To the Graduate Council:

I am submitting herewith a thesis written by Jonathan Duran Willis entitled "Identification and characterization of novel cellulases from *Dissosteira carolina* (Orthoptera: Acrididae) and molecular cloning and expression of an endo-beta-1,4-glucanase from *Tribolium castaneum* (Coleoptera: Tenebrionidae)." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master's of Science, with a major in Entomology and Plant Pathology.

Juan-Luis Jurat-Fuentes, Major Professor

We have read this thesis
and recommend its acceptance:

William Klingeman

Reza Hajimorad

Ranjan Ganguly

Accepted for the Council:

Carolyn R. Hodges, Vice Provost and
Dean of the Graduate School

(Original Signatures are on file with official student records.)

Identification and characterization of novel cellulases from *Dissosteira carolina* (Orthoptera: Acrididae) and molecular cloning and expression of an endo-beta-1,4-glucanase from *Tribolium castaneum* (Coleoptera: Tenebrionidae)

A Thesis Presented for
the Master of Science
Degree
The University of Tennessee, Knoxville

Jonathan D. Willis
December 2009

Dedication

This work is dedicated to my parents, James and Pam, and my heart Diane Larson who stood by, counseled, and supported me throughout my graduate career. Without my parent's guidance growing up I never would have become the person I am today. Diane aided me through her love and devotion making everyday brighter.

Acknowledgements

I gratefully appreciate Dr. Juan-Luis Jurat-Fuentes for financial support, leadership, and most important the patience needed to complete my studies. I appreciate the support from my committee members Dr. William Klingeman, Dr. Reza Hajimorad, and Dr. Ranjan Ganguly. I thank all the members of the Jurat-Fuentes lab for support and discussions in furthering research endeavors. I also thank Dr. Brenda Oppert for her valuable research knowledge and materials. I acknowledge the Southeastern Sun Grant Center for the financial support to pursue novel biocatalysts from insects to improve production of alternative fuels.

Abstract

Cellulosic ethanol holds great potential as biofuel due to its sustainability and renewability, yet recalcitrance of cellulosic feedstocks prevents cost-efficient ethanol production. Enzymatic catalysis of lignocellulosic biomass has the greatest biotechnological potential for cost reductions to the production process. Even though numerous cellulolytic enzymes have been identified in bacteria, plant, and fungi, insects remain as a fairly unexplored prospecting resource. Many insects, either via endogenously or symbiotically derived enzymes, use cellulose as substrate for their energetic needs. Novel cellulases from insects may have the potential to be more efficient than alternative enzymes in the conversion of cellulose to fermentable sugars due to their optimized activity in the highly reducing and extreme pH conditions found in some insect digestive systems.

In this work we present data characterizing cellulolytic activity in the grasshopper *Dissosteira carolina L.* (Orthoptera: Acrididae) and the red flour beetle, *Tribolium castaneum* Herbst (Coleoptera: Tenebrionidae). After a screening for cellulolytic activity in insect populations from the East Tennessee region, *D. carolina* was selected due to relatively high cellulolytic activity compared to documented effective insect cellulolytic species. Cellulolytic activity in digestive fluids from gut and head from juvenile and adult stages of *D. carolina* was measured and an active cellulolytic protein profile demonstrated comparable activities amongst life stages. Partial protein sequences that match those identified from insect and microbial cellulases were obtained from purified 43-kDa and 45-kDa cellulases from *D. carolina* head digestive fluids. Although unsuccessful, attempts were made to purify and clone these enzymes

for recombinant expression. Our research on *D. carolina* is the first report on the purification of endoglucanase activity in a grasshopper species.

Availability of the *T. castaneum* genome allowed for homology searches using reported insect cellulases to identify a predicted cellulase. We cloned the full-length cDNA for this enzyme and named it TcEG1 (for *T. castaneum* endo-glucanase-1). TcEG1 was heterologously expressed in bacterial and insect cell culture systems and its activity against cellulose substrates and thermostability measured. Cloning of a cellulase gene from *T. castaneum* adds to the collection of reported insect cellulases and demonstrates the advantage of using genomic resources for protein discovery.

Table of Contents

Contents

List of Tables

List of Figures

Chapter 1

Review: Discovery and characterization of novel cellulolytic enzymes from insects

A revised version of this chapter titled: “Methods for discovery and characterization of cellulolytic enzymes from insects” has been submitted for publication as part of a special issue on “Insects and Biofuels” in the journal *Insect Science*. Authorship in this paper is as follows: Jonathan D. Willis, Cris Oppert and Juan Luis Jurat-Fuentes. My contributions to this paper include (1) most of the collection and review of published data, and (2) part of the writing.

Abstract:

Cellulose has been identified as a potential resource for ethanol biofuel production due to its sustainability and abundance. However, lignocellulosic biomass is difficult to convert to glucose due to its recalcitrance. Insects have evolved through independent and symbiotic relationships the ability to use cellulose as a sugar source for energy needs. The advancement of molecular biology, genomic, and proteomic research tools has allowed new insights into insect digestion of cellulose. This information is highly relevant to the design of improved industrial processes of biofuel production and to identify potential new targets for development of insecticides. This review describes the diverse methodologies used to detect, quantify, purify, clone, and express cellulolytic enzymes from insects, as well as their advantages and limitations.

