

***Tribolium castaneum* Larval Gut Transcriptome and Proteome: A Resource for the Study of the Coleopteran Gut**

Kaley Morris,[†] Marcé D. Lorenzen,[‡] Yasuaki Hiromasa,[†] John M. Tomich,[†] Cris Oppert,[‡] Elena N. Elpidina,[§] Konstantin Vinokurov,[#] Juan Luis Jurat-Fuentes,[‡] Jeff Fabrick,[⊥] and Brenda Oppert^{*||}

Department of Biochemistry and Biotechnology/Proteomics Core Facility, Kansas State University, Manhattan, Kansas 66506, Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee 37996, A. N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119991, Russia, Institute of Entomology, Biology Center, ASCR, v.v.i., Ceske Budejovice 37005, Czech Republic, USDA, ARS U.S. Arid-Land Agricultural Research Center, 21771 North Cardon Lane, Maricopa, Arizona 85238, and ARS Grain Marketing and Production Research Center, 1515 College Avenue, Manhattan, Kansas 66502

Received February 19, 2009

Tribolium castaneum is an important agricultural pest and an advanced genetic model for coleopteran insects. We have taken advantage of the recently acquired *T. castaneum* genome to identify *T. castaneum* genes and proteins in one of the more critical environmental interfaces of the insect, the larval alimentary tract. Genetic transcripts isolated from the *T. castaneum* larval gut were labeled and hybridized to a custom array containing oligonucleotides from predicted genes in the *T. castaneum* genome. Through a ranking procedure based on relative labeling intensity, we found that approximately 17.6% of the genes represented in the array were predicted to be highly expressed in gut tissue. Several genes were selected to compare relative expression levels in larval gut, head, or carcass tissues using quantitative real-time PCR, and expression levels were, with few exceptions, consistent with the gut rankings. In parallel with the microarrays, proteins extracted from the *T. castaneum* larval gut were subjected to proteomic analysis. Two-dimensional electrophoretic analysis combined with MALDI-TOF resulted in the identification of 37 of 88 selected protein samples. As an alternative strategy, one-dimensional electrophoretic separation of *T. castaneum* larval gut proteins followed by two-dimensional nano-HPLC and ESI-MS/MS resulted in the identification of 98 proteins. A comparison of the proteomic studies indicated that 16 proteins were commonly identified in both, whereas 80 proteins from the proteomic analyses corresponded to genes with gut rankings indicative of high expression in the microarray analysis. These data serve as a resource of *T. castaneum* transcripts and proteins in the larval gut and provide the basis for comparative transcriptomic and proteomic studies related to the gut of coleopteran insects.

Keywords: *Tribolium castaneum* • microarray • proteomics • Coleoptera • insect control • insect gut

Introduction

Recently, the red flour beetle, *Tribolium castaneum*, became the first coleopteran insect to have its genome sequenced.¹ The value of this genome is enhanced by the fact that *T. castaneum* is one of the most serious pests of milled grains and stored products worldwide. Although many other coleopteran insects are major agricultural pests, little information is available

pertaining to the expression profile of genes and proteins in the coleopteran gut, despite the importance of this critical interface between the beetle and its environment. Already a powerful model organism for studying insect development, the availability of this genome sequence, along with previously developed genetic tools such as germline transformation and RNA interference, make *Tribolium* an obvious candidate for profiling genes and proteins in the coleopteran larval gut.

While *Tribolium* is currently the best suited coleopteran for genetic analyses, a closely related beetle, the yellow mealworm (*Tenebrio molitor*), has been the focus of many biochemical studies involving the larval gut (reviewed in refs 2 and 3). Essentially, the *T. molitor* gut has an acidic anterior midgut with predominantly cysteine protease activity and an alkaline posterior midgut with relatively more serine protease activity.^{4–8} This digestive strategy is presumably an adaptation to feeding

* To whom correspondence should be addressed: Email, bso@ksu.edu; Telephone, 1-785-776-2780; Fax, 1-785-537-5584.

[†] Department of Biochemistry and Biotechnology/Proteomics Core Facility, Kansas State University.

[‡] Department of Entomology and Plant Pathology, University of Tennessee.

[§] A. N. Belozersky Institute of Physico-Chemical Biology, Moscow State University.

[#] ASCR.

[⊥] USDA ARS Grain Marketing and Production Research Center.

^{||} USDA ARS U.S. Arid-Land Agricultural Research Center.

on cereals that contain high levels of defense-related inhibitors, whereby cysteine proteases in the anterior midgut degrade cereal serine protease inhibitors before they reach target proteases in the posterior midgut. A similar digestive strategy has been found in the *T. castaneum* larval gut, where the acidic anterior gut contains 80% of the total proteolytic activity, mostly due to cysteine proteases.⁹ However, there are insufficient genomic and proteomic characterizations of the coleopteran gut to fully understand the process of digestion in these insects.

Considerable efforts were made to improve the annotation of the *T. castaneum* genome.¹ However, an EST study suggested that as much as one-third of the genome still lacks sufficient annotation.¹⁰ The goal of this study was to use transcriptomic and proteomic profiling to delineate genes and proteins in the *T. castaneum* gut and provide a tool to enhance the understanding of the coleopteran gut.

Experimental Section

Insects and Dissection. *T. castaneum* were reared on a diet of 95% wheat flour and 5% brewer's yeast at 28 °C, 75% R.H., in complete darkness. For dissection, larvae were ice-anesthetized, the anterior and posterior ends were removed with sterile forceps, and guts were dissected and placed in the appropriate solution, as described for each procedure.

The Tc Genome Array. For transcriptome analysis, a "genome" microarray for *T. castaneum* (Tc Genome Array) was developed using annotated sequences (designated with GLEAN numbers) from the *T. castaneum* genome annotation project.¹ For the array, 35-mer oligonucleotide probes were designed to approximately 12 000 GLEAN sequences for the custom array (CustomArray 12K, Combimatrix, Mukilteo, WA). Although 16 400 genes were estimated in the *Tribolium* genome, the number was reduced to 12 000 by eliminating 4400 sequences of housekeeping genes, those with high sequence identity to transposable elements or retrotransposable elements, and those with a lack of identity to any known sequence. Controls were also included in the array for background signal, including oligos from nontarget species, mismatch oligos for genes with variable expression, and antisense oligos.

Isolation of RNA and Hybridization to the Chip. To synchronize insects, eggs were collected from *T. castaneum* adults over a 24 h period, and neonates were reared on 85% stabilized wheat germ, 10% hard red winter wheat flour, and 5% brewer's yeast. At 13 days of age, guts were extracted from six to eight larvae for each of three biological replicates and one technical replicate, frozen in liquid nitrogen, and total RNA was extracted (RNeasy kit, Qiagen, Valencia, CA). The integrity of RNA was determined (Bioanalyzer, Agilent, Santa Clara, CA), and 1 μ g of total RNA was used to synthesize amino allyl-modified aRNA labeled with either Cy3 or Cy5 dye (Amino Allyl MessageAmp Kit, Ambion, Austin, TX). Labeled aRNA was hybridized to an individual Tc genome array slide (i.e., separate slides for each replicate) for 18 h at 45 °C. Hybridized arrays were scanned at 635 or 532 nm (depending on the dye) at the Gene Expression Facility at Kansas State University (GenePix 4000B scanner, Molecular Devices, Sunnyvale, CA). Data were analyzed by GenePix Pro 6.1 (Molecular Devices), normalized by the ratio of medians after background fluorescence was subtracted from the median intensity. Parameters for scanning and analysis are provided in Supporting Information Table 3, "GenePix Parameters" spreadsheet.

