoes (Figure S3). These genes are candidates for lifestyle adaptations in the line leading to the gall midges.

Chemosensory Expansions and Contractions Suggest a Focused Sensory Response

Dipteran adaptations may be observed as changes in chemoreceptor repertoires [21]. In *M. destructor*, some basal dipteran odorant receptor (OR) genes are absent, but the total number of ORs (122) is twice the number in *Drosophila* and is within the range of ORs found in mosquitoes [22]. Most (107) of the ORs belong to two massively expanded *M. destructor*-specific lineages (Figure S3). Several of these have considerable sequence similarity, are clustered on chromosomes with up to 11 tandem duplicates (Supplemental Experimental Procedures and Document S2), and include positively selected amino acid sites (Figure S3). The majority of these are expressed in adults, and many show sex-specific

expression [23]. These changes are most likely related to the adult's short lifespan (<36 hr), in which both mating, mediated by extreme male attraction to a multi-component sex pheromone [24], and female host identification for oviposition must be accomplished.

Neuropeptide genes, their G-protein-coupled receptors, and opsins showed a notably different pattern of presence and absence compared to other sequenced species (Figure S4 and the Supplemental Experimental Procedures and Document S2). Gustatory receptors (GRs) for carbon dioxide, sugar, and bitter tastants were identified, but *M. destructor* appears to have lost the capacity to perceive several unidentified tastants. This contraction of the GR repertoire (only 28 GRs identified) was notable, although it was not as radical as that of the honey bee (*Apis mellifera*) [25] or the human body louse (*Pediculus humanus*) [26]. Genic reduction was also found in the family of ionotropic receptors (IRs), where only 39 genes were identified (Supplemental Experimental Procedures and Document S2). Most of the conserved olfactory "antennal IRs" with olfactory function are present, whereas the number of "divergent IRs" with putative roles in taste is greatly reduced in comparison to other Diptera [27]. Finally, 32 odorant-binding proteins (OBPs) were identified (Supplemental Experimental Procedures and Document S2), again a comparatively low number for a dipteran insect. As *M. destructor* adults do not feed and larvae have no capability for host selection, a reduced role for chemosensation is consistent with the general loss of chemoreceptors, in tandem with a contrasting expansion of those receptors in critical roles.

Lateral Gene Transfer of Carbohydrate Metabolic Enzymes

Gall midges evolved from detritivores, and their current capacity as plant parasites might be due to the genes they have acquired from bacterial or fungal associates [8, 28]. *M. destructor* also harbors a host of maternally transmitted bacteria [29] whose genomes are evident in the assembly. We therefore used a sequence homology-based pipeline [30] to look for lateral gene transfer (LGT) events in the *M. destructor* genome. Thirty candidate LGT events were identified (Figure S5). Most of these involve carbohydrate metabolic enzymes, an observation that suggests that *M. destructor* acquired these genes to better utilize plant carbohydrates. One candidate LGT has strong identity to the YD toxin from bacteriophage APSE-3 (Figure S5). Strains of the aphid bacterium *Hamiltonella defensa* harbor bacteriophage containing these loci to protect their hosts from parasitoid attack [31].

A Large Fraction of the Genome Contains an Effector Reservoir

To identify genes encoding *M. destructor* effectors, we focused on two features commonly observed among these proteins in eukaryotes [11–13]: N-terminal signal peptides (SPs) and evidence of rapid evolution, e.g., an absence of sequence similarities with other organisms, multiple gene copies, and high rates of amino acid variability. Given the manner in which the insect attacks its host (Figure 1), we also looked for first-instar larval expression. These features were previously observed among first-instar salivary gland transcripts that encode secreted salivary gland proteins (SSGPs) [32, 33]. Here we sought to reveal the genomic organization of all SSGPs, provide physical evidence of protein secretion, and reveal the limits of the complete effector repertoire. We also sought to identify structural motifs that point to effector mode of action and effectors that trigger resistance in wheat.