Introduction:

Current world energy needs demand the development of industrial-scale processes for the production of biofuels as an economically viable and environmentally sound alternative to finite fossil fuels. Biofuels are defined as any liquid, solid, or gaseous fuel produced from a renewable biological resource (U.S.DOE 2006). In the U.S., lignocellulosic ethanol has been suggested as the desirable biofuel carrying the qualities of being a renewable and sustainable resource. Cost-

efficient production of cellulosic ethanol from non-food crops such as switchgrass has become a key factor to increase ethanol sources without diminishing our food stock supplies for fuel resources (U.S.DOE 2006).

Cellulose is one of the main components of the plant cell wall and it is associated with hemicellulose and lignin forming a recalcitrant structure preventing enzyme access to sugar components. Cellulose is structured as a linear biopolymer crystallized and stacked upon itself to form fibrous bundles (Clarke 1997). The most basic unit of cellulose is cellobiose, which is composed of two glucose residues joined and inverted 180° forming a β-(1,4) linkage. High heat and dilute/concentrated acid treatments are used to disrupt the lignin barrier releasing cellulose(Wyman, Dale et al. 2005; U.S.DOE 2006; Lynd, Laser et al. 2008; Somerville 2008). Hydrolysis of cellulose requires multiple glycoside hydrolase (GH) enzymes working synergistically to obtain free glucose for fermentation into ethanol. Chemical methods to convert cellulose into fuel to produce synthesis gas or biohydrogen have been proposed, but currently have issues with efficiency, harmful waste products, and require deeper research efforts (Demirbas 2007). Enzymatic degradation of cellulose for production of ethanol biofuel is considered the pathway with the greatest potential for improvement and cost reduction (Wyman, 1999). Current estimates suggest that reducing the cellulase enzyme amounts during ethanol production by half could decrease processing costs by up to 13% (Lynd, Laser et al. 2008). Currently available enzymes used for cellulosic ethanol are of fungal and bacterial origin (U.S.DOE 2006). Limitations of these current cellulolytic enzymes include inability to withstand high temperatures, low specific activity, generation of inhibitory agents during

hydrolysis, and susceptibility to pH changes (Mousdale 2008). Molecular engineering and bioprospecting efforts should provide novel enzymes with higher specific activity and with lower susceptibility to inhibitors.

Insects have evolved effective strategies to utilize lignocellulosic substrates as sources of energy, which makes them an optimal resource for prospecting for novel cellulolytic enzymes. Extensive efforts have characterized lignocellulose degradation in termites (Cleveland 1924; Watanabe, Noda et al. 1998; Khademi, Guarino et al. 2002; Scharf 2008), yet there is a lack of information on the cellulolytic systems in alternative insect species. The insect digestive system displays conditions that are optimized for degradation of plant biomass, including reducing conditions and high alkaline pH in some insect orders (Lehane and Billingsley 1996). Digestive conditions of insects are similar to that of bioreactors and serve as bioprospecting representatives for efficient enzymes for the hydrolysis of plant biomass. Evidence for the prospecting potential in insects is most notably found in species of termites, which are known to live almost exclusively on substrates with high lignocellulosic content.

Originally, cellulolytic activity in insects was attributed to symbiotic gut bacteria (Cleveland 1924). However, the first cellulase gene of insect origin was discovered and cloned from *Reticulitermes speratus* (Kolbe) (Isoptera: Rhinotermitidae) in the late eighties (Watanabe, Noda et al. 1998). Currently, cellulolytic enzymes have been identified or detected in insects belonging to ten different orders, including Blattodea, Coleoptera, Diptera, Hymenoptera, Isoptera, Lepidoptera, Orthoptera, Phasmatodea, Plecoptera, and Trichoptera (Martin 1983;

Martin 1991; Marana, Terra et al. 1995; Cazemier, Huub et al. 1997.). A majority of this research has only been at the enzymatic detection level, although insect-made cellulases have been cloned from *Apriona germari* (Hope) (Coleoptera: Cerambycidae) (Lee, Kim et al. 2004; Lee, Lee et al. 2005; Wei, Lee et al. 2006), *Psacothera hilaris* (Pascoe) (Coleoptera: Cerambycidae) (Sugimura, Watanabe et al. 2003), *Tenebrio molitor* (L.) (Coleoptera: Tenebrionidae) (Ferreira, Marana et al. 2001), *Mastotermes darwiniensis* (Froggart) (Isoptera: Mastotermitidae) (Li, Frohlich et al. 2003), *Neotermes koshunensis* (Shiraki) (Isoptera: Kalotermitidae) (Ni, Tokuda et al. 2007), *Reticulitermes flaviceps* (Kollar) (Isoptera: Rhinotermitidae) (Zhou, Smith et al. 2007), and *Teleogryllus emma* (Ohmachi et Matsuura) (Orthoptera: Gryllidae) (Kim, Choo et al. 2008). In this review, the specific methodologies used to detect, characterize and clone cellulolytic enzymes from insect systems will be addressed.

Biodegradation of cellulose to glucose: cellulases

The term "cellulase" defines the cooperative activity of three glycoside hydrolases (GH) enzymes: endo-β-1,4-glucanases (EG; EC. 3.2.1.4), exo-β-1,4-cellobiohydrolases (CBH; EC. 3.2.1.91), and β-glucosidases (EC. 3.2.1.21). EG enzymes work by random cleavage of β-1,4 bonds in the internal portions of cellulose strands to reduce the degree of polymerization of the cellulose chain into smaller subunits. CBH enzymes remove subunits at the non-reducing the cellulose chain ends of cellulose, releasing either cellobiose or glucose. Cellobiose is hydrolyzed to glucose by EC enzymes (Clarke 1997). Due to inhibition feedback from cellobiose onto EG enzymes, β-glucosidase activity is important for complete degradation of cellulose (Holtzapple, Cognata et al. 1990; Gruno, Valjamae et al. 2004).