Analysis of Relative Gut Expression. *T. castaneum* larval gut RNA from different biological samples and technical replicates

were labeled with either Cy3 or Cy5 dyes (aRNA), and labeled RNA was hybridized to Tc genome arrays, as previously described. The median labeling intensity of aRNA hybridized to each oligonucleotide probe minus background was used as a relative measure of fluorescence intensity for each arrayed oligonucleotide (Supporting Information Table 3, Gut Rankings Table; biological replicates are columns E, F, and H, and technical replicate G). To compare the average relative fluorescence for each oligonucleotide (representing a particular GLEAN), a ranking system was devised. Intensity values for each oligonucleotide were averaged and rankings were designated by dividing the intensity value for each individual oligonucleotide by the largest value (78 561) and multiplying each by 100 (column I). The fluorescence intensity of each probe (1 probe = 1 GLEAN model) was ranked from 0 to 100 (Supporting Information Table 3, column J). These rankings are an approximation of the relative expression level of gut-specific transcripts represented by an oligonucleotide probe on the array.

Quantitative Real-Time PCR (qRT-PCR). The relative expression of selected *T. castaneum* transcripts in different tissues was evaluated by qRT-PCR. Primer pairs were designed by Primer3 software (v.0.4.0, 11). Primer pairs were synthesized for several genes predicted to be differentially expressed based on their gut ranking (Table 1). Total RNA was obtained from the tissue of eight 13 d old *T. castaneum* larvae from three biological replicates (RNeasy, Qiagen). Tissue dissections were performed by the following procedure: The head and part of the prothorax was severed and stored (head); the posterior end of the larva was severed and added to the "carcass" tube; the gut (with partial foregut and hindgut and some adhering fat body and Malpighian tubules) was removed (gut), and the remaining carcass was combined with the posterior larval segment (carcass), each stored in RNAlater (Ambion). RNA extractions were prepared similarly, except that samples from the second and third biological replicates were treated with DNase prior to RNA extraction.

Comparative, real-time PCR was performed using an Mx3000P QPCR system (Stratagene, La Jolla, CA) in 25 μ L reactions with 2 \times Brilliant SYBR Green QRT-PCR Master Mix Kit (1-step, Stratagene) containing 1 ng of RNA template, 200 nM each of the forward and reverse primer, 2.44 μ L of water, 12.5 units of RNase Block/StrataScript reverse transcriptase, and 12.5 μ L of 2 \times Master Mix containing SureStart Taq DNA polymerase and SYBR Green. Two technical replications were performed for each of the biological replicates. The thermal profile included 30 min at 50 °C for reverse transcription, 10 min at 95 °C for activation of the DNA polymerase followed by 40 cycles (30 s at 95 °C, 60 s at 55 °C for annealing, and 30 s at 72 °C for extension). Fluorescence data were collected during each annealing step. After the final cycle, dissociation curve data was collected using standard protocol (Stratagene) from 55 to 95 °C, 0.2°/s. Dissociation curves were examined for additional peaks (indicating nonspecific amplification; data not shown). The MxPro 4.01 software (Stratagene) was used to calculate the values for threshold cycles (Ct) in the amplification curves and provided the fold-change in gene expression. Two housekeeping genes, α -tubulin (Glean 04873) and RPS6 (Glean 10830), were used to normalize gene expression levels, and log fold change in gene expression levels in the gut and head were

Table 1. Nucleotide Primers and Gene Targets Amplified by qRT-PCR

primer name ^a	sequence	NCBI #	gut rank ^b	predicted protein
Tc00222-F	AACGAGCCAAGGACAGCTAA	XM_966295	2	cadherin
Tc00222-R	TAGATTTGAGCGGTGGCTCT			
Tc00515-F	ATGTTTCGCATTCGCTCTTCT	XM_970082	45	thaumatin
Tc00515-R	GTGACACTTTGGCCCTGATT			
Tc04873-F	CGCTGTTGGGAATTGTACT	XM_961399	34	α -tubulin
Tc04873-R	ACGACAGTGGGTTCCAAGTC			
Tc05376-F	TTCCAACCTTTGACCTACCG	NM_001114379	51	arylphorin
Tc05376-R	TGCTGATGTCGGCAAAGTAG			
Tc07214-F	AACGCCATCAAAACCAACA	XM_965419	1	cathepsin O
Tc07214-R	CCCCCGTCCCCAAAGT			
Tc10930-F	GTGGATTTTAACGGCTGCTC	XM_963301	43	serine protease homologue (SPH)
Tc10930-R	GTTCAAGCGTGAGTGGATCA			
Tc11001-F	TTGGTGGATTGTTCCITCACA	XM_965320	82	cathepsin L
Tc11001-R	TCATCTCCAGAGGGGATGTC			
Tc13574-F	TTCGGGCAAAAAGTACCACAAAG	XR_043134	1	proline oligopeptidase (POP1)
Tc13574-R	CACGCGGCCACCAACA			
Tc15370-F	CGATTTTAAACATGGGGGATG	XM_001810641	1	cellulase
Tc15370-R	TCTTATAGGCCGGTCGTGTC			
Tc16224-F	CCTCAAGGGCGATTATTGAA	XM_962989	1	proline oligopeptidase (POP2)
Tc16224-R	CGCGTCCATTTAAAACCATCT			
Tc10830-F	ACGCAAGTCAGTTAGAGGGTGCAT	XM_963302	39	ribosomal protein (RPS6)
Tc10930-R	TCCTGTTTCGCCTTTACGCACGATA			
Tc13765-F	GCTGGGGCCCCCTCGTGT	XM_968054	2	rhodopsin6
Tc13765-R	GCATTTCGGCGCTGGTTTG			
Tc13834-F	GCGGGGCTTGGGTACGGTGTCT	XM_967074	1	sine oculus
Tc13834-R	CGCCCAACGGTCTGCCTCTG			
Tc12665-F	GAGCCCGGCCGAGAAATACG	XM_965721	3	glass
Tc12665-R	TGCGGAGGCCTAAAGTCAGC			
Tc00920-F	AATCGCCCCACATTAGTCC	NM_001039421	2	dfd
Tc00920-R	CCGTAATCCGACTTGATGGT			

^a Tc is followed by the number that correlates to the GLEAN number in the annotation project. ^b See Supporting Information.

compared to that of the carcass. Data was *ln* transformed and was the average of 3 biological replicates, \pm S.E.

Separation of *T. castaneum* Gut Proteins by 1D or 2D SDS-PAGE Gel Analysis. To identify proteins expressed in the *T. castaneum* gut, proteins were analyzed by 2-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2D SDS-PAGE) and MALDI-TOF or by 1-dimensional SDS-PAGE (1D SDS-PAGE) and 2D nano-HPLC and ESI-MS/MS. Proteins were extracted from 20 *T. castaneum* larval guts, 12–14 d old, following homogenization in 100 μ L of Tris/leupeptin buffer (25 mM Tris, pH 8.0, 0.5 mM leupeptin) with 0.05% PPS surfactant (Protein Discovery, Knoxville, TN) for 1D gel separation or Tris/leupeptin buffer and 4 M urea for 2D separation. Extracts were stored frozen at -80 °C.

For 1D separation, samples were thawed, centrifuged at 15 000 \times *g* and 4 °C for 5 min, and 15 μ L of supernatant was separated in 10–20% Tricine gels (Invitrogen, Carlsbad, CA) under reducing conditions. The gel was stained with Coomassie R-250 (Imperial Protein Stain, Pierce Chemical Co.). For 2D SDS-PAGE, gradients of pH 3–7 and 3–10 in precast gels (Invitrogen) were evaluated, and with the latter there were few proteins in the basic range and lower resolution of proteins in the acidic range. Therefore, pH 3–7 gels were chosen for optimal separation of proteins from the *T. castaneum* larval gut. For separation, 15 μ L of thawed gut extract was mixed with one volume of pH 3–7 IEF sample buffer and subjected to electrophoresis in precast pH 3–7 IEF gels according to the recommended protocol. Individual lanes from IEF stained gels were cut and overlaid on a 4–12% Bis-Tris Zoom gel according to the manufacturer's recommended procedure, and proteins

were separated in the second dimension under reducing conditions using MOPS SDS running buffer (Invitrogen).

