A Horizontal Gene Transfer Led to the Acquisition of a Fructan Metabolic Pathway in a Gall Midge

Xiaoyan Cheng, Sandra Garcés-Carrera, Robert Jeff Whitworth, John P. Fellers, Yoonseong Park,* and Ming-Shun Chen*

Animals are thought to use only glucose polymers (glycogen) as energy reserve, whereas both glucose (starch) and fructose polymers (fructans) are used by microbes and plants. Here, it is reported that the gall midge *Mayetiola destructor*, and likely other herbivorous animal species, gained the ability to utilize dietary fructans directly as storage polysaccharides by a single horizontal gene transfer (HGT) of bacterial levansucrase/inulinase gene followed by gene expansion and differentiation. Multiple genes encoding levansucrases/inulinases have their origin in a single HGT event from a bacterium and they show high expression levels and enzymatic activities in different tissues of the gall midge, including nondigestive fat bodies and eggs, both of which contained significant amounts of fructans. This study provides evidence that animals can also use fructans as energy reserve by incorporating bacterial genes in their genomes.

1. Introduction

The gall midge *Mayetiola destructor*, also called the Hessian fly, is a specialist that inhabits in a single host plant, wheat. Hessian fly genomics has uncovered multiple evolutionary innovations for the adaptation of this species to its fixed habitat, including the contribution of nearly 10% of its genes to the production of effector-like proteins for host manipulation and an unconventional genetic mechanism for the generation of mutations in these effector-like genes.[1–5] The production and secretion of effectors in host tissues allow a single small (0.6 mm long) larva to completely take control of a susceptible wheat plant by reprogramming the metabolic pathways of the host plant,[6,7] inducing the formation of nutritive cells at the feeding site,[8] inhibiting wheat growth irreversibly,[9] and suppressing host defenses.[10–12]

In addition to gaining the ability to manipulate a host plant, the efficient usage of nutritional components in this host plant might have been a challenge in the evolution of this specialist. Wheat vegetative tissues such as wheat leaf sheaths in young seedlings are relatively poor in accessible nutrients for insects.[13] In young seedlings, most carbohydrates are inaccessible to many animal species because they are incorporated into structural components, such as the cell wall, to support plant growth. Among the accessible nutrients in wheat vegetative tissues, nonstructural fructans (levan and inulin) are the most abundant carbohydrates and are used by plants as reserves for later development and/or for stress responses.[14–16] Hessian fly larvae reside in the basal part of wheat seedlings between two leaf sheaths until pupation.[17] For rapid population expansion, multiple (up to 100) larvae are usually crowded in a single seedling in wheat fields. The efficient utilization of host fructans can provide selective advantages in the evolution of such an unusual specialist insect.

Horizontal gene transfer (HGT) is defined as the movement of genetic materials between unrelated species that subsequently undergoes vertical transmission within the species. HGT is one of the major genetic mechanisms for species to adapt to specific living habitats. For herbivores, HGT is one of the mechanisms by which organisms gain additional abilities to digest otherwise indigestible compounds from host plants.[18] For instance, an HGT-derived invertase allows the nematode *Globodera pallida* to digest sucrose.[19] An HGT-derived mannanase in the coffee beetle *Hypothenemus hampei* confers the ability to digest galactomannan to this insect.[20] Beetles with HGT-derived xylanases can digest cell wall polysaccharides.[21] A unique HGT-derived *Mayetiola destructor* levansucrase (MDL) gene has been revealed in the Hessian fly genome.[22] A more systematic analysis revealed that a HGT was followed by gene expansions for at least ten copies of *MDL* genes in this case. We investigated potential differentiations of the multiple copies of *MDL* genes, and found that the Hessian fly *MDL* genes not only function in the digestion of fructans from host plants but also utilize fructans that...
are directly assimilated into their own endogenous metabolic system as storage carbohydrates. Therefore, we propose that expansion of the horizontally transferred levansucrase genes in the Hessian fly resulted in a shift in the endogenous carbohydrate metabolic pathway for utilizing fructans.

2. Results

2.1. Evolution of the Hessian Fly MDL Genes by HGT Followed by Gene Duplication

The gene categorized as bacterial glycose hidrolase family 32 (EC 3.2.1 GH32 or β-fructofuranosidasase) in the Hessian fly was initially discovered in the transcriptomes of the gut and salivary glands.[8,28] Subsequent searches against the Hessian fly genome assembly[29] identified a number of genes with high similarities to these genes likely play critical roles in Hessian fly physiology.[27] The phylogeny of the MDLs shows the recently duplicated pairs of MDL-1 and 3, and 2 and 10, while pseudogenizations may have occurred, which was indicated by the presence of five truncated MDL-like genes (HF-1 to 5 in Figure 1A). Overall, there is ample evidence to support the hypothesis that the MDL genes are duplicates from a single HGT-derived gene and provide the Hessian fly with an adaptive advantage.

In a phylogenetic analysis, GH32 found in arthropods and a nematode were mainly divided into two groups, tentatively
named GH32A and GH32B. The GH32A group includes enzymes that have been biochemically characterized as having levanase/inulinase activities,[28,29] while the GH32B group includes enzymes described as having sucrase activity. [30] HGT of bacterial GH32 has been previously described in Bomby mori[31] and in Sphenophorus levis.[32] This atypical adaptive evolution for the efficient digestion of plant-specific fructans appears to have occurred in multiple species in a punctuated manner, except for the sucrase (GH32B) HGT in Lepidoptera, which apparently occurred in an early evolutionary lineage of Lepidoptera and subsequently was maintained in multiple lepidopteran species (Figure 1C, and Figure S2, Supporting Information). The taxonomically punctuated distribution of levanase/inulinase (GH32A) in at least ten species of animals also supports multiple HGTs in the evolutionary history. GH32A HGTs have occurred independently in Arachnida (mites and spider species), Collembola, Hemiptera, Lepidoptera, and Tardigrada species. Notably, hemipterans with GH32A HGTs include the notorious pest species white fly (Bemisia tabaci) and western flower thrips (Frankliniella occidentalis). Expansion of the MDL gene family in Hessian fly is with a unique pattern among HGT of GH32. We investigated the expanded MDL genes to test the possible functional differentiation in addition to the digestive function of the MDL.