Through DNA and protein sequence homology, 115 GH families (GHF) have been identified (Henrissat 1991; Henrissat and Bairoch 1993; Davies and Henrissat 1995; Henrissat and Davies 1997). Due to their diversity, cellulolytic enzymes fall into numerous GHF families. For example, EGs belong to GHF 5,6,7,8,9,12,44,45,48,51,61,and 74, while CBHs belong to GHF 5, 6, 9, and 48; and β glucosidases belong to GHF 1, 3, 9 (CAZY database).

Most GHs have a catalytic site containing Asp or Glu amino acids to hydrolyze glycosidic bonds. Asp or Glu will serve as either the nucleophile or proton donor and the specific role of Asp versus Glu is different for each GHF. For GHs active on cellulose, a separate important region termed cellulose binding domain (CBD) or cellulose binding module (CBM) is present. This domain is separated from the catalytic domain by a linker sequence to prevent steric hindrances and allowing cellulose to be hydrolyzed. In some instances, a single GH may have multiple catalytic sites which can place it into different GH families (Davies and Henrissat 1995; Bayer, Chanzy et al. 1998).

For insect cellulases, complete sequence data has revealed GHs belonging to only five families: GHF5, GHF7, GHF9, GHF10, and GHF45. Cellulases belonging to the GHF9 have been found in termites (Watanabe and Tokuda 2001) and *T. emma* (Kim, Choo et al. 2008). This family consists of EG, CBH, and β-glucosidases from prokaryote and eukaryote organisms and display characteristic 6 alpha helices in their secondary structure. The catalytic domain is joined to the cellulose binding domain by a linker sequence and is made of conserved aspartic acid residues nearer the N-terminus and a glutamine at the C-terminus. NtEG1, a GHF9 EG from

Nasutitermes takasagoensis (Shiraki) (Isoptera: Termitidae) (Khademi, Guarino et al. 2002) was the first insect cellulase for which the three dimensional structure was resolved, demonstrating the characteristic alpha helical barrel and catalytic domain typical of enzymes in GHF9.

An EG belonging to GHF5 was found in *P. hilaris* and *A. germari* (Sugimura et al. 2003, Wei et al. 2006). GHF5 glycoside hydrolases have a crystal structure composed of eight alpha helices and a beta sheet (Ducros, Czjzek et al. 1995). An EG belonging to GHF 45 was found in *A. germari (Lee, Kim et al. 2004; Lee, Lee et al. 2005)* and *Phaedon cochleariae* (Fabricius) (Coleoptera: Chrysomelidae) (Girard and Jouanin 1999). Unique features of enzymes belonging to GHF45 are the catalytic site (–TTTRYWDCCKPSC–), and 14 cysteines conserved throughout the sequence (Girard and Jouanin 1999; Bourne and Henrissat 2001).

Current evidence suggests that insect cellulolytic systems contain a mixture of GH from multiple families as found in *A. germari* and multiple termite species (Scharf 2008). Mixtures of GHF enzymes may suggest to presence of both endogenous and symbiont-derived enzymes for efficient and complete bioconversion in these systems.

Insect cellulolytic activity detection:

Carboxymethylcellulose (CMC) is a form of cellulose modified with carboxymethyl groups added to improve water solubility. This cellulase substrate is the most commonly documented for detection of EG activity due to its ease of use and solubility. Increased solubility in water

allows for in solution assays or incorporation into agar or polyacrylamide gels for electrophoretic detection and separation of cellulolytic enzymes (Schwarz, Bronnenmeier et al. 1987). Additionally, in-solution assays with CMC as substrate can be performed in microplates, reducing methodological complexity when compared to previous IUPAC test tube methods (Xiao, Storms et al. 2005). Degradation of CMC in solution can be monitored by the detection of generated reducing sugars using the 3,5-dinitrosalicylic acid (DNSA) or tetrazolium blue methods (Miller 1959, Jue and Lipke 1985). The most common buffers used for these assays have a pH of 6, usually based on sodium citrate or sodium acetate, although pH can also be altered to test pH range for activity of EGs. Table 1 provides a listing of insects with detected EG activity and demonstrates the popularity of CMC as the dominant substrate used in conjunction with DNSA.

Table 1: Insects with documented β-1,4-endoglucanase activity and the detection methodology used. Where reported, purification and size information are provided.

Order (Family) *Genus species*	***Substrate (s)***	***Detection Method***	***Purification***	***Size***	***Reference***
Blattodea					
(Blaberidae)					
Panesthia cribrata	CMC	DNSA	Partial column chromatography	53.6, 48.8kDa	(Scrivener 1993)
Pycnoscelus surinamensi	CMC	DNSA			(Cazemier, Huub et al. 1997.)
(Blattidae)					
Blaberus fuscus	CMC	DNSA			(Cazemier, Huub et al. 1997.)
Periplaneta americana	CMC	DNSA			(Cazemier, Huub et al. 1997.)
Periplaneta americana	CMC, Filter Paper	DNSA			(Gijzen, van der Drift et al. 1994)
Periplaneta australasia	CMC	DNSA			(Cazemier, Huub et al. 1997.)
Coleoptera					
(Buprestidae)					
Agrilus planipennis	CMC	CMC plates			(Vasanthakumar, Handelsman et al. 2008)
(Cerambycidae)					