Compared to most sequenced insect genomes (Figure 2), *M. destructor* has a large fraction of genes (34%) lacking homologs in other organisms. Within this fraction, 919 SSGPs had a perfect match with a MAKER2 gene model; 284 were in the single-copy "no-homology" fraction, and 635 were in the multicopy "self-homology-only" fraction of *M. destructor* genes (Document S2). Many of the genes in the latter category are tandem duplications in short chromosomal regions that display the effects of diversifying selection, in which the regions encoding the mature proteins have as little as 25% similarity [33]. Taken together, these observations suggest that the fraction of genes unique to *M. destructor* contains a large reservoir of effector encoding sequences. To test this possibility, we used a computational pipeline for effector gene identification modeled after the one used to identify effectors in the aphid *Myzus persicae* [10]. The pipeline uses the SignalP v3.0 program [34] to predict the presence of signal peptides in the amino acid sequences encoded in genes expressed in first-instar larvae. It identified 466 putative effectors in the no-homology fraction (Table S1). Combined, these putative effectors represent over 7% of the genes in the *M. destructor* genome. Because presence at the site of action is evidence of effector biology, we used mass spectrometry to identify proteins at first-instar larval feeding sites, and several SSGPs were identified (Figure S6).

The Largest SSGP Family Encodes F-Box-LRR E3-Ligase-Mimicking Proteins

We next concentrated our analysis on the multi-gene family (SSGP-71) that had the greatest representation in our proteomic analysis. We found 426 SSGP-71 members (Document S2) that fit MAKER2 gene models, making it the largest arthropod gene family yet identified. Motifs in these proteins (described below) placed 59 SSGP-71 proteins within the "orthology-to-other-arthropods" fraction. The remaining 366 SSGP-71 genes reside in the self-only fraction. Like most other SSGP genes, the majority of SSGP-71 genes are composed of two exons, the first encoding the SP and the second the mature protein. SSGP-71 members usually encode longer proteins than other SSGP genes and, unlike other SSGPs, are relatively well dispersed throughout the genome, often in triplets. We found 50 SSGP-71 genes on chromosome A1, 30 on A2, 60 on X1, and 137 on X2 (the remaining 148 are on scaffolds that have yet to be assigned to chromosomes). Telomeric scaffold X2.16 has the greatest number of copies (69) and the greatest SSGP-71 gene density (29.5 copies/Mb). Another telomeric scaffold (X1.1) had both the second-largest copy number (31) and the second-largest gene density (7.7 copies/Mb).

Protein alignments combined with structural analyses [35] indicate that SSGP-71 mature proteins contain a cyclin-like F box domain near the N terminus and a series of leucine-rich repeats (LRRs) (Figures 3 and S7). Variability among SSGP-71 members results from amino acid substitutions, insertions and deletions, the presence or absence of the F box, and the number of LRRs. We found 419 SSGP-71 genes encoding an SP and 357 encoding an F box. A majority of SSGP-71 genes (293) encode an SP, an F box, and 13 LRRs. Not surprisingly, greater sequence similarity was typically observed among copies in close proximity; for example, 23 of the 31 SSGP-71 copies on scaffold X1.1 lack an F box.

F box domains are commonly associated with LRRs, and both domains mediate protein-protein interactions in a variety of contexts [37]. *Drosophila* and mosquito genomes