2.2. Expressions Patterns of the MDL Genes in Digestive and Nondigestive Tissues

Spatiotemporal transcript levels of the MDL genes indicated that the highest expression level occurred in the 1st- and 2nd-instar larvae, the most active feeding stages, with some

![Figure 2](https://www.advancedsciencenews.com)
exceptions (Figure 2A). Similarly, nearly all examined MDL transcripts were at the highest levels in the tissues involved in food digestion, i.e., midgut and salivary glands, in the 1st-and 2nd-instar feeding stages (Figure 2B).

However, a number of MDL transcripts were also present at significantly high levels in nonfeeding stages. For example, MDL-4 and MDL-7 were expressed at high levels in the 3rd instar (nonfeeding prepupal stage) and pupal stages; MDL-8, the only copy lacking a secretion signal peptide, showed high transcript levels in the pupal stage; and MDL-1 was generally high in every stage, including eggs, pupae, and female adults. The MDL transcript level in female adults was similar to the level in eggs, implying maternal deposition of the mRNA into eggs. Likewise, MDL transcripts were present at significantly high levels in the tissues that were not directly involved in digestion: MDL-1 and MDL-2 transcripts were present at high levels in the Malpighian tubes, and transcript levels of MDL-3, 4, 6, 7, and 9 were moderate in fat bodies (Figure 2B). The significant levels of MDL transcripts in nondigestive tissues and nonfeeding stages suggest that these genes have evolved to play additional roles in the endogenous physiology of the Hessian fly.

The immunohistochemistry (IHC) data were also consistent with the transcript levels measured by RT-PCR for a number of selected MDL proteins. Affinity-purified antibodies showed specific activities toward the corresponding antigens via Western blotting (Figure S3B, Supporting Information). The results for MDL-1, 4, and 7 were generally consistent with the transcript levels in different tissues (Figure 2C), namely, the lack of staining for MDL-1 in the salivary glands and fat bodies and the ubiquitous presence of MDL-4 and MDL-7. Preimmune sera did not stain any tissues at the same gain level in confocal microscopy (Figure S3C, Supporting Information).

2.3. Fructans and MDL Enzyme Activities in Both Wheat Plants and the Hessian Fly

To determine if fructans (levan and inulin) were abundant in wheat seedlings and contributed nutritionally to Hessian fly larvae, the amounts of levan and inulin at the feeding site of wheat tissues were measured. Wheat plants contained significant levels of levan and inulin (Figure 3A). Surprisingly, the wheat infested with Hessian fly larvae had significantly higher levels of levan and inulin (>2×) at the feeding site than the corresponding tissues without Hessian fly infestation or tissues of infested plants distant from the feeding site, indicating that Hessian fly infestation either induced the production of levan and inulin or altered the composition of polysaccharides at the feeding site.

The presence of significant amounts of fructans in the nondigestive tissues of Hessian fly worms was confirmed and compared to that in other insect species, namely, the fruit fly Drosophila melanogaster (FF) and the tobacco hornworm Manduca sexta (TH) (Figure 3A). Hessian fly fat bodies and eggs contained significantly higher amounts of levan and inulin than the tobacco hornworm fat bodies. The levels of levan and inulin were essentially undetectable in fruit flies (Figure 3A).

We tested whether the Hessian fly has levanase, inulinase, and sucrase activities in different stages. The hydrolysis of both levan and inulin by levanases and inulinases yields difructose, which can be converted into fructose to enter fructolysis. We found significant levels of levanase and inulinase activities in every developmental stage, including nonfeeding 3rd instars, pupae, adults, and eggs (Figure 3B). Although the levels of enzymatic activity in the pupae and males were low, there were significant levels of enzymatic activity in female adults and eggs. In fact, the tissues from female adults and eggs possessed almost identical, significantly high levels of enzymatic activities. The enzymatic activity was also tested in the nondigestive tissues of feeding larvae of the Hessian fly and other insect species. Hessian fly fat bodies and eggs showed high levels of activity, while fat bodies from the fruit fly and tobacco hornworm showed mainly sucrase activities (Figure 3C).

Overall, we found high levels of fructans in the host wheat and even higher levels in the infested wheat at the feeding site than in the uninfested or nonfeeding tissue of wheat. Levanase, inulinase, and sucrase activities were at peak levels in the feeding stages, while the nonfeeding stages (eggs and females) also showed significant levels of the enzymatic activities. Hessian fly eggs and fat bodies contained high levels of fructans.

2.4. Recombinant MDLs Possess Levanase, Inulinase, and Sucrase Activities

We tested whether the MDL genes encode functional levases and/or inulinas. The bacterial production of recombinant MDLs (rMDLs) and purification using the His-tag were successful, as determined by detection of the expected protein size via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Figure S3B, Supporting Information) with two exceptions, namely, MDL-2 and MDL-8, for which the production of recombinant proteins was unsuccessful. Assays with the three substrates levan, inulin, and sucrase revealed that all eight rMDLs showed inulinase activity, and all rMDLs except rMDL-3 showed levanase activity (Figure 4A). A strong substrate specificity was observed in rMDL-3, which exhibited activity against only inulin. rMDL-5 and 10 showed activities toward levan and inulin but not sucrase. Other rMDLs had lower sucrase activity than levanase and inulinase activities (Figure 4A,B). The optimal pH and temperature for enzymatic activities ranged from 5 to 9 for pH and 20 to 37 °C, respectively.