Trypsin Digestion of Proteins. Following 1D SDS-PAGE separation, gels were stained with Coomassie R-250, and lanes were cut horizontally between molecular mass markers (nine sections total) and transferred to microcentrifuge tubes. Protein spots were excised from stained 2D gels as 2–3 mm diameter pieces and transferred to 1.5 mL microcentrifuge tubes. For both 1D and 2D samples, protein-free regions of the gels were excised as background controls. Stained gel pieces were destained and the SDS removed by washing three times for 1 h in 100 μ L 1:1 acetonitrile/water at 30 °C. Gel pieces were dried for 30 min under vacuum and rehydrated with sequencing grade trypsin (Trypsin Gold, Promega, Madison, WI), 10 ng/ μ L in 40 mM ammonium bicarbonate, pH 8.0, in 10 μ L aliquots to completely cover the dried gel pieces. Hydrolysis was at 30 °C for 17 h. The digestion solution was transferred to 1.5 mL microcentrifuge tubes, and tryptic peptides remaining within the gel plugs were recovered by another extraction with 50 μ L of 50% acetonitrile in water and 2% trifluoroacetic acid (TFA) at 30 °C for 2 h. Fractions were combined and dried by speed vacuum concentration, and tryptic peptides were resuspended in 5 μ L of either 30 mg/mL 2,5-dihydroxybenzoic acid (DHB) (Sigma, St. Louis, MO) or 15 mg/mL α -cyano-4-hydroxycinnamic acid (HCCA) (Bruker Daltonics, Billerica, MA), each dissolved in 33% acetonitrile/0.1% TFA.

MALDI-TOF. To measure the tryptic peptide profile for each sample, 1.5 μ L of the peptide/matrix solution was dispensed onto a stainless steel target plate. MALDI TOF spectra were acquired with a Bruker Ultraflex III (Bruker Daltonics), and

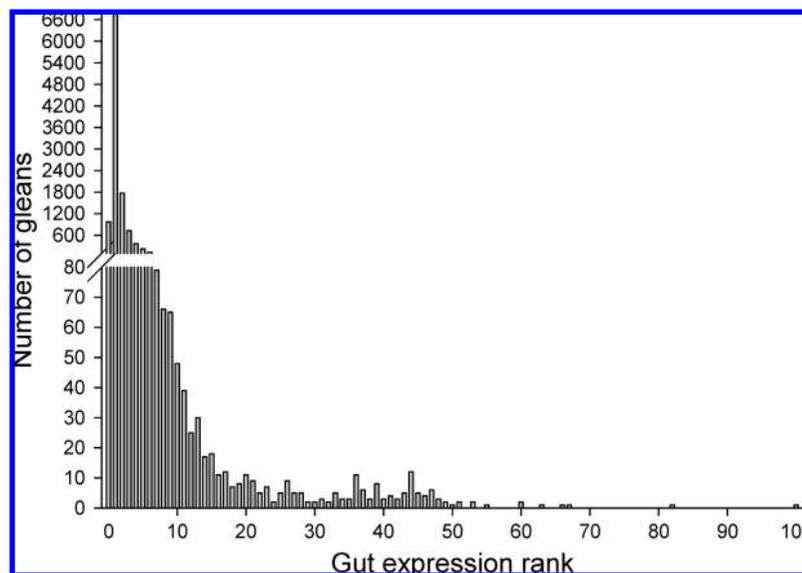


Figure 1. Relative number of GLEANS found in each ranking in the microarray analysis, from 0 to 100.

positively charged ions were acquired in the reflector mode over a m/z range of 500–5000 using delayed extraction. The spectra were processed by the FlexAnalysis 3.0 and Biotoools software 3.0 (Bruker Daltonics) without further smoothing or spectra processing. Monoisotopic masses were obtained using the SNAP algorithm with at least 5 for the signal/noise ratio of peak intensity. Measurements were externally calibrated with 8 different peptides ranging from 757.39 to 3147.47 (Peptide Calibration Standard II, Bruker Daltonics). The generated peak lists were transferred to ProteinScape 1.3 software that used the MASCOT 2.2 search engine for protein identification. We used the ProteinScape Score Booster feature to improve database search results by automatic iterative recalibrations and background eliminations.

2D nano-HPLC and Electrospray Ionization Tandem Mass Spectrometry (ESI–MS/MS). For 2D nano-HPLC, peptide fragments from 1D gels were injected onto a strong cation exchange microprecolumn (500 μm inner diameter (ID) \times 15 mm, BioX-SCX; LC Packings). Peptides were eluted from the column as fractions after injecting a step gradient of 50, 100, 500, and 1000 mM ammonium acetate. Each fraction was injected automatically in line onto a C18 reverse-phase micro column (300 μm ID \times 1 cm, PepMap; LC Packings) to remove salts. Peptides were separated with a C18 reversed-phase nanocolumn (75 μm ID \times 15 cm, PepMap; LC Packings) by a nanoflow linear acetonitrile/formic acid gradient. The 2D-nanoLC was performed automatically using a microcolumn switching device (Switchos; LC Packings) coupled to an autosampler (Famos; LC Packings) and a nanogradient generator (UltiMate Nano HPLC; LC Packings). The system control software, Hystar 3.2, was used to control the entire process. The eluted peptides were injected into an HCT Ultra Ion Trap Mass Spectrometer (Bruker Daltonics). The mass spectrometer was set up in the data dependent MS/MS mode to alternatively acquire full scans (m/z acquisition range from 300 to 1500 Da/e). The four most intense peaks in any full scan were selected as precursor ions and fragmented by collision energy. MS/MS spectra were interpreted and peak lists were generated by DataAnalysis 3.4 and Biotoools 3.0 software (Bruker Daltonics).

Bioinformatics. Peptide masses were compared to NCBIInr.2008.04.16 using MASCOT (<http://www.matrixscience.com>).

The following parameters were used in all searches: the maximum number of missed cleavages allowed was 1; the mass tolerance was 0.5 Da; minimum peptides required to match were 4; and the monoisotopic masses of observation were used to match the calculated monoisotopic fragment mass for protein identification. Mascot scores greater than 79 (NCBI) were significant in searches with MALDI-TOF peptides, whereas scores greater than 47 (NCBI Metazoa) were significant with peptides from ESI–MS/MS ($P < 0.05$).

For the microarray data, we used a custom annotated database developed in-house from the *Tribolium* genome annotation project,¹ and functional categories of highly expressed genes in the *T. castaneum* larval gut were compared using GeneSifter (Geospiza, Seattle, WA; Supporting Information Figure 5).

Results

Microarray Analysis. To identify abundant transcripts in the *T. castaneum* larval gut, we used custom microarrays (i.e., Tc genome array) containing oligonucleotides representing approximately 75% of the annotated “GLEAN” sequences from the *T. castaneum* genome.¹ Rankings were made for each GLEAN in the array based on the relative labeling intensity of each oligonucleotide probe (Supporting Information Table 3). We predicted that the rankings would provide a measure of the relative expression of each GLEAN transcript in the gut tissue. There was some variation in the hybridization of individual oligonucleotides among the biological replicates, which may reflect natural variation in gene expression among individual larva, or the lack of synchronicity in larval ages. Although there were inconsistencies within the technical replications, the relative ranking within the technical replicates was similar.