Anoplophora glabripennis	CMC, MCC	DNSA, Zymogram			(Xiao-juan, Xiong-fei et al. 2008)
Apriona germari	CMC	DNSA, CMC Plates	cDNA library SF9 cells	29, 36, 47 kDA	(Lee, Kim et al. 2004; Lee, Lee et al. 2005; Wei, Lee et al. 2006)
Hylotrypus bajules	CMC	DNSA			(Cazemier, Huub et al. 1997.)
Psacothera hilaries	CMC, MCC	Tetrazolium blue	Zymogram purification	47 kDa	(Sugimura, Watanabe et al. 2003)
Saperda vestita	CMC, Filter Paper	CMC Plate, FPA Visible degradation			(Delalibera, Handelsman et al. 2005)
(Chrysomelidae)					
Aulacophora foveicollis	CMC	Zymogram, DNSA	Zymogram purification	NR	Sami 2008
(Curculionidae)					
Dendroctonus frontalis	CMC, Filter Paper	CMC Plate, FPA Visible degradation			(Delalibera, Handelsman et al. 2005)
Ips pini	CMC, Filter Paper	CMC Plate, FPA Visible degradation			(Delalibera, Handelsman et al. 2005)
(Scarabeidae)					
Pachnoda marginata	CMC	DNSA			(Cazemier, Huub et al. 1997.)
Diptera					
(Psychodidae)					
Phlebotomus papatasi	OBRHC, cellulose azure				(Jacobson and Schlein 1997)
(Tipulidae)					
Tipula abdominalis	CMC	DNSA			(Walters 1991)
Hymenoptera					
(Apidae)					
Apis mellifera	NR	Whole Genome annotation	NR	1497bp	(Kunieda, Fujiyuki et al. 2006)
Isoptera					
(Kalotermitidae)					
Cryptotermes pingyangensis	CMC	Arsenomolybdate chromotography			(Mo, Yang et al. 2004)
Neotermes koshunensis	CMC, MCC	Tetrazolium blue			(Tokuda, Lo et al. 2005)
(Macrotermitidae)					
Odontotermes formosanus	CMC	Arsenomolybdate chromatography, Somoygi-Nelson, Tetrazolium blue	Anion column purification	80 kDa	(Yang, J. et al. 2003; Mo, Yang et al. 2004; Tokuda, Lo et al. 2005)
Mastotermes darwiniensis	CMC	DNSA, Zymogram	Degenerate PCR, RACE	36kDa	(Cazemier, Huub et al. 1997.; Li, Frohlich et al. 2003)
(Rhinotermitidae)					
Coptotermes formosanus	CMC	Arsenomolybdate chromatography, tetrazoium blue			(Mo, Yang et al. 2004; Tokuda, Lo et al. 2005)
Reticulitermes flaviceps	CMC, pNPG, Beechwood Xylose	Arsenomolybdate chromatography, DNSA	cDNA library	48.7kDa, 35.7, 31.8, 34.5	(Mo, Yang et al. 2004; Zhou, Smith et al. 2007)
Reticulitermes leptomandiublaris	CMC	Arsenomolybdate chromotography			(Mo, Yang et al. 2004)

Reticulitermes speratus	CMC, MCC	Tetrazolium blue	Phage library	448 AA	(Watanabe, Noda et al. 1998; Tokuda, Lo et al. 2005)
(Termitidae)					
Nasutitermes takasagoensis	CMC	Glucose-oxidase-mutarotase, tetrazolium blue, Zymogram	Column chromatography	45kDa	(Tokuda, Watanabe et al. 1997; Tokuda, Lo et al. 2005; Tokuda and Watanabe 2007)
Nasutitermes walkeri	MCC, CMC	Zymogram, tetrazolium blu	NR	45kDa	(Tokuda and Watanabe 2007)
(Termopsidae)					
Hodotermopsis sjostedti	CMC, MCC	Ttetrazolium blue			Tokuda et al 2005
Lepidoptera					
(Papilionidae)					
Parnassius apollo ssp. frankenbergeri	CMC	DNSA			(Nakonieczny, Michalczyk et al. 2006)
***(*Saturniidae)**					
Philosamia ricini	CMC	Glucose-Oxidase Peroxidase			(Pant and Ramana 1989)
Orthoptera					
(Acrididae)					
Schistocerca gergaria	CMC	DNSA			(Cazemier, Huub et al. 1997.)
***(*Gryllidae)**					
Acheata domesticus	CMC	DNSA			(Cazemier, Huub et al. 1997.)
Teleogryllus emma	CMC	DNSA	cDNA library SF9 cells	47 kDa	(Kim, Choo et al. 2008)
Plecoptera					
(Peltoperlidae)					
Peltoperla arcuata	CMC	DNSA			(Walters 1991)
(Pternoarcidae)					
Allonarcys proteus	CMC	DNSA			(Walters 1991)
Phasmatodea					
(Phasmatidae)					
Eurycantha calcarata	CMC	DNSA			(Cazemier, Huub et al. 1997.)
Trichoptera					
(Limnephilidae)					
Pycnopsyche spp.	CMC	DNSA			(Walters 1991)

Agar plates infused with CMC have been used to localize EG activity in guts of *Rhagium inquisitor* (L.) (Coleoptera: Cerambycidae) (Zverlov, Holl et al. 2003) and as early screening tool for cellulolytic symbionts from *Saperda vestita* (Say) (Coleoptera: Cerambycidae), *Ips pini* (Say) (Coleoptera: Curculionidae), and *Dendroctonus frontalis* (Zimmerman) (Coleoptera: Curculionidae) (Delalibera, Handelsman et al. 2005). In this method, plates are incubated with digestive fluids and activity detected by staining CMC with Congo red dye. After washing with high salt (1M NaCl), regions where CMC has been degraded are observed as clear areas. The diameter of this activity area has been utilized as a relative quantitative measurement to compare activity of heterologously expressed enzymes from *Coptotermes formosanus* (Shiraki) (Isoptera: Rhinotermitidae) (Zhang, Lax et al. 2009).