3. Discussion

Acquisition and functionalization of HGT-derived MDL genes in the Hessian fly conferred the ability to efficiently use host fructans for nutrition. In addition, we propose that the Hessian fly adapted the carbohydrate metabolic pathway for utilizing fructans as its own storage carbohydrates for endogenous energy reserves. This innovative evolutionary path started with the HGT of a levanase/inulinase from a bacterium (likely Bacillus sp.) followed by gene family expansion and diversification. The adoption of the fructan metabolic pathway by the recipient of the HGT is likely to have occurred by sequential evolutionary steps of multiple duplications followed by diversification in multiple MDL genes.
HGT allows the exchange of genetic materials between reproductively isolated organisms, and has been recognized as a powerful driving force in adaptive evolution, although the understanding of its impact on eukaryotic evolution is still at an early stage. The most successful eukaryotic phylum of arthropods has been shown for a number of interesting cases of HGT of bacterial and fungal genes, leading to the acquisition of a new ability for the detoxication of phytochemicals from host.
One of the expected sizes. Reducing sugars produced by recombinant proteins were verified on an SDS/PAGE gel (Figure S3B, Supporting Information). The expressed recombinant proteins for Figure S3A (Supporting Information). We were unable to obtain sufficient recombinant proteins for MDL-2 and MDL-8 for enzymatic activity assays. Purified recombinant proteins were verified on an SDS/PAGE gel (Figure S3B, Supporting Information). The expressed recombinant proteins were of the expected sizes. Reducing sugars produced by recombinant protein digestion were determined by the DNS method.[33] The enzymatic activities of recombinant MDLs (rMDLs) in vitro. A) Enzymatic activities of rMDLs toward levan, inulin, and sucrose. Blue, yellow, and red bars represent levanase, inulinsase, and sucrase activities, respectively. Enzymatic activities were determined based on the amounts of fructose generated by the corresponding recombinant enzyme with the unit per mg protein. Data are shown as the mean ± standard deviation (SD) (n = 3). Recombinant proteins with full-length cDNAs excluding the signal peptide-coding region were expressed in the Escherichia coli strain Rosetta with constructs based on the expression vector pET28a(+) (Novagen). Detailed procedures for making constructs are given in the Experimental Section. The structures of each construct for MDL-1, MDL-3, MDL-4, MDL-5, MDL-6, MDL-7, MDL-9, and MDL-10 are shown in Figure S3A (Supporting Information). We were unable to obtain sufficient recombinant proteins for MDL-2 and MDL-8 for enzymatic activity assays. Purified recombinant proteins were verified on an SDS/PAGE gel (Figure S3B, Supporting Information). The expressed recombinant proteins were of the expected sizes. Reducing sugars produced by recombinant protein digestion were determined by the DNS method.[33] One unit of enzyme activity was defined as the amount of enzyme required to release 1 μmol of fructose min⁻¹ under the assay conditions. B) Kinetic parameters of the enzymatic activity of each recombinant protein toward levan, inulin, and sucrose. The enzymatic activities of the recombinant proteins were also determined under different pH and temperature conditions. The parameters included K_M (Michaelis constant) and V_max, which were calculated by a Lineweaver–Burk plot. All parameter data are listed in Table S3 (Supporting Information), presented as the mean ± SD (n = 3). In this panel, mean data were used to generate the figure. The tested pH range was from 3.0 to 12.0, and the temperature range was from 4 to 60 °C. Ten MDL proteins were separated into different clades with the same analysis as Figure 1A.

Enzymatic activities for rMDL-1 under neutral pH suggest roles in energy production and the locations of introns and signal peptides (Figure 1C): the branch a (four genes in Leptotrombidium delense) is with no introns and no signal peptide, the branch b (two genes in Tetranichus urticae) has one intron at a location and phase that differs from those in MDL, the branch c (two genes in Orchesella cincta) is with no introns and no signal peptide, and the branch d (two genes in Folsomia candida) has five introns in different locations and phases.

The MDLs and multiple GH32A in other arachnid species in this cluster could have shared the same origin from a single HGT that was followed by independent acquisitions of the introns and signal peptides in different lineages and independent losses of the GH32A genes in majorities of branches in taxonomically punctuated groups. Alternatively, the same observation could be interpreted by multiple HGTs from related bacterial taxa in only a number of arthropod species. In this case, lack of closely related bacterial GH32A to this cluster could be a result of an incomplete sampling of the bacterial genes. Finally, but with very low probability, GH32A in arthropods may have undergone a convergence evolution in the arthropod digestive system, which may offer optimal enzymatic activity in the arthropod digestive systems. More extensive sampling with genome sequences in the future will provide information for understanding the evolutionary origin of the MDLs.

The acquisition of the fructan metabolic pathway in the Hessian fly is supported by the expression patterns of the MDL genes (Figure 2) with levanase/inulinsase activities (Figure 3B) and high fructan contents (Figure 3A) in nondigestive organs, including the fat bodies and eggs of Hessian flies. The rMDLs showed significantly different enzymatic properties (Figure 4), along with different expression patterns in different stages and tissues (Figure 2). MDL-8 uniquely lacks a secretion signal peptide and is basally located in the phylogeny of the MDLs (Figure 1), implying that MDL-8 might be the ancestral copy after the acquisition of the intron sequences but before the evolution of the secretion signal peptide, although an alternative scenario could be specific loss of secretion signal peptide in the MDL-8. MDL-7 showed constitutively moderate to high transcript levels in different stages and tissues (Figure 2A,B). The optimal conditions for rMDL7 was pH 5 and 20 °C, respectively, for all three substrates. The data imply that rMDL-7 is likely to play multiple roles including extraoral digestion with the optimal pH of 5, which is close to that of wheat at the feeding site (Figure S5, Supporting Information). MDL-1 showed high transcript levels in eggs with optimal enzymatic conditions, such as pH 7 and 37 °C and pH 8 and 30 °C for levanase and inulinsase activities, respectively. The highest activities of rMDL-1 under neutral pH suggest roles in energy production by hydrolysis of endogenous fructans in developing eggs, which exhibit approximately neutral pH values. MDL-6 was mainly expressed in fat bodies, which also exhibited a neutral pH value. Enzymatic activities for rMDL-6 were optimal at pH 7–8, supporting its role in hydrolysis of endogenous fructans in fat bodies. Variations in expression patterns, enzymatic properties, and the evidence of purifying selection on different MDLs support the hypothesis that different MDL homologs have evolved diversified functions in different tissues.