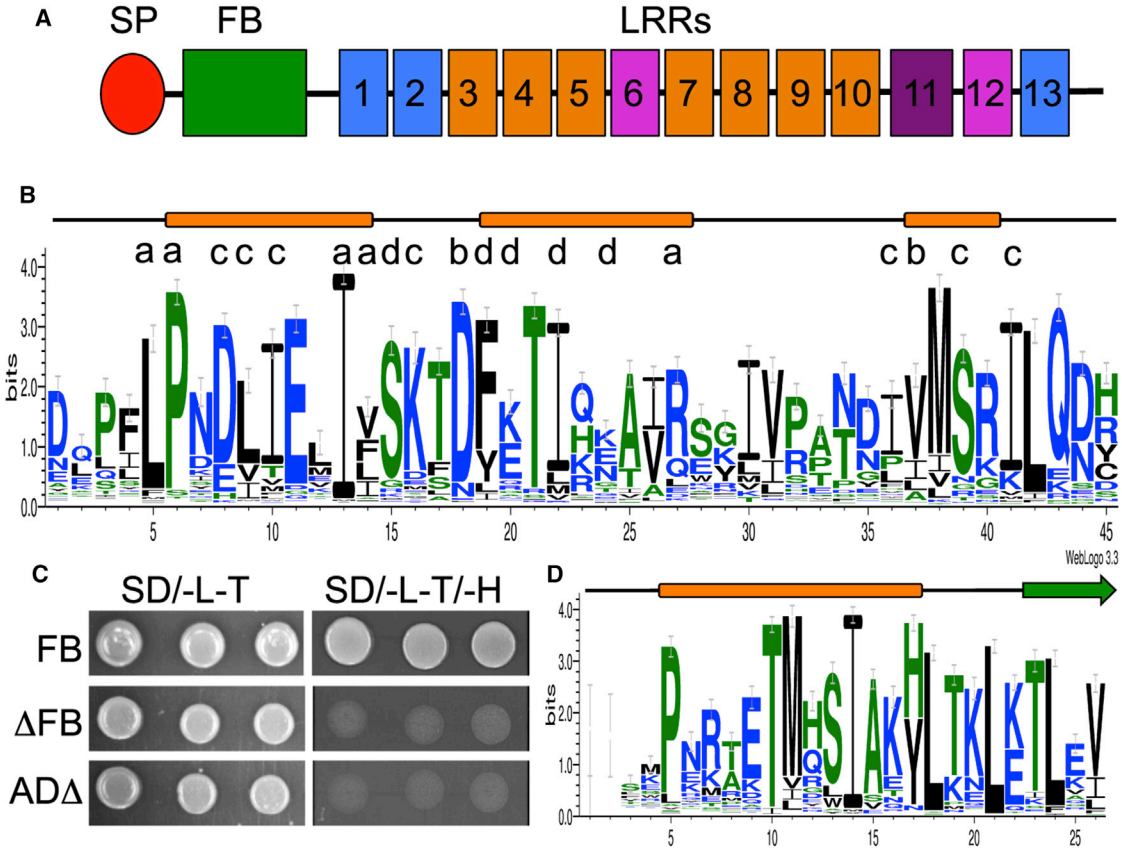

Figure 3. Predicted Structures of Proteins Encoded by SSGP-71 Family Genes

(A) Organization of the typical SSGP-71 protein (SP, signal peptide; FB, F box; LRR, leucine-rich repeat). The LRRs are colored according to the sequence of L residues: blue, LLxxLxxxLxxLxL; orange, Lx(1–3)LxxLxxLxL; pink, Lx(3–5)LxxxLxxxxxLxxLxL; purple, LxxLx(2–3)LxxL (where "L" indicates leucine, valine, isoleucine, phenylalanine, tyrosine, and methionine and "x" indicates any amino acid).

(B) Logo showing the consensus sequence of the SSGP-71 F box based on the alignment of 270 predicted proteins. Above the logo, correspondence of the consensus residue with conserved F box amino acid residues is indicated, where "a" indicates >40%, "b" indicates 20%–40%, "c" indicates 15%–20%, and "d" indicates 10%–15% conservation [36].

(C) Yeast two-hybrid interaction between Mdes005073-RA constructs (bait) and wheat SKP6 (TaSKP6) on nonselective (SD/-L-T) and selective (Sd/-L/-T/-H) media. SSGP-71-142 (Mdes013725-RA) construct containing the F box (FB) had a positive interaction with TaSKP6. A negative interaction was observed when the F box was removed (ΔFB) and in the absence of TaSKP (ADΔ).

(D) Logo showing the consensus sequence of the seventh LLR based on the alignment of 323 predicted proteins. Predicted secondary structures above the F box and LRR consensus sequences are indicated, where the orange bars represent alpha helices and the green arrow represents a beta sheet.