The “brightest” spots on the hybridized array were those with increased labeling intensity and represented GLEANS with higher gut expression rankings (Figure 1). Most of the rankings were 0, 1, or 2. GLEANS with rankings ≥ 3 represented 17.6% of the total oligonucleotides on the array, and it is reasonable to assume that a higher ranking was indicative of the relative expression of the GLEAN gene in the gut. We arbitrarily chose

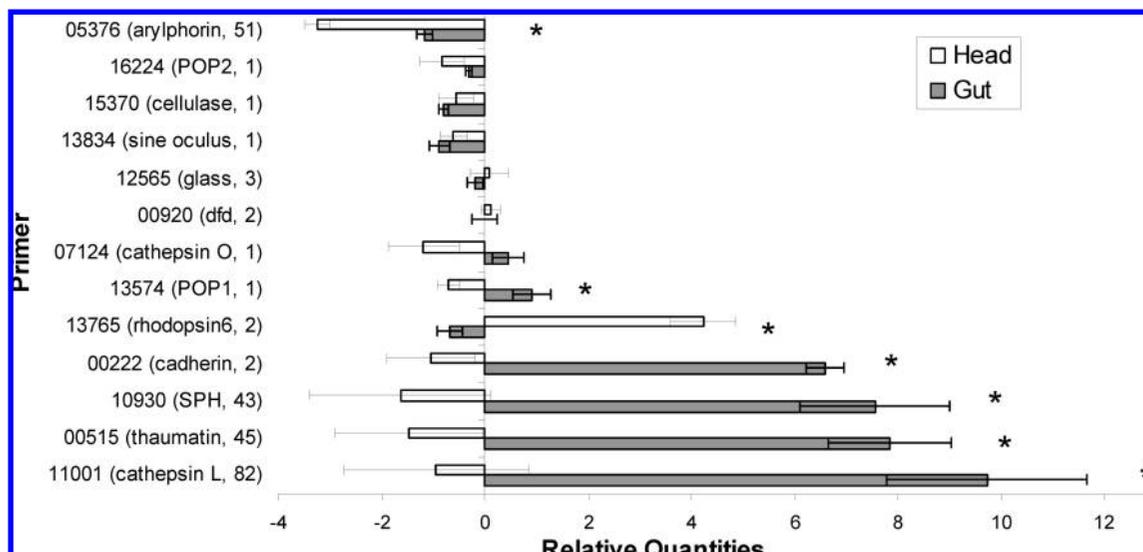


Figure 2. qRT-PCR comparison of the quantities of transcripts from gut and head relative to that of carcass templates (data is \ln transformed, average \pm S.E., $n = 3$ biological replicates), using primers specific for selected transcripts (Table 1, with predicted gene function and gut ranking in parentheses). Data with asterisks are significantly different (ANOVA of transcript quantities in head and gut templates, $p < 0.05$).

the cutoff ranking of 3, containing 2033 GLEAN transcripts more likely to be expressed at higher levels in the gut.

qRT-PCR. To further examine gut expression profiles, qRT-PCR was used to compare relative transcript abundance in total RNA templates from larval gut, head, and carcass. We chose primers corresponding to genes of interest: those predicted to be highly expressed in gut tissue (GLEANs 00515, 05376, 10930, 11001) and those predicted to have lower levels of expression in the gut (GLEANs 00222, 00920, 07214, 12565, 13574, 13765, 13834, 15370, 16224) based on rankings from microarray results (Table 1, Supporting Information Table 3).

The results of qRT-PCR were in agreement with the microarray predictions for GLEANs 00515, 10930, and 11001. The relative expression of these transcripts was more in the gut than in the carcass or head (Figure 2). GLEAN 00515 (gut ranking of 45) encodes a thaumatin-like protein, and GLEANs 10930 and 11001 (ranked 43 and 82, respectively) encode a serine protease homologue and cathepsin L, respectively. Therefore, the prediction that these genes were highly expressed in the larval gut based on microarray gut rankings was supported by qRT-PCR analyses, and all were significantly different from expression in the head.

GLEAN 00222 encodes a cadherin with homology to lepidopteran cadherins and is the ortholog of a cadherin in the *T. molitor* larval gut that confers sensitivity to *Bacillus thuringiensis* Cry3Aa toxin.¹² The *T. molitor* ortholog was cloned from larval midgut cDNA, suggesting that GLEAN 00222 also is expressed in the *T. castaneum* larval gut. Although the microarray ranking of 2 indicated that GLEAN 00222 was expressed at relatively low levels, qRT-PCR indicated significantly higher expression level in the gut than the head when compared to expression in the carcass of *T. castaneum* larvae (Figure 2).

Microarray predictions for GLEANs 00920, 07214, 12565, 13574, 13765, 13834, 15370, and 16224 were supported by qRT-PCR, indicating that the relative expression of these genes in the gut was either lower or similar to expression in the head and/or carcass (Figure 2). GLEAN 07214 encodes a cysteine cathepsin related to cathepsin O; GLEAN 15370 encodes a

cellulase-like protein; GLEANs 13574 and 16224 encode proteins similar to proline oligopeptidases; GLEANs 00920, 12565 (glass), 13765 (rhodopsin6), and 13834 (sine oculus) were chosen in an effort to identify transcripts more highly expressed in the head. All were ranked from 1–3 in the gut microarray, and all were expressed at levels similar to or lower than that of the carcass and/or head in the PCR assay, except for GLEAN 13765, which was expressed at significantly higher levels in the head than the gut.

GLEAN 05376, encoding arylphorin, ranked highly at 51 in the microarray experiment. However, qRT-PCR indicated that the level of expression of GLEAN 05376 was lower in the gut than in the carcass. Nonetheless, the expression of GLEAN 05376 in the gut was an order of magnitude higher than that of the head.

2D SDS-PAGE and MALDI TOF. For proteomic analysis, proteins were extracted from the *T. castaneum* larval gut and separated by 2D SDS-PAGE (Figure 3). A total of 88 protein spots were hand-picked and trypsinized, and peptides were analyzed as peptide mass fingerprints by MALDI-TOF/TOF. Of the 88 proteins picked for analysis, 57 were tentatively identified, and 37 had MASCOT scores that were significant (Table 2). Many proteins identified in the proteomic 2D gel correlated to highly expressed genes in the gut, with 69% of these correlating to genes having a ranking ≥ 3 in the microarray analysis. Proteins identifications included glycosyl hydrolases and amylases representing different families, as well as those associated with chitin metabolism and catabolism. Cysteine and serine proteases also were found and had high gut rankings in the transcriptome study. Other proteins included a sodium/solute symporter, serpin, SCP-like extracellular protein, tropomyosin, and an oxidoreductase. Several proteins were identified as hypothetical with unknown functions.

1D SDS-PAGE, 2D nano-HPLC, and ESI-MS/MS. For proteomic analysis using 2D nano-HPLC and ESI-MS/MS, *T. castaneum* gut proteins were extracted and separated by 1D SDS-PAGE (Supporting Information Figure 4). The 1D lane containing proteins was cut horizontally into equal-sized gel fragments, and each gel fragment was trypsinized and subjected

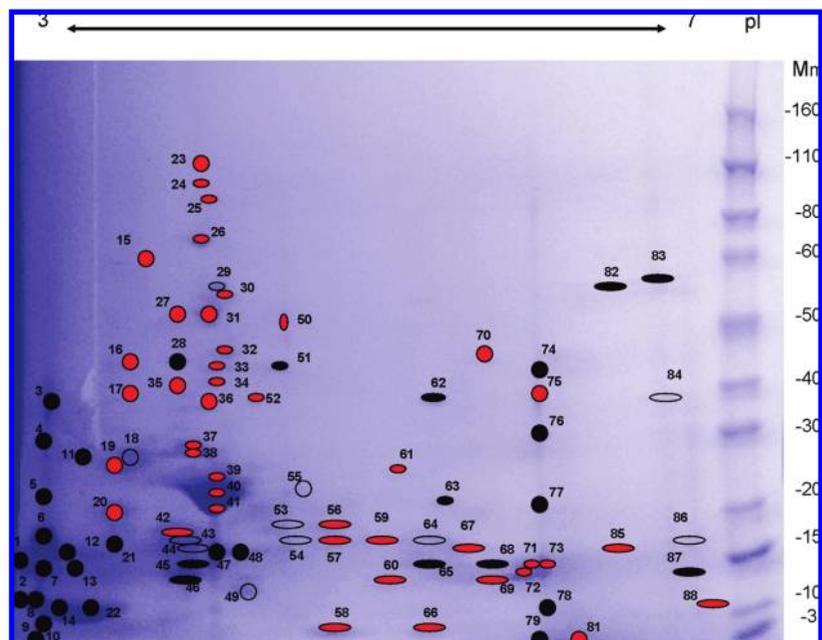


Figure 3. Depiction of Coomassie-stained 2D SDS-PAGE, with numbered spots picked for MALDI-TOF analysis (Table 2). Red spots represent identifications based on significant MASCOT scores; clear spots have nonsignificant MASCOT scores but are reasonable based on M_m and/or pI ; black spots were not identified.

to separation by 2D HPLC. Peaks were analyzed by ESI-MS/MS and filtered by MASCOT scoring of proteins in the NCBI database (Supporting Information Table 4). From this analysis, 98 proteins were identified with significant MASCOT scores, and 16 overlapped with those from the 2D SDS-PAGE MALDI-TOF analysis. Sixty-one of the proteins identified in this analysis had gut rankings ≥ 3 . Of the remainder, 18 were identified as NCBI proteins, but had no match to GLEANS and presumably were not represented on the chip. A few transcripts with low rankings in the transcriptome study, for example XP_976106 (larval serum protein 2) with a gut ranking of 1, were also identified through MALDI-TOF.