The utilization of fructans as storage sugars requires not only enzymes for fructan hydrolysis but also enzymes for
fructan synthesis in storage organs after the uptake of fructose from diet. Although a search of the draft Hessian fly genome database identified a short DNA contig with high sequence identity to a bacterial levansucrase (*Pseudomonas fluorescens*, Figure S4, Supporting Information), confirmation is needed to exclude the possibility of bacterial contaminating sequences. It is possible that there is another levansucrase gene(s) in the Hessian fly genome, but has not been identified due to large gaps in the current Hessian fly draft genome.\(^{[11]}\) Alternatively, levansucrases in symbionts may function in the production of storage fructans inside the Hessian fly. It is also possible, but very unlikely, that dietary fructans are transported from the gut to other tissues directly. The Hessian fly is known to host various symbiotic bacteria in nondigestive tissues.\(^{[13,44]}\) A number of other herbivorous arthropods acquired levane/ inulase genes similarly, but without extensive gene expansion yet. Western flower thrips (*Hemithira in Figure 1C*) and mite species also have similar levane/inulase genes. The Western flower thrips is known to contain symbionts with the levan-sucrase gene,\(^{[45]}\) implying that the fructan metabolic pathways could be acquired by those species via multiple independent HGT events.

The use of fructans as storage carbohydrates in the Hessian fly instead of or in addition to glycogen, the typical energy reserve in animals, may provide a selective advantage to the Hessian fly, specialized in the sole host wheat plant. Hessian fly larvae feed on wheat plants that are rich in fructans (Figure 3A). The fructan diet provides abundant fructose that can be directly polymerized and stored without expensive metabolic conversion of fructose to glucose-6-phosphate and glycogen. Therefore, we propose that a single HGT event followed by gene expansions and differentiation resulted in the acquisition of the fructan metabolic pathway for the endogenous storage carbohydrate in the adaptive evolution of the Hessian fly, which has specially adapted to its sole preferred host wheat.

4. Experimental Section

**Insect Rearing and Sample Collection:** Hessian fly used in this study was biotype GP, which was derived from a field collection from Scott Country, Kansas in 2005.\(^{[46]}\) Stock of the insect has been continuously maintained on seedlings of the susceptible wheat cultivar “Karl 92” in a greenhouse. Hessian fly samples were collected throughout the whole life cycle including 1st-instar larvae, 2nd-instar larvae, 3rd-instar larvae, pupae, male adults, and female adults. Where larvae were about to hatch, plants were dissected every 2 h to check if larvae had reached the feeding site. When the first larva was found, that time was set as zero for time course collection. Insect samples were collected on day 1, 3, 5, 7, and 12 d and the samples were designated 1\(^{st}\)-instar (day 1), 1\(^{st}\)-3\(^{rd}\) (day 3), 2\(^{nd}\)-instar (day 5), 3\(^{rd}\)-instar (day 7), and pupae (day 12), respectively. Insects were harvested by exposing dissected plants to water. Hessian fly larvae on the wheat leaf sheath fell into the water during exposure. After collection, water in a microcentrifuge tube was removed and Hessian fly larvae or pupae were frozen in liquid nitrogen immediately and stored at \(-80^\circ\)C until use. Male and female adults were collected randomly from a stock, in which flies were just emerging. Three independent biological replicates for each sample were collected.

**Insect Tissue Dissection:** For dissection of larval tissues, 3 and 5 d old larvae were collected separately for dissection, and the collected tissue samples were designated 1\(^{st}\)-instar and 2\(^{nd}\)-instar, respectively. For tissue dissection, individual larvae were placed in a drop of deionized water on a petri dish. Dissection was carried out under a dissecting microscope (Leica M165C, Leica Microsystems, Buffalo Grove, IL). Larval tissues were pulled out of the insect body with two pairs of forceps (Dumont \#5), with one pair of forceps holding the posterior tip of a larva and the other pair of forceps holding the forehead. When the two pairs of forceps pull away from each other, the larval exoskeleton was broken and the tissues inside flowed out into the water. Individual tissues including gut, salivary glands, and Malpighian tubules were separated with a dissecting pin. Individual tissues were then transferred to different tubes using forceps. Fat bodies remained inside the carcass and were squeezed out with two pairs of forceps into collecting tubes directly. Extra liquid that leaked out along with fat bodies was removed with a syringe. Each tissue sample contained pooled collections from \( \approx 300 \) larvae. The tissues were frozen immediately in liquid nitrogen and stored at \(-80^\circ\)C until use. Three independent biological replicates were collected for each tissue sample.

To collect predeposited eggs, freshly emerged female adults were caught with an aspirator. After immobilizing the females by chilling at \(-20^\circ\)C for 5 min, the females were dissected in phosphate-buffered saline (PBS) buffer. Long oval-shaped eggs are reddish and could be easily distinguishable from other tissues. The eggs along with other ovaries were transferred to a microtuge tube with distilled water. After a brief centrifuge, eggs were precipitated to the bottom whereas other ovary tissues remained on the top of the solution. After removing other ovary tissues and excess water, the eggs were immediately frozen in liquid nitrogen for further use.

To collect a large quantity of postdeposited eggs, \( \approx 5000 \) female adults were collected and caged onto an agar plate in a 10 cm petri dish. Females were forced to deposit eggs onto the surface of the agar plate. The plate with females was kept in a growth chamber maintained at 20 °C for 24 h. Eggs on the plate were washed with distilled water and frozen in liquid nitrogen for further processing. For each sample, three independent biological replicates were collected.