See also Figure S6 and Table S1.

contain ~25 F-box-LRR-encoding genes [37]. None of these encode an SP. Plant genomes contain over 700 F-box-LRR-encoding genes [38], but these also lack SPs. In plants, the proteins facilitate the transfer of ubiquitin to target proteins destined for degradation in the proteasome. They play essential roles in hormonal signaling, plant development, and plant immunity. The F box interacts with Skp1-like proteins, a component of the Skp-Cullin-F-box-E3-ubiquitin-RING-ligase complex that targets proteins for degradation. The LRR domain gives the complex target specificity. As with the F-box-LRR E3-ligase-mimicking effectors that have evolved in bacterial plant pathogens [3, 39, 40], we suspect that SSGP-71 proteins are a novel class of F-box-LRR mimics that enable the insect to hijack the plant proteasome in order to directly produce nutritive tissue and additionally defeat basal plant immunity and stunt plant growth. In support of this hypothesis, we note that first-instar larval feeding increases both Skp and ubiquitin expression in susceptible wheat [41] and that a direct protein-protein interaction between F-box-containing SSGP-71 proteins and wheat Skp proteins is observed in vivo (Figures 3C and S7).

M. destructor Virulence Mutations Are Associated with SSGP-71 Genes

Effectors evolve to give the parasite an advantage over its host. However, effectors can also expose the parasite to plant defense mechanisms [4]. Plant resistance (*R*) genes encode proteins that survey plant cells for the presence of specific effectors. These proteins elicit effector-triggered immunity (ETI) when their cognate effector is perceived. Thus, when the parasite attacks a plant carrying an *R* gene, the cognate effector is a liability, and selection favors the effector's loss or modification. Identifying the mutations associated with these modifications is the classical approach to effector discovery. We took this approach to search for further evidence that SSGP-71 proteins are effectors.

Wheat *R* genes named Hessian fly resistance genes 6 and 9 (*H6* and *H9*) elicit ETI that is lethal to "avirulent" *M. destructor*