Genes in the top ranked category (≥ 3) were annotated using homology to proteins with defined function (data not shown) to obtain a summary of functional categories in the data set. Functional categories of *T. castaneum* gut-expressed genes were compared using a custom annotation database in GeneSifter (Geospiza, Seattle, WA; Supporting Information Figure 5). The categories of “binding” and “catalytic activity” contained the greatest number of sequences (more than 70%), whereas activities associated with “transporter activity” and “structural molecule activity” contained approximately 16% of the total sequences. These data are reasonable representations of activities associated with the insect gut, such as food digestion, energy production, and defensive responses.

Discussion

Incorporating microarray and proteomic approaches has greatly facilitated our studies of gut transcripts and proteins in coleopteran larvae. Combined with our previous biochemical data from the tenebrionid gut, we now can begin to assemble pathways of proteins involved in normal biological process, such as protein digestion, or in pathological response to bacterial toxins, such as *B. thuringiensis* Cry toxins.

This is the first report of thaumatin in the insect gut and may be related to previous reports of antifungal effects in *T.*

*castaneum*¹³ and/or thermal regulation in other coleopterans.¹⁴ The importance of high levels of thaumatin in the gut may be related to the relative refractivity of *T. castaneum* larvae to fungal pathogens compared to *T. molitor* (J. Lord, personal communication).

Cysteine proteases, such as cathepsin L, and serine proteases, such as trypsin and chymotrypsin, are important digestive enzymes in *T. castaneum* and *T. molitor* (reviewed in refs 2, 3, and 7–9). In *T. castaneum* larvae, 80% of the total proteolytic activity is found in the anterior midgut, and 97% of that activity is due to cysteine proteases.⁹ Cathepsin L transcripts have been found in the gut of *T. molitor*^{2,15} and *Diabrotica virgifera virgifera*,^{16,17} whereas cathepsin B transcripts also were found in *T. molitor*² and *Anthonomus grandis*.¹⁸ Of the 25 C1 cysteine peptidase family genes identified in our *Tribolium* annotation study,¹ seven cathepsin B and five cathepsin L encoding transcripts were found to be highly expressed (with a gut ranking ≥ 3 , Supporting Information Table 5). Five of the seven sequences encoding cathepsin B-like proteases are likely inactive enzymes (referred to as homologues), because they lack critical residues necessary for an active peptidase, whereas only one of five cathepsin L-like sequences was a homologue. The importance of highly expressed protease-like homologues in the gut has not been delineated, but we hypothesized that they may be a protective mechanism against cereal protease inhibitors.² A similar strategy to overcome dietary protease inhibitors has also been predicted for Lepidoptera.¹⁹ Four presumably active cathepsin L and two cathepsin B proteases were highly expressed in the *T. castaneum* larval gut according to the microarray analysis, and two of the cathepsin L (GLEANS 11000 and 11001) and one cathepsin B (GLEAN 02952) had proteomic support. Previous biochemical analysis of cysteine protease activities in *T. castaneum* larval midgut⁹ revealed the same number of proteases predicted by the microarray rankings.

At least 75 of the approximately 150 S1 serine peptidase genes in *T. castaneum* were found in the highly expressed gene

Table 2. Mascot Identifications of *T. castaneum* Gut Proteins Separated by 2D SDS-PAGE, Trypsin Digested, and Subjected to MALDI-TOF Analysis

spot number ^a	accession number ^b	protein family ^c	predicted function	Mascot score (# peptides/total)	M _m ^d (% Coverage)	pI	glean ^e	gut rank
15	AAW67571 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 16	85 (10/23)	42050.4 (30)	4.57	none	
	NP_001034515 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 3	70 (9/23)	42032.4 (26)	4.57	09176	37
16	XP_975666	Glycosyl hydrolase family 1 (PF00232)	β-glucosidase	102 (10/18)	58130.4 (19)	4.68	10052	55
17	NP_001103905	Polysaccharide deacetylase (PF01522)	chitin deacetylase 6	120 (13/20)	45047.5 (38)	4.61	13662	14
18	XP_971663	Dehydrogenase E1 component (PF00676)	oxidoreductase	51 (9/32)	47587.2 (25)	8.68	04386	1
19	AAW67571	Glycosyl hydrolases family 18 (PF00704)	Chitinase 16	93 (11/26)	42050.4 (34)	4.57		
	NP_001034515 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 3	76 (10/26)	42032.4 (29)	4.57	09176	37
20	XP_974298	Papain family cysteine protease (PF00112)	cathepsin B	62 (7/41)	36370.8 (33)	4.76	02952	43
23	XP_967022	Glycosyl hydrolases family 31 (PF01055)	glycosyl hydrolase	85 (11/19)	96188.0 (15)	5.12	16218	8
24	XP_967022	Glycosyl hydrolases family 31 (PF01055)	glycosyl hydrolase	85 (10/21)	96188.0 (13)	5.12		
25	XP_967022	Glycosyl hydrolases family 31 (PF01055)	glycosyl hydrolase	75 (14/33)	96188.0 (17)	5.12		
	XP_968058	Glycosyl hydrolases family 35 (PF01301)	β-galactosidase	55 (10/33)	74787.2 (20)	4.81	14348	9
26	XP_967022	Glycosyl hydrolases family 31 (PF01055)	glycosyl hydrolase	143 (20/36)	96188.0 (21)	5.12	16218	8
27	XP_975220	Alpha amylase (PF00128)	maltase 1	87 (16/51)	65305.4 (36)	4.66	08357	
	XP_975228	Alpha amylase (PF00128)	α-amylase	86 (16/51)	65366.6 (35)	4.70	08358	43
	XP_972326	Glycosyl hydrolases family 31 (PF01055)	glycosyl hydrolase	73 (13/51)	70808.2 (23)	4.54	08276	9
28	None	PP2AC superfamily	phosphoesterase	79 (12/40)	68905 (23)	4.67	07030	
29	XP_968058	Glycosyl hydrolases family 35 (PF01301)	β-galactosidase	88 (12/25)	74787.2 (23)	4.81	14260	1
30	NP_001038095 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	113 (9/44)	39944.8 (36)	4.81	9627	
31	NP_001038095 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	116 (8/28)	39944.8 (37)	4.81		
32	XP_975656	Glycosyl hydrolase family 20 (PF00728)	hexosaminidase A	77 (7/22)	61681.1 (14)	5.20	08778	22
33	XP_971482	Alkaline phosphatase (PF00245)	alkaline phosphatase	100 (10/19)	55131.3 (29)	5.32	06659	42
34	XP_971568	Alpha-L-fucosidase (PF01120)	α-fucosidase	75 (8/23)	52070.0 (36)	4.44	none	
	XP_974849	Immunoglobulin domain (PF00047)	IgG-like	40 (5/23)	48914.5 (9)	6.28	00861	0
35	NP_001103905	Polysaccharide deacetylase (PF01522)	chitin deacetylase 6	127 (13/30)	45047.5 (36)	4.61	13662	14
36	NP_001103905	Polysaccharide deacetylase (PF01522)	chitin deacetylase 6	127 (13/19)	45047.5 (36)	4.61	none	
	XP_973190	Polysaccharide deacetylase (PF01522)	chitin deacetylase 6	127 (13/19)	45047.5 (36)	4.61	02475	1
37	NP_001103906	Noggin (PF05806)	chitin deacetylase 8	110 (14/62)	42650.9 (50)	4.64	14147	15
	XP_973227	Polysaccharide deacetylase (PF01522)	chitin deacetylase 8	110 (14/62)	42650.9 (50)	4.64	14147	15
38	NP_001103906	Noggin (PF05806)	chitin deacetylase 8	88 (10/46)	42650.9 (36)	4.64	14147	15
	XP_973227	Polysaccharide deacetylase (PF01522)	chitin deacetylase 8	88 (10/46)	42650.9 (36)	4.64	14147	15
	NP_001034516 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 2	83 (12/46)	42082.5 (38)	4.58	09626	1
39	NP_001038095	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	94 (9/42)	39944.8 (43)	4.81	09627	49
	XP_972892 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	120 (11/48)	39944.8 (48)	4.81		
40	NP_001038095	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	120 (11/48)	39944.8 (48)	4.81		
	XP_972892 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	94 (8/36)	39944.8 (37)	4.81		
41	NP_001038095	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	94 (8/36)	39944.8 (37)	4.81		
	XP_972892 ^f	Glycosyl hydrolases family 18 (PF00704)	Chitinase 11	94 (8/36)	39944.8 (37)	4.81		
42	XP_967464	Trypsin (PF00089)	trypsin	73 (6/10)	48275.9 (22)	5.47	04041	1
43	XP_967464	Trypsin (PF00089)	trypsin	54 (5/23)	48275.9 (20)	5.47		
44	XP_967464	Trypsin (PF00089)	trypsin	50 (4/35)	48275.9 (18)	5.47		
49	(XP_966945)	SCP-like extracellular protein (PF00188)	SCP-like extracellular protein	80 (8/21)	21397.0 (43)	4.87	00599	
	XP_969295	Ras family (PF00071)	Ras	51 (5/21)	24621.8 (36)	5.53	05229	2
50	NP_001038094	Glycosyl hydrolases family 18 (PF00704)	Chitinase 8	73 (7/15)	54065.5 (20)	4.89	09624	41
	XP_972678 ^f	SSF (PF00474)	sodium/solute symporter	108 (17/21)	262417.7(10)	4.55	13040	20
53	XP_968394	Trypsin (PF00089)	trypsin	54 (4/13)	28468.0 (32)	4.96	10930	43
54	XP_969289	Glycosyl hydrolases family 31 (PF01055)	glycosyl hydrolase	36 (3/11)	96188.0 (14)	5.12	10907	6
55	XP_972660	Serpin (PF00079)	serpin	47 (6/21)	40451.9 (21)	5.36	00760	1
56	XP_976366	NONE	unknown	107 (7/17)	14222.7 (70)	5.09	14914	1
57	XP_968167	Trypsin (PF00089)	trypsin	73 (6/16)	28391.3 (24)	5.19	10927	44
58	XP_976366	NONE	unknown	122 (8/20)	14222.7 (70)	5.09	14914	1
59	XP_968167	Trypsin (PF00089)	trypsin	68 (6/20)	28391.3 (24)	5.19	10927	44
60	XP_970898	SCP (PF00188)	SCP-like extracellular protein	89 (8/17)	21707.6 (35)	5.40	00598	7