**Wheat Cultivation, Infestation, and Tissue Collection:** The wheat cultivar “Karl 92” was used to rear Hessian fly. Twelve wheat seeds were planted in each 10-cm-diameter pot filled with PRO-MIX “BX” potting mix (Hummert Inc., Earth City, MO). The pots were maintained in a growth chamber (AR 66L, Percival) programmed at 20:18 °C (L:D) with a photoperiod of 14:10 (L:D) h. After 7 d when seedlings were at the 1.5-leaf stage (one full grown leaf and the second leaf just emerged), the seedlings were infested with female adults (two plants per female) confined within a screened cage. When larvae were first observed at the feeding site, it was counted as time 0 for the collection time course. At a specific time point, wheat seedlings were dug out, washed clean, and dissected by peeling off the first leaf sheath. The 10 mm region of the second leaf sheath next to the base is the feeding site where multiple Hessian fly larvae were colonized. The 10 mm region of the feeding site was exposed to water to remove any Hessian fly larvae on it. After removing extra water with a Kimwipe, the 10 mm wheat tissue region at the feeding site was collected. For each sample, ten leaf sheaths were collected and combined. The pooled samples were frozen in liquid nitrogen and stored at \(-80^\circ\)C until use.

**DNA and RNA Extraction:** Total insect DNA was extracted from larvae with a Quick-DNA Tissue/ Insect Kit (Zymo Research Corp.) following the manufacturer’s instructions. Quantity and quality of insect DNA were determined on a NanoDrop 2000 (Thermo Scientific).

Total RNA was extracted from insects with the TRIzol Reagent (Invitrogen) following the manufacturer’s instructions. RNase-free DNase I (Qiagen) was used to remove potential genomic DNA contamination in RNA samples. The integrity and quantity of total RNA were determined on an Agilent 2200 TapeStation System (Agilent).

**Reverse Transcription and cDNA Cloning:** First-strand cdNA was synthesized using oligo-dT as a primer with a SuperScript III First-Strand Synthesis Kit (Invitrogen) following the user’s manual. Primers were synthesized based on genomic sequences in the Hessian fly draft genome sequence database. PCR was carried out with 2 µL of a 1/50 (V/V) dilution of the first-strand cdNA, 5 µL of each primer, 25 µL Q5 High-Fidelity 2× Master Mix (New England Biolabs), and 22 µL distilled water in a total volume of 50 µL. All amplification reactions were carried out on
a PTC-200 thermal cycler (Bio-Rad) under the following conditions: 98 °C for 1 min, followed by 28 cycles at 98 °C for 10 s, 55 °C for 30 s, and 72 °C for 45 s, with a final extension step at 72 °C for 2 min. PCR products were separated by electrophoresis on 1.2% agarose gels. DNA bands were stained with ethidium bromide and visualized on a UV-transilluminator. DNA bands with the right estimated sizes were cut from the agarose gel and purified with the QIAquick gel Extraction Kit (Qiagen) following the manufacturer’s protocol. The purified products were ligated into the pMiniT cloning vector (New England Biolabs). The ligated mixture was transformed into chemically treated XL1-Blue competent cells (Fisher Scientific). Insert-positive clones were selected based on blue/white color. Plasmid DNA was extracted and sequenced by the DNA Genotyping Facility at Plant Pathology of Kansas State University.

**Primer Design and Synthesis:** All primers used in this research are listed in Table S4 (Supporting Information). Primers for cDNA cloning were manually designed based on genomic DNA fragments in the Hessian fly genome sequence database. Primers for qPCR were designed by the Real-Time qPCR Assay Entry program (https://www.idtdna.com/scitools/Applications/RealTimePCR/Results.aspx). Primers for expression constructs were designed by the program OligoAnalyzer (https://www.idtdna.com/calc/analyzer). All primers were synthesized by Integrated DNA Technologies company (IDT Technologies).

**Real-Time PCR Analysis:** Real-time PCRs were performed on an iCycler system (Bio-Rad). Each reaction was performed with 1 µL of 15 ng µL⁻¹ first-strand cDNA as a template, 3 µL of each primer, 10 µL of iQ SYBR Green super mix (Bio-Rad), and 3 µL distilled deionized water in a total volume of 20 µL. Each reaction was done in triplicate using PCR tube strips with optical flat caps (Bio-Rad). The amplification reaction was carried out under the following steps: reaction mixture was incubated at 95 °C for 2 min to denature template DNA, and was then amplified by 39 cycles of denaturation at 95 °C for 10 s and annealing and extension at 62 °C for 30 s. Melt curve analyses were subsequently done by heating the PCR mixtures from 65 to 95 °C to ensure amplicon specificity and exclude primer–dimer formation. Relative expression levels were determined by comparing two internal references including ribosomal protein S6 kinase beta-2 (Mdes007562, abbreviated RPSK) and ATP-dependent RNA helicase (Mdes014483, abbreviated ARH). These two genes were expressed fairly equally in all stages and tissues of Hessian fly according to previous RNA-seq data. Relative expression levels of MDL family are indicated by Mean ± standard deviation (SD) (n = 3), which are listed in Table S2 (Supporting Information). Each sample contained three biological replicates with three technical replicates for each biological replicate.

**Construct Design and Production of Recombinant Proteins:** The expression vector pET28a (+) (Novagen) was used to make the expression constructs. Primers were designed to cover exactly the mature protein-coding region. The 5’-primers carried nine extra nucleotide residues which contained a BamH I restriction site except for 3′, which contained a Hind III restriction site (Table S2, Supporting Information). The 5′-primers carried nine extra nucleotide residues which contained either a Hind III or Xho I restriction site for cloning. Each of the primer pairs was used for PCR to amplify the mature protein-coding region for each cDNA. After amplification, the PCR products were double-digested with the respective combinations of restriction enzymes and were then ligated to the pET28a (+) vector, which was also double-digested with the same combination of restriction enzymes. After confirmation of correct ligation by sequencing, the resulting expression constructs were transformed into chemically treated competent cells of the E. coli strain Rosetta 2 (DE3) (Novagen). Bacterial cells carrying expression constructs were maintained in LB medium containing kanamycin (50 µg mL⁻¹). For the production of recombinant proteins, bacterial cells carrying each expression construct were grown overnight at 37 °C in 10 mL liquid LB containing 50 µg mL⁻¹ of kanamycin. The overnight culture was then diluted by 1:100 in fresh terrific broth medium (Sigma) with 3% (v/v) ethanol. After dilution, the culture was continuously incubated at 37 °C until OD₆₀₀ reached 0.6 as described previously by Chhetri et al. [48] when IPTG was added into the culture at a concentration of 1 × 10⁻⁴ m L⁻¹ culture. The culture was further incubated at 20 °C overnight to induce the production of the recombinant protein to high levels.