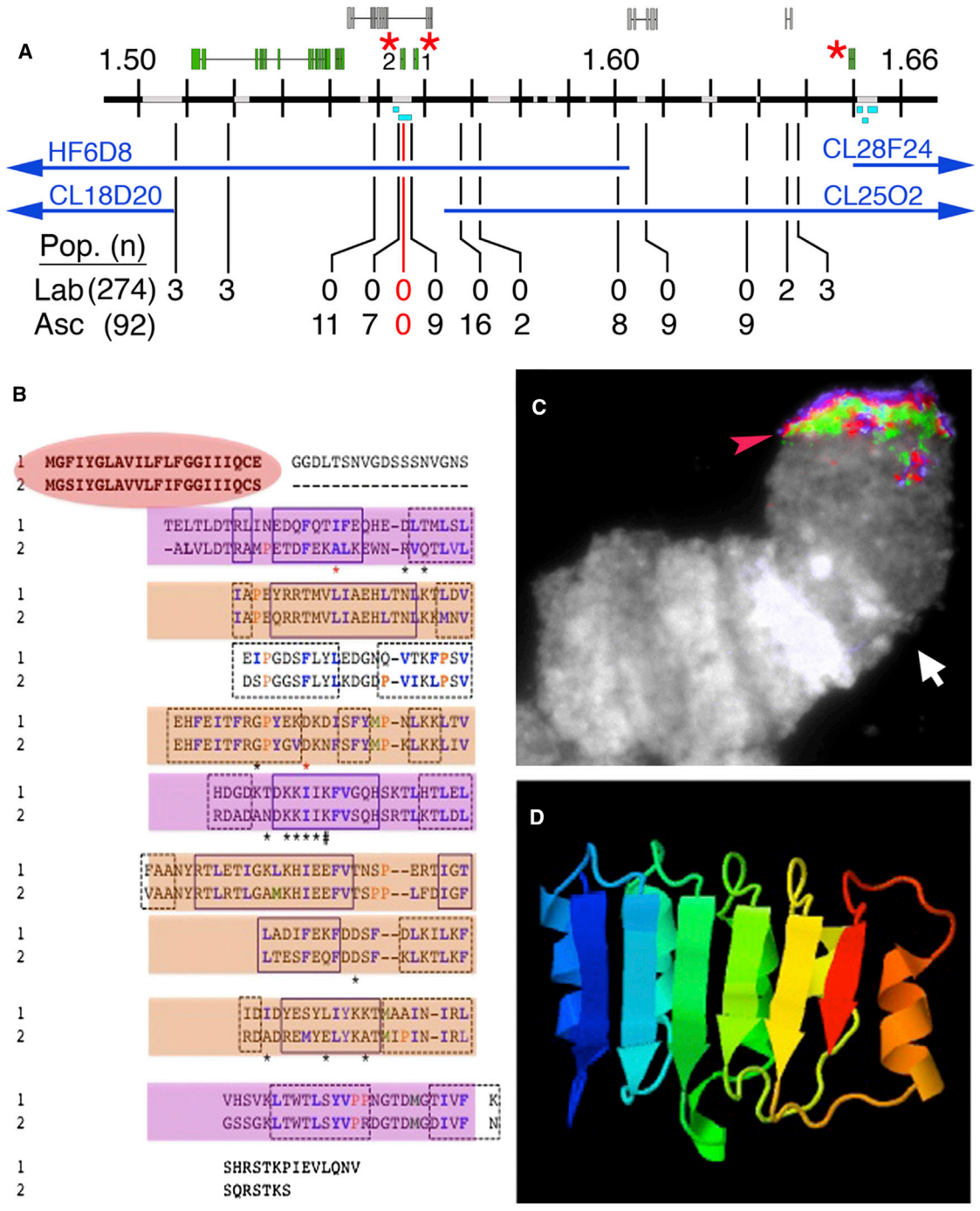

Figure 4. An SSGP-71 Gene Is Responsible for *H9*-Directed ETI

(A) Physical map of a 160-kb segment of the HF genome scaffold X1.1 (black horizontal line; units are Mb). A sequence of HF BAC clones (blue arrows) was used to fill gaps (gray spaces on the black line) in the scaffold sequence. Exons encoded in the sequence are shown as gray (forward strand) and green (reverse strand) boxes above the map. Red asterisks indicate the positions of SSGP-71 genes. Below the map, the number of individuals in structured (Lab) and association mapping (Asc) populations that were recombinant for *H9* virulence and avirulence and polymorphisms within the sequence are shown.
(B) Amino acid alignment of the two *H9* cognate effector candidate SSGP-71 genes (1 and 2; Mdes001118-RA and Mdes015365-RA) showing their SPs (red oval), the alignment of nine LRRs (orange and pink boxes colored as in Figure 3A), and the positions of predicted alpha helices (boxed with solid lines) and beta sheets (boxed with dashed lines). *H9*-virulence-associated mutations in *H9* cognate effector candidate 2 (Mdes015365-RA) are indicated with asterisks (base substitutions) and a number sign (frame shift). *H9*-virulence-associated mutations were absent in candidate 1.
(C) FISH showing the positions of three X1.1 bacterial artificial chromosomes (blue fluorescence, HF13J3; red fuorescence, CL18D20; green fluorescence, CL19E22) near the telomere on the short arm of chromosome X1. The position of the centromere (white arrow) is indicated.
(D) Predicted (Phyre2) three-dimensional conformation of SSGP-71 gene 2 showing the helices and sheets characteristic of LRR proteins.

larvae. "Virulent" genotypes that defeat *H6*- and *H9*-directed ETI have recessive mutations in different effector-encoding genes. These mutations allow the insects to survive on otherwise resistant plants. We found that mutations associated with *H6* virulence in four independent structured mapping populations reside within a 300-kb fragment on scaffold X2.12 (Figure S7). Two SSGP-71 genes, both encoding an F box and 13 LRRs, are the only SP-encoding genes residing within this segment. The expression of one of the two genes (Mdes009086-RA) is lost in *H6*-virulent larvae, indicating that Mdes009086-RA is the *H6* cognate effector.

Mapping the mutations associated with *H9* virulence provided further evidence that SSGP-71 genes are effectors. Using two independent structured mapping populations and four independent field populations, we mapped *H9* virulence within a 20-kb segment on scaffold X1.1 (Figure 4). The only genes within this segment are SSGP-71 genes (*H9* cognate effector candidates 1 and 2). Like some bacterial E3-ligase mimics [15], neither candidate encodes an F box. A frame-shift mutation creating a putative null allele within candidate 2 (Mdes015365-RA) was perfectly associated with *H9* virulence in laboratory and field populations (Figure 4), making Mdes015365-RA the best candidate *H9* cognate effector. Because more than 30 *R* genes elicit ETI against *M. destructor* in wheat [8], we expect that wheat uses additional *R* genes that guard against SSGP-71 effectors. Interestingly, wheat cultivar R gene resistance in agricultural mono-culture tends to last only 5 to 10 years. We speculate that *M. destructor* populations maintain a large reservoir of SSGP-71 effectors to enable redundancy such that any given SSGP-71 gene contains null alleles at high enough minor allele frequencies to avoid *R*-gene-trigged immunity for a small fraction of the population. This enables the *M. destructor* population to rapidly overcome any cultivar-specific resistant gene.