In a similar strategy, zymograms can be used to detect proteins with cellulolytic activity after separation in SDS-PAGE (Schwarz, Bronnenmeier et al. 1987). These zymograms are performed by separating proteins in digestive fluids by SDS-PAGE gels (Laemmli 1970) containing CMC, and staining for the presence of undigested CMC as for agar plate assays. The advantage of this method over activity assays in agar plates is that specific protein bands with cellulolytic activity can be visualized, and their molecular weight estimated. Because cellulolytic enzymes may degrade CMC during the electrophoretic run, it is important to partially denature these enzymes by heating at 70°C for 20 min. so that discrete activity bands, rather than smears, are visualized. Using this technique, a 32-kDa cellulase was detected in digestive fluids of *R. inquisitor* (Zverlov, Holl et al. 2003). Zymograms were also used to establish cellulolytic isozyme profiling in *Anoplophora glabripennis* (Motschulsky) (Coleoptera: Cerambycidae) larvae after feeding on different tree species (Xiao-juan, Xiong-fei et al. 2008).

Cellulase substrates including filter paper, microcrystalline cellulose (MCC), and cotton, have been used to determine total cellulolytic activity due to their low levels of chemical modification, which makes them resemble more closely natural cellulose substrates. A filter paper assay (FPA) demonstrated cellulase activity from symbionts in the hindgut of *Periplaneta americana* (L.) (Blattaria: Blattidae) (Gijzen, van der Drift et al. 1994). FPA was also used to qualitatively determine the existence of cellulase activity in gut symbionts isolated from *S. vestita*, *I. pini*, and *D. frontalis* (Delalibera, Handelsman et al. 2005).

MCC (or Avicel) is a microcrystalline form of cellulose of varying pore sizes, which due to its crystalline structure better represents cellulose found in plant biomass. MCC can be used in cellulase screening for a more realistic substrate, but its use is limited to assays in solution with continuous mixing due to its insolubility in water (Tokuda and Watanabe 2007). A microplate assay using MCC and FPA as substrates and DNSA as detection method was developed to generate high throughput data (Xiao, Storms et al. 2005) (Chundawat, Balan et al. 2008). MCC and FPA do not work exceptionally well in cases where cellobiose, accumulates, which is the product of the enzymatic reaction because this substrate inhibits EG activity. This inhibition limitation is overcome by addition of β-glucosidases to degrade cellobiose to glucose.

A variety of substrates are used for detecting and quantifying β-glucosidase enzymatic activity. Cellobiose is used to determine activity of β-glucosidases by measuring release of glucose using the DNSA assay (Chan et al. 1989). Salicin and helicin are β-glycosides containing a D-glucose residue that can be cleaved by β-glucosidases and detected using spectrophotometry using a glucose standard curve as reference. In addition, 4-Nitrophenyl β-D-glucopyranoside (pNPG) can be used for β-glucosidase activity detection by liberation of p-nitrophenol or by fluorescence

using 4-methyl umbelliferyl β-D-glucopyranoside (MUG) (Jacobson and Schlein 1997). Zymograms using MUG as substrate can also be used to detect β-glucosidase activity protein bands in insect digestive fluids (Bhatia, Mishra et al. 2002). Table 2 provides a list of insects prospected for β-glucosidase activity and the specific methodology used for detection of this activity.

CBHs are the most difficult of the cellulolytic enzymes to detect due to their inactivity against modified (water-soluble) substrates such as CMC (Clarke 1997). These enzymes have activity on micro-crystalline cellulose (MCC) and amorphous cellulose (MCC treated with phosphoric acid), but these two substrates are difficult to use due to their low solubility. Additionally, β-glucosidases have the potential to react to similar linkage sites due to their broad catalytic range resulting in a high false detection rate (Clarke 1997). Table 3 provides a list of insects in which CBH activity has been detected and the methods used for detection. Interestingly, the parasite *Leishmania major* (Ross) was found to produce EG and CBH enzymes in sand fly vectors (Jacobson and Schlein 1997). It has also been demonstrated that termites with gut flagellates tend to degrade MCC in the hindgut, while those lacking symbionts degrade this substrate in the midgut or not at all (Tokuda, Lo et al. 2005). Localized digestion of MCC to the hindgut suggests that CBH

Table 2: Insects with documented β-D-glucosidase activity and detection methodology used. Where reported, purification and protein size information are provided.