**Purification and Quantification of Recombinant Proteins:** E. coli cells containing a specific recombinant protein were harvested by centrifuging at 8000 rpm for 15 min. The pellet was suspended in 1× PBS. The suspension was sonicated (EpiShear Probe Sonicator) on ice, and the resulting mixture was centrifuged at 12 000 rpm for 15 min at 4 °C. Recombinant proteins stayed in the supernatant were purified by passing through HisPur Ni-NTA Spin columns (Thermo Scientific) following the manufacturer’s instructions. Homogeneity of the purified proteins was determined on SDS/PAGE gels. Protein quantification was carried out by staining with Brilliant Blue G-250 and subsequently measured using a Protein Quantification Kit-Rapid (Cat. No. 51254-1KT, Sigma).

**Analysis of Enzymatic Activity of Purified Recombinant Proteins:** Each recombinant protein was tested against different substrates including levan, inulin, and sucrose. Enzymatic activities of the recombinant proteins were determined as described by Dahech et al. [49] with some modifications. Briefly, different amounts of recombinant proteins as specified in each assay were dissolved into 500 µL solution buffered by 20 × 10⁻³ m phosphate (PBS, pH 6.5) containing 1 mg mL⁻¹ of either levan, inulin, or sucrose as the substrate (Sigma). The assay solution was incubated at 30 °C for 30 min. After that, the reaction was inactivated by boiling for 5 min. Reducing sugars produced by enzymatic activity in the reaction mixture were determined by the 3,5-dinitrosalicylic acid (DNS) method. [33] One unit of enzymatic activity was defined as the amount of enzyme required to release 1 µmol of fructose min⁻¹ under the assay conditions.

**Measurement of Sugar Content in Wheat Seedlings:** Fresh wheat tissue samples were collected from the feeding site (10 mm of the second leaf sheath next to the base) and nonfeeding site (10 mm of the second leaf sheath next to the leaf ligule) of uninfested control and HF-infested wheat seedlings. After the weight was recorded, plant samples were ground into powder, and homogenized with PBS buffer (pH 6.5). The supernatant was collected after centrifuged and denatured in boiling water for 10 min. Levan and inulin standard solutions were prepared at 0, 5, 10, 15, 20 mg mL⁻¹. Each supernatant was divided into two equal aliquots. The first aliquot was added with 3% (V/V) trichloroacetic acid (TCA, Sigma) at 55 °C for 1 h to hydrolyze levan only, releasing fructose, but not inulin. [51] The levan content in the wheat was calculated based on the standard curve. Another aliquot was added with inulinase (10 U mL⁻¹, Sigma) at 37 °C for 1 h, which digests both levan and inulin, releasing reducing sugar determined by the above DNS method. [31] Inulin content was calculated by the total (measurement in the second aliquot) subtracted by the levan content (measurement in the first aliquot). Three biological replicates were used for the statistics.

**Antigen Preparation and Antibody Production:** Recombinant proteins used as antigens for antibody production were purified in two steps. The first step was affinity purification by passing through a HisPur Ni-NTA Spin column. The eluted protein was further purified by running through a 4–12% gradient SDS/PAGE gel. After staining with Coomassie Brilliant Blue R-250 (Sigma), the major band corresponding to the target protein was cut out. Several gel slides for the same recombinant protein were combined and placed into dialysis tubing (Thermo Scientific), which contained 200 × 10⁻³ m Tris-acetate buffer (pH 7.4) containing 1% SDS and 1.5% DTT. The dialysis tubing containing gel slides was put into an Oval Horizontal Electrophoresis System (ThermoFisher) for horizontal electrophoresis in a buffer containing 50 × 10⁻³ m Tris-acetate (pH 7.4) and 0.1% SDS. After running at 90 mA for 4–5 h, the dialysis tubing was then transferred into a buffer containing 20 × 10⁻³ m Tris-HCl (pH 6.8) and 0.02% SDS for dialysis overnight at 4 °C. The buffer containing proteins in the dialysis tubing were then transferred into a 1.5 mL microtube and proteins in the tube were precipitated with a 10% TCA (4 volume of cold acetone per 1 volume of protein mix) at −20 °C overnight. After centrifugation, the protein pellet was washed with cold acetone, dried, and redissolved in PBS. [51] 5 mg of the purified protein with a concentration ≥0.5 mg mL⁻¹ was used for antibody production. Polyclonal antibodies were produced via a commercial contract with GenScript (Piscataway, NJ). Antibodies were affinity purified against
the original antigen. Possible immunoglobulin against His-tag in the recombinant protein was eliminated by passing through a HisPur Ni-NTA Spin column.

**IHC Staining:** The indirect immunohistochemical staining was carried out following a revised protocol described by Simo et al. [32] Briefly, salivary glands, midguts, Malpighian tubules, and fat bodies were dissected from Hessian fly larvae in PBS buffer. The dissected tissues were fixed overnight in 4% paraformaldehyde in PBS, followed by three washes in PBS containing 1% Triton X-100 (PBST). The tissues were blocked in 1× PBS, 0.2% Triton X-100, 10% DMSO, and 6% normal goat serum (Sigma) for 60 min. The tissues were incubated in 1× PBS containing 0.2% Triton X-100, 10% DMSO, and 1:1000 dilution of a primary antibody for 2 d. After washing with 1× PB containing 0.2% Tween-20 and 10 µg/ml heparin for 75 min with a buffer change every 15 min, the tissues were incubated overnight with 1:1000 dilution Alexa 488-labeled secondary antibodies (Molecular Probes). After 75 min of washing with 1× PB containing 0.2% Tween-20 and 10 µg/ml heparin with a buffer change every 15 min, the tissues were then stained with the DNA-staining dye 4',6-diamidino-2-phenylindole, dihydrochloride (DAPI). After three washes with PBS, the stained tissues were mounted onto a glass slide in PBS. Negative controls included preabsorption of each antibody with its respective antigen have been carried for each antibody staining. Images for the stained tissues were captured using a confocal microscope (Zeiss LSM-700).