In conclusion, our analysis of the *M. destructor* genome showed a number of insect-related genic adaptations to life as a plant parasite. Most prominently, we identified a large reservoir of potential effector proteins that have expanded due to the arms race between gall midge and host plant. We also show that two members of the largest gene family within this reservoir are effector proteins by performing genetic mapping and demonstrating an in vivo physical interaction with plant host proteins. Members of this family have remarkable structural similarities to both bacterial effectors and their putative plant archetypes, the E3 ubiquitin ligases. Thus, as others have anticipated [3], the *M. destructor* SSGP-71 gene family provides an impressive example of the kind of molecular convergence that exists among unrelated plant parasitic organisms.

Experimental Procedures

Details of *M. destructor* husbandry, genome sequencing, sequence assembly, gene annotation, orthology analysis, LGT and gene mapping are described in the Supplemental Experimental Procedures.

Accession Numbers

The NCBI BioProject accession number for the *M. destructor* genome assembly reported in this paper is PRJNA45867, and that for the RNA-sequencing transcriptome data used for the gene annotation reported in this paper is PRJNA247377. The NCBI SRA accession numbers for *M. destructor* genomic sequence before and after somatic chromosome elimination reported in this paper are SRX540852 and SRX049257. *M. hordei* adult genomic sequence reported in this paper has NCBI SRA accession number SRX049256.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, seven figures, one table, and gene sequences and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.12.057.

Author Contributions

Data generation: C.Z., L.N.E., T.R.B., and M.E.-B. performed the genetic mapping, E3 ligase analysis, and Hessian fly husbandry. G.M.W., Y.Z., L.-L.P., F.L., L.P., K.P.B., S.L.L., and L.V.N. constructed libraries. The genome sequence was generated using the 454 platform by M.C., L.F., M.H., L.J., V.K., D.S., R.S., D.N., G.O., R.T., T.M., and C.K. F.O., M.J., and Y.H. generated both genomic and transcriptome sequences using the Illumina platform. Genome assembly and associated informatics were performed by J.Q., D.Z., D.K., M.B., K.C.W., and S.R. Work performed at the Baylor College of Medicine Human Genome Sequencing Center was managed by S.N.J., S.E.S., D.M.M., and R.A.G. Data analysis: The agripestBase automated gene annotation was performed by S.C., D.C., T.G., N.H., S.N., V.V., and S.J.B. R.M.W., P.I., and E.M.Z. generated OrthoDB protein orthology analysis. Manual gene annotation and analysis: D.W. and J.H.W. searched for and characterized lateral gene transfers. M.N.A., C.L., and H.M.R. curated and characterized chemosensory proteins. SSSGPs were curated and characterized by H.C., B.F.P., X.L., and M.-S.C. F.H. and C.J.P.G. curated and characterized neuropeptides, protein hormones, and biogenic amines. R.B. and M.F. curated and characterized light-detecting molecules. Transcription factors and repeats were curated and characterized by S.K.B. and N.G. J.T.S., A.J.J., S.S., B.J.S., C.E.W., and R.H.S. curated and characterized antioxidants, metabolism, carboxyl esterases, polyamines, and neuromuscular genes. Proteomic analysis was performed by H.C., J.C.C., and K.X. Finally, M.N.A., M.O.H., R.H.S., J.H.W., E.M.Z., M.-S.C., S.J.B., J.J.S., and S.R. designed and led the project and played major roles in the writing of the manuscript and its publication.