Order (Family) *Genus species*	**Substrate (s)**	**Cellulase Detection Method**	**Purifciation**	**Size**	**Reference**
Blattodea					
(Blaberidae)					
Panesthia cribrata	pNPG	p-nitrophenol	Partial column chromatography	53.6, 48.8kDa	(Scrivener 1993)
Gromphadorrhinna portentosa	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
(Blattidae)					
Blaberus fuscus	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
Periplaneta americana	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
Periplaneta australasia	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
(Blattodea)					
Pycnoscelus surinamensi	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
Coleoptera					
(Cerambycidae)					
Anoplophora glabripennis	Salicin	DNSA, Zymogram			(Xiao-juan, Xiong-fei et al. 2008)
Hylotrypus bajules	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
(Curcolinidae)					
Rhynchophorus palmarum	pNPG	p-nitrophenol	Anion, Gel Filtration, Hydrophobic chromatography	58 kDa	(Yapi, Gnakri et al. 2009)
(Scarabeidae)					
Pachnoda marginata	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
(Tenebrionidae)					
Tenebrio molitor	pNPG, laminarin	glucose curves	FPLC immunoprobe to cDNA library	59 , 64 kDA	(Ferreira, Marana et al. 2001)
Diptera					
(Nematocera)					
Ptychoptera paludosa	MUG	MUG			(Wolf, Zwick et al. 1997)
(Sciaridae)					
Rhynchosciara americana	pNPG, cellobiose, salicin	pNBG	Electrophoresis purified	106, 65kDa	(Ferreira and Terra 1983)
Isoptera					
(Kalotermitidae)					
Cryptotermes pingyangensis	Salicin	Arsenomolybdate chromotography			(Mo, Yang et al. 2004)
Neotermes koshunensis	Cellobiose	Glucose-oxidase-mutarotase	cDNA RACE, E.coli His Tag	60kDA	(Ni, Tokuda et al. 2007)
(Macrotermitdidae)					
Odontotermes formosanus	Salicin	Arsenomolybdate chromotography			(Mo, Yang et al. 2004)
(Mastermitidae)					
Mastotermes	pNPG	p-nitrophenol			(Cazemier, Huub et al.

darwiniensis					1997.)
(Rhinotermitidae)					
Coptotermes formosanus	Salicin	Arsenomolybdate chromotography			(Mo, Yang et al. 2004)
Reticulitermes flaviceps	Salicin	Arsenomolybdate chromotography			(Mo, Yang et al. 2004)
Reticulitermes leptomandiublaris	Salicin	Arsenomolybdate chromatography			(Mo, Yang et al. 2004)
(Termitidae)					
Nasutitermes takasagoensis	Cellobiose	Glucose-oxidase-mutarotase, tetrazolium blue	Column chromatography		(Tokuda, Watanabe et al. 1997)
Lepidoptera					
(Noctuidae)					
Anticarsia gemmatalis	pNPG, helicin, salicin, MUG, cellobiose	Glucose-Oxidase Peroxidase			(Yu 1989)
Heliothis zea	pNPG, helicin, salicin, MUG, cellobiose	Glucose-Oxidase Peroxidase			(Yu 1989)
Spodoptera frugiperda	pNPG, helicin, salicin, MUG, cellobiose	Glucose-Oxidase Peroxidase			(Yu 1989)
Trichoplusia ni	pNPG, helicin, salicin, MUG, cellobiose	Glucose-Oxidase Peroxidase			(Yu 1989)
(**Papilionidae)**					
Parnassius apollo ssp. Frankenbergeri	pNPG, Cellobiose	DNSA and p-nitrophenol			(Nakonieczny, Michalczyk et al. 2006)
Orthoptera					
(Acrididae)					
Abracis flavolineata	pNPG, Cellobiose, alkyyl-β-glucosidase	degradation	FPLC SE and Anion	82kDa	(Marana, Terra et al. 1995)
Camnula pellucida	Salicin	Chromotagram			(Davis 1963)
Locusta migratoria	pNPG, Cellobiose	degradation	Partial via gel electrophoresis	65kDa	(Morgan 1975)
Melanoplus bivittalus	Salicin	Chromotagram			(Davis 1963)
Melanoplus infantalis	Salicin	Chromotagram			(Davis 1963)
Melanoplus packardiee	Salicin	Chromotagram			(Davis 1963)
Melanoplus sanguinipes	Salicin	Chromotagram			(Davis 1963)
Schistocerca gergaria	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
(Gryllidae)					
Acheata domesticus	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)
Phasmatodea					
(Phasmatidae)					
Eurycantha calcarata	pNPG	p-nitrophenol			(Cazemier, Huub et al. 1997.)

Table 3: Insects with documented cellobiohydrolases and detection methodology used.

Order (Family) *Genus species*	**Substrate (s)**	**Cellulase Detection Method**	**Reference**
Coleoptera			
(Cerambycidae)			
Anoplophora glabripennis	MCC	DNSA,	(Xiao-juan, Xiong-fei et al. 2008)
Isoptera			
(Kalotermitidae)			
Neotermes koshunensis	MCC	Tetrazolium Blue	(Tokuda, Lo et al. 2005)
(Rhinotermitidae)			
Reticulitermes speratus	MCC	Tetrazolium Blue	(Tokuda, Lo et al. 2005)
Coptotermes formosanus	MCC	Tetrazolium Blue	(Tokuda, Lo et al. 2005)
(Termitidae)			
Odontotermes formosanus	MCC	Tetrazolium Blue	(Tokuda, Lo et al. 2005)
Nasutitermes takasgoensis	MCC	Tetrazolium Blue	(Tokuda and Watanabe 2007)
(Termopsidae)			
Hodotermopsis sjoestedti	MCC	Tetrazolium Blue	(Tokuda, Lo et al. 2005)
Diptera			
(Psychodidae)			
Phlebotomus papatasi	Ostazin brillian red hydroxyethylcellulose, cellulose azure, MUG		(Jacobson and Schlein 1997)

enzymes are being secreted by symbionts, while midgut degradation may be due to enzymes secreted by the insect cells. Insect cellulase purification, cloning and expression in heterologous systems:

As in the case of fungal and bacterial cellulases, insect cellulases are usually less abundant than non-cellulolytic proteins, making fractionation and subsequent screening essential for purification and characterizing of cellulolytic enzymes from raw insect samples. An additional hindrance of insect cellulases is that in some cases, hydrolases are produced in limited quantity, requiring additional effort to obtain the necessary amounts of insect samples when compared to cellulases propagated in bacterial or fungal cultures. Liquid chromatography has been readily used to purify and concentrate insect cellulolytic proteins. Using size exclusion and anion exchange chromatography, six EG and two β-glucosidases were purified from gut fluids of *Panesthia cribrata* (Saussure) (Blattodea: Blaberidae) (Scrivener, Slaytor et al. 1989). In *T. molitor* midguts, two highly similar β-glucosidases were purified using hydrophobic liquid chromatography separation (Ferreira, Marana et al. 2001). In most instances, multiple purification steps are necessary for enzyme purification. For instance, a purification protocol including anion exchange, size exclusion, and hydrophobic-interaction chromatography was needed to purify a β-glucosidase from gut fluids of *Rhynchophorus palmarum* (L.) (Coleoptera: Curculionidae) (Yapi, Gnakri et al. 2009). An alternative to liquid chromatography, zymograms were used to purify an EG enzyme from *P. hilaris* (Sugimura, Watanabe et al. 2003; Sami 2008) and *Aulacophora foveicollis* (L.) (Coleoptera: Chrysomelidae) (Sami 2008).

Once detected and identified, cloning of insect cellulases is pursued to allow for characterization in heterologous systems. EG and β-glucosidases of insect origin have been cloned using cDNA

libraries, EST libraries, or degenerate PCR to conserved or sequenced N-terminal regions (listed in Table 1 and 2). Currently, no CBH enzymes have been cloned from insects, probably due to the difficulties related to their detection and purification.

PCR cloning using primers designed to N-terminal sequences from purified cellulases has been successfully used to clone a full length EG protein from *P. hilaris* (Sugimura, Watanabe et al. 2003). Degenerate primers to conserved cellulase regions have also been successfully used to clone genes from *Mastotermes darwiniensis* (Froggart) (Isoptera: Mastermitidae) (Li, Frohlich et al. 2003; Schiott, De Fine Licht et al. 2008). Limitations of this approach include the amounts of pure protein needed for accurate sequencing, presence of contaminating proteins in sequence reads, and sequences that are too degenerate to design specific primers.

With the availability of whole insect genomes, the number of identified insect-derived cellulases is expected to increase. Examples include predicted cellulases identified in the *Apis mellifera* (L.) (Hymenoptera: Apidae) (Kunieda, Fujiyuki et al. 2006) and *Tribolium castaneum* (Herbst) (Coleoptera: Tenebrionidae) (Morris, Lorenzen et al. 2009) genomes. Further research is needed to demonstrate functionality and specificity of these cellulases. cDNA and EST libraries have also been used as a source of partial sequences of predicted cellulase that can be used in for Rapid Amplification of cDNA Ends (RACE) to generate full-length cDNA clones. Using this method, cellulolytic enzymes have been identified and cloned from *T. molitor* (Ferreira, Marana et al. 2001), *Neotermes koshunensis* (Shiraki) (Isoptera: Kalotermitidae) (Ni, Tokuda et al. 2007), *A. germari* (Lee, Kim et al. 2004; Lee, Lee et al. 2005; Wei, Lee et al. 2006), *T. emma* (Kim, Choo et al. 2008), and *Reticulitermes flavipes* (Kollar) (Isoptera: Rhinotermitidae) (Zhou, Smith et al. 2007). Even though these genomic approaches to cellulase identification allow for

rapid cloning, they are usually expensive and may result in identification of enzymes with low activity or with a secondary role for cellulose digestion in the insect.

Conclusion/Insights from insect cellulases and their applications to biofuels

Optimization of cellulase mixtures for biofuel production will be dependent on the type of feedstock and pre-treatment used. Cellulolytic enzymes being used in bioreactors for ethanol production would optimally need to retain activity through the high heat and acidic pretreatment conditions used to make the cellulose in the lignocellulosic biomass available. However, because many of the microbial and fungal cellulolytic enzymes have optimal activity in the pH 4-6 range and temperatures around 50°C (Clarke 1997), research has been focused on optimizing enzymatic mixtures for the saccharification process (Kumar, Singh et al. 2008; Mekala, Singhania et al. 2008). As in bacterial and fungal cellulases, EG enzymes from *A. germari* had optimum activity at 50-55°C with a pH 6.0 (Lee, Kim et al. 2004; Lee, Lee et al. 2005; Wei, Lee et al. 2006), while a CBH from *N. koshunensis* worked optimally at pH 5.0 and 45°C (Ni, Tokuda et al. 2007). Interestingly, some insect enzymes display novel features that may be of interest for biofuel production. For example, an EG from *Aulacophora foveicollis* (L.) (Coleoptera: Chrysomelidae), displays an alkaline pH tolerance of 7.8, not seen in other animal cellulases (Watanabe and Tokuda 2001; Sami 2008).

The source of cellulolytic activity in insects has been explored to characterize how cellulolytic enzymes are produced in the insect digestive system. Feeding trials with antibiotics provide an inference into the need of symbionts for proper cellulose digestion {Tokuda, 2007 #97;Treves,

1994 #136}. However, it has been shown in the cockroach *P. americana* that when fed antibiotics it was able to maintain its cellulolytic activity, suggesting that in the event of disrupted symbiosis the cockroach can up-regulate endogenous hydrolase genes (Wharton, Wharton et al. 1965). A dual digestion sytem was found in *C. formosanus* whereas endogenous cellulases were responsible for degradation in the midgut, and symbiont-derived enzymes further degraded cellulose in the hindgut (Nakashima, Watanabe et al. 2002). Since antibiotic feeding yields indirect evidence of the importance of symbiotic cellulases, more direct approaches to determine the source of digestive enzymes are needed. Sequencing of cellulases and alignment to cellulase genes of known origin may provide, at least in some cases, a more definitive answer to their origin.