**Sequence Alignment and Phylogenetic Analysis:** NCBI BLAST search was carried out to identify homologous sequences (https://blast.ncbi.nlm.nih.gov/BLAST.cgi). The identities between protein pairs were also generated by BLAST two-sequence alignment. Conserved domains were analyzed using CD-Search in NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Gene structure was generated on Gene Structure Display Server (GSDS). [33] Hessian fly MDLS and homologous sequences were used to produce a phylogenetic tree based on the neighbor-joining method [34] using the Molecular Evolutionary Genetics Analysis version 7.0 (MEGA7) [35] with 500 bootstrap tests.

**Statistical Analysis:** All data represent three independent experiments. Data are presented as mean ± SD (n = 3). Statistics were using Student’s t-test or ANOVA test. Different letters show significant differences (p < 0.05). Groups with the same letter are not significantly different from one another. *p* represents significant differences at *P* < 0.05, and **p** indicates at *P* < 0.01.

**Supporting Information**

Supporting Information is available from the Wiley Online Library or from the author.

**Acknowledgements**

This paper is a joint contribution from the U.S. Department of Agriculture-Agriculture Research Service at Manhattan, KS, and the Kansas Agricultural Experiment Station. Mention of a commercial or proprietary product does not constitute an endorsement or recommendation for its use by the USDA. USDA is an equal opportunity provider and employer.

**Conflict of Interest**

The authors declare no conflict of interest.

**Author Contributions**

M.S.C. conceived the study, X.C. and M.S.C. designed the research plan. X.C. and S.G.C. conducted the experimental work. X.C., J.P.F., Y.P., and M.S.C. analyzed the data. R.J.W., Y.P., and M.S.C. provided reagents and other resources for the research. X.C., R.J.W., J.P.F., Y.P., and M.S.C. wrote the paper. The authors thank Dr. Susan Brown, Dr. Michael Kanost, and Dr. Raymond Cloyd for reviewing an earlier version of this article.

**Keywords**

animal fructan metabolism, horizontal gene transfer, inulinase, levanase

Received: November 19, 2019
Revised: January 30, 2020
Published online: February 21, 2020

[1] C. Zhao, L. N. Escalante, H. Chen, T. R. Benatti, J. Qu, S. Chellapill, R. M. Waterhouse, D. Wheeler, M. N. Andersson, R. Bao, M. Batterson, S. K. Behura, K. P. Blankenburg, D. Caragea, J. C. Carolan, M. Coyle, M. El-Bouhssini, L. Francisco, M. Friedrich, N. Gill, T. Grace, C. J. P. Grimmlerikzuizen, Y. Han, F. Hauser, N. Herndon, M. Holder, P. Ioannidis, L. Jackson, M. Javaid, S. N. Jhangiani, *Curr. Biol.* 2015, 25, 613.

[2] M.-S. Chen, X. Liu, Z. Yang, H. Zhao, R. H. Shulke, J. J. Stuart, S. Hulbert, *BMC Evol. Biol.* 2010, 10, 296.

[3] M.-S. Chen, H.-X. Zhao, Y. C. Zhu, B. Scheffler, X. Liu, X. Liu, S. Hulbert, J. J. J. Stuart, *J. Infect. Physiol.* 2008, 54, 1.

[4] M.-S. Chen, J. P. Fellers, J. J. Stuart, J. C. Reese, X. Liu, *Insect Mol. Biol.* 2004, 13, 101.

[5] Z. Wang, J.-Q. Ge, H. Chen, X. Cheng, Y. Yang, J. Li, R. J. Whitworth, *M.-S. Chen*, *Insect Biochem. Mol. Biol.* 2018, 100, 30.

[6] J. J. Stuart, M.-S. Chen, R. Shulke, M. O. Harris, *Annu. Rev. Physiol.* 2012, 50, 339.

[7] L. Zhu, X. Liu, X. Liu, R. Jeannotte, J. C. Reese, M. Harris, J. J. Stuart, M.-S. Chen, *Mol. Plant-Microbe Interact.* 2008, 21, 70.

[8] M. O. Harris, T. P. Freeman, O. Rohlfritsch, K. G. Anderson, S. A. Payne, J. A. Moore, *Ann. Entomol. Soc. Am.* 2006, 99, 305.

[9] R. A. Byers, R. L. Gallun, *J. Econ. Entomol.* 1972, 65, 955.

[10] C. Khajuria, H. Wang, C. Khajuria, J. C. Reese, R. J. Whitworth, R. Welti, M.-S. Chen, *Mol. Plant-Microbe Interact.* 2012, 25, 920.

[11] L. Zhu, X. Liu, H. Wang, C. Khajuria, J. C. Reese, R. J. Whitworth, R. Welti, M.-S. Chen, *Mol. Plant-Microbe Interact.* 2004, 27, 920.

[12] X. Liu, J. Bai, L. Huang, L. Zhu, X. Liu, N. Weng, J. C. Reese, M. Harris, J. J. Stuart, M.-S. Chen, J. J. Stuart, M.-S. Chen, *J. Insect Physiol.* 2007, 33, 2171.

[13] Z. L. Tan, H. P. Chen, L. H. He, R. J. Fang, T. X. Xing, *Anim. Feed Sci. Technol.* 1995, 53, 337.

[14] G. A. F. Hendry, *New Phytol.* 1993, 123, 3.

[15] I. Vijn, S. Smeekens, *Plant Physiol.* 1999, 120, 351.

[16] V. B. McGrath, A. B. Blakeney, G. D. Batten, *New Phytol.* 1997, 136, 145.

[17] J. H. Hatchett, G. L. Kreitner, R. J. Elzinga. *Ann. Entomol. Soc. Am.* 1990, 83, 1137.

[18] J. C. D. Hotopp, M. E. Clark, D. C. S. G. Oliveira, J. M. Foster, P. Fischer, M. C. M. Torres, J. D. Giebel, N. Kumar, N. Ishmael, S. Wang, J. Ingram, R. V. Nene, J. Shepard, J. Tomkins, S. Richards, D. J. Spirel, E. Chedin, B. E. Stlasko, H. Tettelin, J. H. Werren, *Science* 2007, 317, 1753.