Acknowledgments

Genome sequencing was supported by USDA-NIFA AFRI grant 2008-35302-18816 to S.R. Virulence mutation mapping performed in the J.J.S. lab was funded by USDA-NIFA AFRI grant 2010-03741 to J.S. Work by J.H.W. and D.W. was supported by NSF DEB0821936 and DEB1257053. M.N.A. and C.L. were funded by grants from the Swedish Research Council (V.R.) and the Royal Physiographic Society. R.M.W. was supported by Marie Curie International Outgoing Fellowship PIOF-GA-2011-303312. R.M.W. and P.I. were supported by Swiss National Science Foundation awards 31003A-125350 and 31003A-143936 to E.M.Z. Denise Caldwell and Andrew Katz (Purdue University) provided the photographs shown in Figure 1.

Received: October 30, 2014
Revised: December 7, 2014
Accepted: December 23, 2014
Published: February 5, 2015

References

1. Stone, G.N., and Schönrogge, K. (2003). The adaptive significance of insect gall morphology. Trends Ecol. Evol. *18*, 512–522.
2. Price, P.W. (2005). Adaptive radiation of gall-inducing insects. Basic Appl. Ecol. *6*, 413–421.
3. Mukhtar, M.S., Carvunis, A.R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M.T., et al.; European Union Effectoromics Consortium (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science *333*, 596–601.
4. Jones, J.D.G., and Dangl, J.L. (2006). The plant immune system. Nature *444*, 323–329.
5. Torto-Alalibo, T., Collmer, C.W., Lindeberg, M., Bird, D., Collmer, A., and Tyler, B.M. (2009). Common and contrasting themes in host cell-targeted effectors from bacterial, fungal, oomycete and nematode plant symbionts described using the Gene Ontology. BMC Microbiol. *9* (1), S3.
6. Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R., and Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. Mol. Plant Microbe Interact. *22*, 115–122.
7. Hogenhout, S.A., and Bos, J.I. (2011). Effector proteins that modulate plant—insect interactions. Curr. Opin. Plant Biol. *14*, 422–428.