Table 2. Continued

spot number ^a	accession number ^b	protein family ^c	predicted function	Mascot score (# peptides/total)	M _m ^d (% Coverage)	pI	glean ^e	gut rank
61	XP_970838	Cathepsin propeptide inhibitor domain (I29) (PF08246) Papain family cysteine protease (PF00112)	cathepsin L	91 (8/11)	35560.7 (16)	5.33	11001	82
64	XP_968167	Trypsin (PF00089)	trypsin	62 (5/15)	28391.3 (21)	5.19	10927	44
66	XP_976366	NONE	unknown	114 (7/17)	14222.7 (64)	5.09	14914	1
67	XP_968504	KID (PF02524) Tropomyosin (PF00261)	tropomyosin	84 (16/35)	52340.3 (23)	5.08	11019	1
69	XP_976015	NONE	unknown	89 (6/7)	25445.2 (26)	5.43	00288	50
70	XP_973737	Flavin containing amine oxidoreductase (PF01593)	pyridine nucleotide-disulfide oxidoreductase	87 (8/11)	55023.6 (16)	6.13	09641	3
71	XP_976412	NONE	unknown	76 (7/18)	25403.2 (27)	5.74	00289	45
72	XP_976015	NONE	unknown	102 (9/22)	25445.2 (38)	5.43	00288	50
73	XP_976412	NONE	unknown	75 (12/73)	25403.2 (51)	5.74	00289	45
	XP_967475	GST N (PF02798) GST C (PF00043)	Glutathione S-transferase	48 (9/73)	23571.3 (38)	6.12	none	50
	XP_976015	NONE	unknown	46 (9/73)	25445.2 (37)	5.43	00288	50
74	None	Myosin	Myosin heavy chain	134 (25/30)	262094 (12)	5.49	05924	1
75	(XP_975828)	Hexokinase 1 (PF00349) Hexokinase 2 (PF03727)	hexokinase	120 (16/21)	50149.7 (29)	6.00	none	
	(XP_975717)	Hexokinase 1 (PF00349) Hexokinase 2 (PF03727)	hexokinase	100 (11)	50122.7	6.10	none	
80	NP_001107851	Hemocyanin N (PF03722)	arylphorin	65 (10/25)	91020.2 (11)	6.35	05376	51
81	NP_001107851 XP_967140 XP_976053	Hemocyanin N (PF03722) Hemocyanin M (PF00372) Hemocyanin C (PF03723)	arylphorin	129 (18/37)	91020.2 (26)	6.35	05376	51
	XP_967228	Hemocyanin N (PF03722) Hemocyanin M (PF00372) Hemocyanin C (PF03723)	hexamerin	43 (10/37)	84466.2	6.11	06769	24
	(XP_976106)	Hemocyanin N (PF03722) Hemocyanin M (PF00372) Hemocyanin C (PF03723)	hemocyanin	43 (10/37)	85264.1	6.01	06947	1
83	None	Lipoprotein	Retinoid/fatty acid binding protein	80(19/20)	421443 (5)	7.79	15029	1
84	XP_973475	Multiple families (21)	V-ATPase	72 (9/35)	58668.9 (22)	8.23	06249	1
85	XP_967228	Hemocyanin N, M, C (PF00372, PF03722, PF03723)	hemocyanin	83 (11/24)	84466.2 (16)	6.11	06769	24
	(XP_976106)			82 (11/24)	85264.1(16)	6.01	06947	1
86	NP_001034522 ^g	Hemocyanin N (PF03722) Hemocyanin M (PF00372) Hemocyanin C (PF03723)	prophenoloxidase	76 (14/34)	79369.7 (18)	7.67	15849	2
	XP_967179	Hemocyanin N (PF03722) Hemocyanin M (PF00372) Hemocyanin C (PF03723)	prophenoloxidase	61 (14/34)	87730.6	8.58	15848	1
88	XP_968284	Cu/Zn superoxide dismutase (PF00080)	Cu-Zn superoxide dismutase 1	98 (7/14)	15689.5 (62)	6.23	07011	9

^a Number corresponds to the number in Figure 3; entries in bold have mascot scores greater than the significance threshold of 62 using custom databases or 79 for NCBI databases. ^b NCBI protein accession numbers (those that were removed from NCBI after spot identification are noted in parentheses). ^c Protein families were obtained by searching NCBI fasta sequences <http://pfam.sanger.ac.uk/>.²⁸ ^d M_m(molecular mass) and pI (isoelectric point) were predicted from the NCBI fasta sequence by ProtParam (ExPASy Proteomics tools). ^e Glean numbers were obtained during the annotation of the *Tribolium* m genome (1) and have a gut ranking as determined by the microarray analysis (see Supporting Information Table 3). ^f Reference 29. ^g Reference 30.

set from the microarray. In a biochemical study of the digestive proteases in the larval gut of *T. castaneum*, activities were similar to chymotrypsin and trypsin and were found mostly in the posterior midgut.⁹ In the present study, 26 and 21 of the 75 transcripts encoded presumptive chymotrypsin and trypsin, respectively; the remainder corresponded to inactive serine protease-like homologues. The finding that only a few trypsin and chymotrypsin proteins were identified through proteomics may indicate that these enzymes are expressed at low levels in the gut. Under normal dietary conditions, *T. castaneum* larvae use primarily cysteine peptidases for protein digestion.^{20,21} However, when challenged with cysteine peptidase inhibitors, larvae respond by shifting to increased serine peptidase activity.²² We speculate that the large number of gut-expressed serine proteases may be associated with an inhibitor compensation response. Essentially, larvae are “primed” to respond quickly to inhibitor cereal defenses.