[19] E. G. J. Danchin, E. A. Guzeeva, S. Mantelin, A. Berepiki, J. T. Jones. *Mol. Biol. Evol.* 2016, 33, 1571.

[20] R. Acuña, B. E. Padilla, C. P. Flórez-Ramos, J. D. Rubio, J. C. Herrera, P. Benavides, S.-J. Lee, T. H. Yeats, A. N. Egan, J. D. Joyce, K. C. Rose, *Proc. Natl. Acad. Sci. USA* 2012, 109, 4197.

[21] E. G. J. Danchin, M.-N. Rosso, P. Vieira, J. de Almeida-Engler, P. M. Coutinho, B. Henriassat, P. Abad, *Proc. Natl. Acad. Sci. USA* 2010, 107, 16751.
[22] S. Zhang, R. Shukle, O. Mittapalli, Y. C. Zhu, J. C. Reese, H. Wang, B.-Z. Hua, M.-S. Chen. *Insect Physiol.* **2010**, *56*, 1198.
[23] N. Saitou, M. Nei. *Mol. Biol. Evol.* **1987**, *4*, 406.
[24] S. Kumar, G. Stecher, K. Tamura. *Mole. Biol. Evol.* **2016**, *33*, 1870.
[25] J. Felsenstein. *Evolution* **1985**, *39*, 783.
[26] E. Zuckerkandl, L. Pauling. *Evolutionary Divergence and Convergence in Proteins*, Elsevier, Amsterdam **1965**, p. 97.
[27] L. Zhang, W.-H. Li. *Mol. Biol. Evol.* **2004**, *21*, 236.
[28] J. L. Parrent, T. Y. James, R. Vasaitis, A. F. S. Taylor. *BMC Evol. Biol.* **2009**, *9*, 148.
[29] T. Pons, O. Olmea, G. Chinea, A. Beldarrain, G. Márquez, N. Acosta, L. Rodríguez, A. Valencia. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 383.
[30] K. M. Trollope, N. van Wyk, M. A. Kotjomela, H. Volschenk. *FEBS J.* **2015**, *282*, 4782.
[31] Daimon, T. Taguchi, Y. Meng, S. Katsuma, K. Mita, T. Shimada. *J. Biol. Chem.* **2008**, *283*, 15271.
[32] R. Pedezzi, F. P. P. Fonseca, C. D. Santos Júnior, L. T. Kishi, W. R. Terra, F. Henrique-Silva. *Insect Biochem. Mol. Biol.* **2014**, *55*, 31.
[33] G. L. Miller. *Anal. Chem.* **1959**, *31*, 426.
[34] J. M. Rippe, T. J. Angelopoulos. *Adv. Nutr.* **2013**, *4*, 236.
[35] J. R. Brown. *Nat. Rev. Genet.* **2003**, *4*, 121.
[36] P. J. Keeling, J. D. Palmer. *Nat. Rev. Genet.* **2008**, *9*, 605.
[37] C. Pál, B. Papp, M. J. Lercher. *Nat. Genet.* **2005**, *37*, 1372.
[38] E. V. Koonin, K. S. Makarova, L. Aravind. *Annu. Rev. Microbiol.* **2001**, *55*, 709.
[39] S. Choudhuri. *Bioinformatics for Beginners*, Elsevier, Amsterdam **2014**.
[40] J. C. D. Hotopp. *Trends Genet.* **2011**, *27*, 157.
[41] R. Kirsch, L. Gramzow, G. Theißen, B. D. Siegfried, R. H. ffrench-Constant, D. G. Heckel, Y. Pauchet. *Insect Biochem. Mol. Biol.* **2014**, *52*, 33.
[42] N. Wybouw, Y. Pauchet, D. G. Heckel, T. Van Leeuwen. *Genome Biol. Evol.* **2016**, *8*, 1785.
[43] R. Bansal, S. Hulbert, B. Schmerhorn, J. C. Reese, R. J. Whitworth, J. J. Stuart, M.-S. Chen. *PloS One* **2011**, *6*, e23170.
[44] R. Bansal, S. H. Hulbert, J. C. Reese, R. J. Whitworth, J. J. Stuart, M.-S. Chen. *Pathogens* **2014**, *3*, 459.
[45] P. D. Facey, G. Méric, M. D. Hitchings, J. A. Pachebat, M. J. Hegarty, X. Chen, L. V.A. Morgan, J. E. Hoepppner, M. M. A. Whitten, W. D. J. Kirk, P. J. Dyson, S. K. Sheppard, R. D. Sol. *Genome Biol. Evol.* **2015**, *7*, 2188.
[46] M.-S. Chen, E. Echegaray, R. J. Whitworth, H. Wang, P. E. Sloderbeck, A. Knutson, K. L. Giles, T. A. Royer. *J. Econ. Entomol.* **2009**, *102*, 774.
[47] M.-S. Chen, S. Liu, H. Wang, X. Cheng, M. El Bouhssini, R. J. Whitworth. *PloS One* **2016**, *11*, e0155616.
[48] G. Chhetri, P. Kalita, T. Tripathi. *MethodsX* **2015**, *2*, 385.
[49] I. Dahech, H. B. Ayed, K. S. Belghith, H. Belghith, H. Mejdoub. *Int. J. Biol. Macromol.* **2013**, *60*, 128.
[50] A. C. Muro, E. Rodríguez, C. M. Abate, F. Síveriz. *Folia Microbiol.* **1999**, *44*, 647.
[51] A. J. Link, J. LaBaer. *Cold Spring Harbor Protocols* **2011**, *6*, 993.
[52] L. Simó, M. Slovák, Y. Park, D. Zitňan. *Cell Tissue Res.* **2009**, *335*, 639.
[53] B. Hu, J. Jin, A.-Y. Guo, H. Zhang, J. Luo, G. Gao. *Bioinformatics* **2015**, *31*, 1296.