8. Stuart, J.J., Chen, M.S., Shukle, R., and Harris, M.O. (2012). Gall midges (Hessian flies) as plant pathogens. Annu. Rev. Phytopathol. *50*, 339–357.
9. Aggarwal, R., Subramanyam, S., Zhao, C., Chen, M.S., Harris, M.O., and Stuart, J.J. (2014). Avirulence effector discovery in a plant galling and plant parasitic arthropod, the Hessian fly (*Mayetiola destructor*). PLoS ONE *9*, e100958.
10. Bos, J.I.B., Prince, D., Pitino, M., Maffei, M.E., Win, J., and Hogenhout, S.A. (2010). A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). PLoS Genet. *6*, e1001216.
11. Kämper, J., Kahmann, R., Bölker, M., Ma, L.J., Brefort, T., Saville, B.J., Banuett, F., Kronstad, J.W., Gold, S.E., Müller, O., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature *444*, 97–101.
12. Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., Di Stasio, M., Rössel, N., Mendoza-Mendoza, A., Pester, D., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. Science *330*, 1546–1548.
13. Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., Ver Loren van Themaat, E., Brown, J.K.M., Butcher, S.A., Gurr, S.J., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science *330*, 1543–1546.
14. Harris, M.O., Freeman, T.P., Moore, J.A., Anderson, K.G., Payne, S.A., Anderson, K.M., and Rohfritsch, O. (2010). H-gene-mediated resistance to Hessian fly exhibits features of penetration resistance to fungi. Phytopathology *100*, 279–289.
15. Harris, M.O., Freeman, T.P., Anderson, K.M., Harmon, J.P., Moore, J.A., Payne, S.A., Rohfritsch, O., and Stuart, J.J. (2012). Hessian fly Avirulence gene loss-of-function defeats plant resistance without compromising the larva's ability to induce a gall tissue. Entomol. Exp. Appl. *145*, 238–249.
16. Aggarwal, R., Benatti, T.R., Gill, N., Zhao, C., Chen, M.-S., Fellers, J.P., Schemerhorn, B.J., and Stuart, J.J. (2009). A BAC-based physical map of the Hessian fly genome anchored to polytene chromosomes. BMC Genomics *10*, 293.
17. Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M., and Kriventseva, E.V. (2013). OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. *41*, D358–D365.
18. Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics *12*, 491.
19. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.
20. Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J.W., Lambkin, C., Bertone, M.A., Cassel, B.K., Bayless, K.M., Heimberg, A.M., et al. (2011). Episodic radiations in the fly tree of life. Proc. Natl. Acad. Sci. USA *108*, 5690–5695.
21. McBride, C.S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proc. Natl. Acad. Sci. USA *104*, 4996–5001.
22. Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F., et al. (2010). Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science *330*, 86–88.
23. Andersson, M.N., Videvall, E., Walden, K.K.O., Harris, M.O., Robertson, H.M., and Löfstedt, C. (2014). Sex- and tissue-specific profiles of chemosensory gene expression in a herbivorous gall-inducing fly (Diptera: Cecidomyiidae). BMC Genomics *15*, 501.
24. Andersson, M.N., Haftmann, J., Stuart, J.J., Cambron, S.E., Harris, M.O., Foster, S.P., Franke, S., Francke, W., and Hillbur, Y. (2009). Identification of sex pheromone components of the hessian fly, *Mayetiola destructor*. J. Chem. Ecol. *35*, 81–95.
25. Robertson, H.M., and Wanner, K.W. (2006). The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. Genome Res. *16*, 1395–1403.
26. Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., et al. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc. Natl. Acad. Sci. USA *107*, 12168–12173.
27. Croset, V., Rytz, R., Cummins, S.F., Budd, A., Brawand, D., Kaessmann, H., Gibson, T.J., and Benton, R. (2010). Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. *6*, e1001064.
28. Dunning Hotopp, J.C., Clark, M.E., Oliveira, D.C., Foster, J.M., Fischer, P., Muñoz Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., et al. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science *317*, 1753–1756.
29. Bansal, R., Hulbert, S.H., Reese, J.C., Whitworth, R.J., Stuart, J.J., and Chen, M.-S. (2014). Pyrosequencing reveals the predominance of pseudomonadaceae in gut microbiome of a gall midge. Pathogens *3*, 459–472.
30. Wheeler, D., Redding, A.J., and Werren, J.H. (2013). Characterization of an ancient lepidopteran lateral gene transfer. PLoS ONE *8*, e59262.
31. Degnan, P.H., Yu, Y., Sisneros, N., Wing, R.A., and Moran, N.A. (2009). *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. Proc. Natl. Acad. Sci. USA *106*, 9063–9068.
32. Chen, M.-S., Fellers, J.P., Stuart, J.J., Reese, J.C., and Liu, X. (2004). A group of related cDNAs encoding secreted proteins from Hessian fly [*Mayetiola destructor* (Say)] salivary glands. Insect Mol. Biol. *13*, 101–108.
33. Chen, M.-S., Liu, X., Yang, Z., Zhao, H., Shukle, R.H., Stuart, J.J., and Hulbert, S. (2010). Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge. BMC Evol. Biol. *10*, 296.
34. Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.
35. Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.
36. Kipreos, E.T., and Pagano, M. (2000). The F-box protein family. Genome Biol. *1*, REVIEWS3002.
37. Ho, M.S., Tsai, P.I., and Chien, C.T. (2006). F-box proteins: the key to protein degradation. J. Biomed. Sci. *13*, 181–191.
38. Vierstra, R.D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. Nat. Rev. Mol. Cell Biol. *10*, 385–397.
39. Hicks, S.W., and Galán, J.E. (2010). Hijacking the host ubiquitin pathway: structural strategies of bacterial E3 ubiquitin ligases. Curr. Opin. Microbiol. *13*, 41–46.
40. Angot, A., Peeters, N., Lechner, E., Vailleau, F., Baud, C., Gentzbittel, L., Sartorel, E., Genschik, P., Boucher, C., and Genin, S. (2006). *Ralstonia solanacearum* requires F-box-like domain-containing type III effectors to promote disease on several host plants. Proc. Natl. Acad. Sci. USA *103*, 14620–14625.
41. Liu, X., Bai, J., Huang, L., Zhu, L., Liu, X., Weng, N., Reese, J.C., Harris, M., Stuart, J.J., and Chen, M.-S. (2007). Gene expression of different wheat genotypes during attack by virulent and avirulent Hessian fly (*Mayetiola destructor*) larvae. J. Chem. Ecol. *33*, 2171–2194.