A gene encoding a cellulase-like protein, GLEAN 15370, was expressed at low levels in the gut according to gut ranking and qRT-PCR. A cellulase from the honeybee, *Apis mellifera*, is similar to GLEAN 15370 and is expressed in salivary glands.²³ It was hypothesized that the honeybee cellulase is associated with the hypopharyngeal gland and participates in the digestion of the inner wall of pollen grains. Therefore, cellulase in *T. castaneum* larvae may aid in the digestion of cereal bran. However, data from qRT-PCR with GLEAN 15370 was inconclusive and did not indicate higher expression in the head, which may have been because of the developmental stage or larval diet.

One concern in biochemical or proteomic analysis of the gut is the inadvertent identification of proteins that are not of insect origin (e.g., are from the diet). Searches of plant databases with peptides from the proteomic study did not yield any positive matches (data not shown). The gut microarray analysis is

superior in this respect because oligonucleotides were specific to *T. castaneum* gene sequences, assuming that the material used for sequencing was not contaminated. Therefore, proteomic and genomic techniques combined increase the accuracy of tissue-specific protein identifications.

A potential problem with all tissue-specific analyses is obtaining clean gut tissues free of associated tissues, such as fat body, tracheae, and hemolymph. This is particularly a problem when using extremely sensitive detection methods. In this respect, we suspected that, contrary to the high gut ranking, GLEAN 05376, encoding arylphorin, was not a gut transcript. Transcript abundance in the gut was not supported by qRT-PCR data. Arylphorin is associated with the hemolymph and fat body in lepidopterans,^{24,25} although there is one report of a similar protein, hexamerin, in the lepidopteran midgut.²⁶ An alternative explanation of the high gut ranking for arylphorin in *T. castaneum* is that it is difficult to extract gut tissue completely devoid of fat body, where arylphorin and hexamerin are abundant. Therefore, it is possible that the increased signal for GLEAN 05376 in the microarray was due to gut samples contaminated by fat body or hemolymph. This also may be problematic for highly abundant proteins, such as the number of glycosyl hydrolases and chitin-associated proteins identified in the proteomic study of the gut, which instead may be associated with tracheae. As a precaution, the gut rankings and gut-specific proteomic identifications should be used as a guide and should be verified by additional experimental data.

Several hypothetical proteins with unknown functions were found, supported by microarray, MALDI-TOF, and/or ESI-MS/MS. Most are low molecular mass proteins with acidic isoelectric points, and a few have chitin-binding 3 domains.²⁷ Some of these hypothetical proteins have been found in other insect genomes, but their function is unknown.

Conclusions

We provide herein the first comprehensive evaluation of an insect gut lumen and epithelium using a combined genomic and proteomic approach. These results demonstrate that proteins from the *T. castaneum* larval gut can be identified by complementary approaches using genomic and proteomic techniques, each with advantages and disadvantages. The acquisition of the *Tribolium* genome was the most critical factor for success of this project. Without the genome sequences, oligonucleotide microarrays could not have been produced and fewer proteomic identifications with significant MASCOT scores would have been obtained.

The microarray analysis provided a quick screen of putative gut transcripts, where fluorescence intensity was used as an indicator of relative expression. Proteins identified in the gut tissue by both MALDI-TOF and ESI-MS/MS generally correlated with higher gut rankings, providing additional evidence of gut origin. However, previous biochemical data also were invaluable in predictions of gut specificity.

Many of the previously intangible research questions regarding the *T. castaneum* larval gut can now be addressed with additional genetic, genomic, and proteomic data. These data will be a dynamic resource for researchers to compare the gut transcriptome and proteome of other coleopteran insects. We also are using the data to evaluate changes in the *T. castaneum* gut in response to various biotic and abiotic challenges for the development of new control products for stored product pests.

Acknowledgment. We thank Kent Shelby and Steve Garczynski for their thoughtful comments, and Richard

Beeman, Jeff Lord, and Yoonseong Park for cooperation to develop the microarray chip. Support for the project was made possible in part by a KSU Targeted Excellence Award (JMT) and an NSF Major Research Instrumentation Program Grant No. 0521587 (J.M.T.). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Supporting Information Available: Table 3. Intensity of each labeled oligonucleotide in the Tc genome chip (GLEAN#) for each *Tribolium castaneum* pooled larval gut sample (biological replicates 1, 2a, or 3, or technical replicate 2b). Intensities were averaged for all four samples (AVERAGE), and relative rankings were made by dividing each averaged intensity for each GLEAN by the largest average intensity value (78 561) multiplied by 100. NCBI numbers corresponding to each GLEAN are provided for the gene (NCBI # Gene) or protein (NCBI # Protein). Proteomic support, either ESI-MS/MS and/or MALDI-TOF, is cross-referenced to Table 2 and Table 4. GenePix parameters are given for each gut biological and technical replicate on a separate spreadsheet.

Table 4. NCBI protein identifications (accession numbers and annotations) from *Tribolium castaneum* larval gut extract analyzed by 2D nano-HPLC and ESI-MS/MS. IDs are based on significant MASCOT scores (with the sequence coverage percentage and number of matched peptides). Gut rankings from the microarray (Table 3) and those identified through MALDI-TOF (Table 2) are provided for corresponding GLEANS.

Table 5. Predicted peptidases from the C1 cysteine peptidase family, clan CA, in the *T. castaneum* genome (1), with their relative gut rankings. Bolded entries have a gut ranking ≥ 3 .

Figure 4. Coomassie-stained 1D SDS-PAGE of *T. castaneum* larval gut proteins (lane Tc); molecular mass markers on the left.

Table 5. Functional categories of the top ranked genes (≥ 3) in the *Tribolium castaneum* larval genome. Categories of Molecular Function (MF) obtained with GeneSifter software and are presented as a percentage of the total (GeoSpiza, Seattle, WA). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) *Tribolium* Genome Sequencing Consortium. The first genome sequence of a beetle, *Tribolium castaneum*, a model for insect development and pest biology. *Nature* **2008**, *452*, 949–955.
- (2) Prabhakar, S.; Chen, M.-S.; Elpidina, E. N.; Vinokurov, K. S.; Smith, C. M.; Oppert, B. Molecular characterization of digestive proteinases and sequence analysis of midgut cDNA transcripts of the yellow mealworm, *Tenebrio molitor* L. *Insect Mol. Biol.* **2007**, *16*, 455–468.
- (3) Elpidina, E. N.; Goptar, I. A. Digestive peptidases in *Tenebrio molitor* and possibility of use to treat celiac disease. *Entomol. Res.* **2007**, *37*, 139–147.
- (4) Terra, W. R.; Ferreira, C.; Bastos, F. Phylogenetic consideration of insect digestion. Disaccharidases and the spatial organization of digestion in the *Tenebrio molitor* larvae. *Insect Biochem.* **1985**, *15*, 443–449.
- (5) Thie, N. M. R.; Houseman, J. G. Cysteine and serine proteolytic activities in larval midgut of yellow mealworm, *Tenebrio molitor* L. (Coleoptera: Tenebrionidae). *Insect Biochem.* **1990**, *20*, 741–744.
- (6) Terra, W. R.; Ferreira, C. Insect digestive enzymes: properties, compartmentalization and function. *Comp. Biochem. Physiol.* **1994**, *109B*, 1–62.
- (7) Vinokurov, K. S.; Elpidina, E. N.; Oppert, B.; Prabhakar, S.; Zhuzhikov, D. P.; Dunaevsky, Y. E.; Belozersky, M. A. Diversity of digestive proteinases in *Tenebrio molitor* (Coleoptera: Tenebrionidae) larvae. *Comp. Biochem. Physiol.* **2006a**, *145B*, 126–137.

- (8) Vinokurov, K. S.; Elpidina, E. N.; Oppert, B.; Prabhakar, S.; Zhuzhikov, D. P.; Dunaevsky, Y. E.; Belozersky, M. A. Fractionation of digestive proteinases from *Tenebrio molitor* (Coleoptera: Tenebrionidae) larvae and role in protein digestion. *Comp. Biochem. Physiol.* **2006b**, *145B*, 138–146.
- (9) Vinokurov, K. S.; Elpidina, E. N.; Zhuzhikov, D. P.; Oppert, B.; Kodrik, D.; Sehna, F. Digestive proteolysis organization in two closely related tenebrionid beetles: red flour beetle (*Tribolium castaneum*) and confused flour beetle (*Tribolium confusum*). *Arch. Insect Biochem. Physiol.* **2009**, *70*, 254–279.
- (10) Park, Y.; Aikins, J.; Wang, L. J.; Beeman, R. W.; Oppert, B.; Lord, J.; Brown, S. J.; Lorenzen, M. D.; Richards, S.; Weinstock, G.; Gibbs, R. Analysis of transcriptome data in the red flour beetle *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* **2008**, *38*, 380–386.
- (11) Rozen, S.; Skaletsky, H. J. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Krawetz, S., Misener, S., Eds.; Humana Press: Totowa, NJ, 2000; pp 365–386.
- (12) Fabrick, J.; Oppert, C.; Lorenzen, M. D.; Morris, K.; Oppert, B.; Jurat-Fuentes, J. L. A novel *Tenebrio molitor* cadherin is a functional receptor for *Bacillus thuringiensis* toxin Cry3Aa. *J. Biol. Chem.* **2009**, *284*, 18401–18410.
- (13) Altincicek, B.; Knorr, E.; Vilcinskas, A. Beetle immunity: Identification of immune-inducible genes from the model insect *Tribolium castaneum*. *Dev. Comp. Immunol.* **2008**, *32*, 585–595.
- (14) Wang, L.; Duman, J. G. A thaumatin-like protein from larvae of the beetle *Dendroides canadensis* enhances the activity of anti-freeze proteins. *Biochemistry* **2006**, *45*, 1278–1284.
- (15) Cristofaletti, P. T.; Ribeiro, A. F.; Terra, W. R. The cathepsin L-like proteinases from the midgut of *Tenebrio molitor* larvae: Sequence, properties, immunocytochemical localization and function. *Insect Biochem. Mol. Biol.* **2005**, *35*, 883–901.
- (16) Koiwa, H.; Shade, R. E.; Zhu-Salzman, K.; D'Urzo, M. P.; Murdock, L. L.; Bressan, R. A.; Hasegawa, P. M. A plant defensive cystatin (soyacystatin) targets cathepsin L-like digestive cysteine proteinases (DvCALs) in the larval midgut of western corn rootworm (*Diabrotica virgifera virgifera*). *FEBS Lett.* **2000**, *471*, 67–70.
- (17) Bown, D. P.; Wilkinson, H. S.; Jongasma, M. A.; Gatehouse, J. A. Characterisation of cysteine proteinases responsible for digestive proteolysis in guts of larval western corn rootworm (*Diabrotica virgifera*) by expression in the yeast *Pichia pastoris*. *Insect Biochem. Mol. Biol.* **2004**, *34*, 305–320.
- (18) de Oliveira Neto, O. B.; Batista, J. A.; Rigden, D. J.; Franco, O. L.; Fragoso, R. R.; Monteiro, A. C.; Monnerat, R. G.; Grossi-De-Sa, M. F. Molecular cloning of a cysteine proteinase cDNA from the cotton boll weevil *Anthonomus grandis* (Coleoptera: Curculionidae). *Biosci. Biotechnol. Biochem.* **2004**, *68*, 1235–1242.
- (19) Brito, L. O.; Lopes, A. R.; Parra, J. R.; Terra, W. R.; Silva-Filho, M. C. Adaptation of tobacco budworm *Heliothis virescens* to proteinase inhibitors may be mediated by the synthesis of new proteinases. *Comp. Biochem. Physiol.* **2001**, *128B*, 365–375.
- (20) Oppert, B.; Morgan, T. D.; Culbertson, C.; Kramer, K. J. Dietary mixtures of cysteine proteinase and serine proteinase inhibitors exhibit increased toxicity toward the red flour beetle *Tribolium castaneum*. *Comp. Biochem. Physiol.* **1993**, *105C*, 379–385.
- (21) Oppert, B.; Morgan, T. D.; Hartzler, K.; Lenarcic, B.; Galesa, K.; Brzin, J.; Turk, V.; Yoza, K.; Ohtsubo, K.; Kramer, K. J. Effects of proteinase inhibitors on growth and digestive proteolysis of the red flour beetle, *Tribolium castaneum*(Herbst) (Coleoptera: Tenebrionidae). *Comp. Biochem. Physiol.* **2003**, *134C*, 481–490.
- (22) Oppert, B.; Morgan, T. D.; Hartzler, K.; Kramer, K. J. Compensatory proteolytic responses to dietary proteinase inhibitors in the red flour beetle, *Tribolium castaneum*(Herbst) (Coleoptera: Tenebrionidae). *Comp. Biochem. Physiol.* **2005**, *140C*, 53–58.
- (23) Kunieda, T.; Fujiyuki, T.; Kucharski, R.; Foret, S.; Ament, S. A.; Toth, A. L.; Ohashi, K.; Takeuchi, H.; Kamikouchi, A.; Kage, E.; Morioka, M.; Beye, M.; Kubo, T.; Robinson, G. E.; Maleszka, R. Carbohydrate metabolism genes and pathways in insects: insights from the honey bee genome. *Insect Mol. Biol.* **2006**, *15*, 563–576.
- (24) Telfer, W. H.; Kunkel, J. G. The function and evolution of insect storage hexamers. *Ann. Rev. Entomol.* **1991**, *36*, 205–228.
- (25) Ryan, R. O.; Schmidt, J. O.; Law, J. H. Arylphorin from the haemolymph of the larval honey bee. *Insect Biochem.* **1984**, *14*, 515–520, 1984.
- (26) Ma, G.; Roberts, H.; Sarjan, M.; Featherstone, N.; Lahnstein, J.; Akhurst, R.; Schmidt, O. Is the mature endotoxin Cry1Ac from *Bacillus thuringiensis* inactivated by a coagulation reaction in the gut lumen of resistant *Helicoverpa armigera* larvae? *Insect Biochem. Mol. Biol.* **2005**, *35*, 729–739, 2005.
- (27) Oppert, B.; Elpidina, E. Old problems, new approaches. In *4th Meeting of Comparative Physiologists & Biochemists in Africa - Mara 2008*; Morris, S., Vosloo, A., Eds.; Medimond: Bologna, Italy, 2008; pp 173–182.
- (28) Finn, R. D.; Mistry, J.; Schuster-Böckler, B.; Griffiths-Jones, S.; Hollich, V.; Lassmann, T.; Moxon, S.; Marshall, M.; Khanna, A.; Durbin, R.; Eddy, S. R.; Sonnhammer, E. L. L.; Bateman, A. Pfam: clans, web tools and services. *Nucleic Acids Res.* **2006**, *34* (Database issue), D247–D251.
- (29) Zhu, Q.; Arakane, Y.; Banerjee, D.; Beeman, R. W.; Kramer, K. J.; Muthukrishnan, S. Domain organization and phylogenetic analysis of the Chitinase-like family of proteins in three species of insects. *Insect Biochem. Mol. Biol.* **2008**, *38*, 452–466.
- (30) Arakane, Y.; Muthukrishnan, S.; Beeman, R. W.; Kanost, M. R.; Kramer, K. J. Laccase 2 is the phenoloxidase gene required for beetle cuticle tanning. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11337–11342.

PR900168Z