Insect cellulases are crucial to insect survival due to their role in obtaining nutritional sugars from plant biomass. RNA interference (RNAi) of the *cell-1* cellulase gene in the termite *R. flavipes* resulted in reduced feeding and fitness (Zhou, Wheeler et al. 2008). Additionally, evidence for the role of plant compounds to inhibit EG activity in *A. foveicollis* has been reported (Sami 2008). This information is relevant to biofuels, especially considering that numerous cellulolytic inhibitors could cause difficulties in biomass reactors (Juge 2006). Evolution of insect cellulases to avoid inhibition by plant chemicals may provide clues for engineering enzymes with improved activity during plant biomass degradation. The effect of cellulase knockdown on insect viability (Zhou et al., 2008) suggests the potential use of these enzymes not only as new biocatalysts, but also as targets for the design of insecticidal technologies. Interestingly, cellulolytic activity may also be relevant for resistance to insecticides. Resistant

strains of *Sitophilus zeamais* (Motschulsky) (Coleoptera: Curculionidae) were found to have increased cellulolytic and protease activity (Araujo, Guedes et al. 2008), possibly due to increased energy needs.

Current evidence suggests that the insect digestive system provides interesting environments to bioprospect for novel biofuel catalysts. Broader understandings of cellulose digestion in insects will continue to grow under the advent of genomic and proteomic technologies to help characterize complex symbiotic interactions that result in efficient assimilation of sugars from plant biomass. The results from research on enzymes from insect digestive systems have already demonstrated their use outside of biofuels as possible targets for insecticides. The lignocellulose barrier has evolved recalcitrance to resist degradation. Considering that insects have adapted their digestive system to overcome plant defensive strategies, characterization of insect cellulolytic systems may provide information crucial to efficient deconstruction of lignocellulosic materials for biofuel production.

Chapter 2

Identification and characterization of novel cellulolytic activity from salivary and digestive fluids of *Dissosteira carolina* (Orthoptera: Acrididae)

This chapter is a revised version of a paper by the same title to be submitted to the *Insect Biochemistry and Molecular Biology* by Jonathan D. Willis, William Klingeman, Brenda Oppert, Cris Oppert, and Juan Luis Jurat-Fuentes. The use of "our" in this chapter refers to my co-author and me. My primary contributions to this paper include (1) the experimental work, (2) the collection and analysis of data, (3) the gathering and interpretation of literature, and (4) most of the writing.

Abstract:

Novel cellulolytic enzymes are needed as cellulose degradation catalysts for biotechnological processing of plant biomass into biofuel. Screening of insect salivary and gut fluids using carboxymethylcellulose (CMC) and microcrystalline cellulose (MCC) as substrates demonstrated relatively high cellulolytic activity in digestive fluids from the gut and head of Carolina grasshoppers, *Dissosteira carolina* (L.) (Orthoptera: Acrididae). In zymograms using CMC as substrate, we detected four distinct cellulolytic protein bands in *D. carolina* gut fluids, and compared cellulase expression through developmental stages and between the gut regions. Using size exclusion and anion exchange liquid chromatography we purified and obtain a partial sequence for two cellulolytic protein bands from *D. carolina* head and one from gut fluids. NCBI database searches confirm high similarity with insect, bacterial and plant β–1,4-endoglucanases for the head bands, however the gut band matched only to bacterial enzymes. Results from our work demonstrate the presence of cellulolytic activity in the digestive system of *D. carolina*. Considering that this grasshopper species feeds on diverse grassy plant material, and the proposed use of switchgrass as feedstock for bioethanol production, characterization of

these novel cellulolytic systems may help develop applications in plant biomass biodegradation for biofuel production.

Introduction:

Due to its abundance and renewability, cellulose is considered the optimal biopolymer for production of ethanol biofuel as alternative to fossil fuels. Cellulose is found in plant cell walls and is degraded by the synergistic effect of three enzymatic activities: endo-β-1,4-glucanases (EG; EC. 3.2.1.4), exo-β-1,4-cellobiohydrolases (CBH; EC. 3.2.1.91), and β-glycosidases (EC. 3.2.1.21) (Clarke 1997). High costs associated with enzymatic lignocellulose degradation during ethanol biofuel production have increased interest in prospecting for more efficient cellulolytic systems. Even though plants, bacteria, and fungi have traditionally been the focus of cellulase research efforts, reports on insect cellulolytic activity have increased interest in these organisms as bioprospecting resource for novel enzymes.

Insect cellulolytic activity has been reported in ten different orders of insects, including Blattodea, Coleoptera, Diptera, Hymenoptera, Isoptera, Lepidoptera, Orthoptera, Phasmatodea, Plecoptera, and Trichoptera (Martin 1991; Marana, Terra et al. 1995; Cazemier, Huub et al. 1997.). Cellulolytic activity has been mostly characterized in termites, resulting in the first identification of an endogenous insect cellulase in *Reticulitermes speratus* (Kolbe) (Isoptera: Rhinotermitidae) (Watanabe, Noda et al. 1998). Additional research has resulted in cloning of cellulases from *Apriona germari* (Hope) (Coleoptera: Cerambycidae) (Lee, Kim et al. 2004; Lee, Lee et al. 2005; Wei, Lee et al. 2006), *Psacothea hilaris* (Pascoe) (Coleoptera; Cerambycidae) (Sugimura, Watanabe et al. 2003), *Apis mellifera* (L.) (Hymenoptera: Apidae) (Kunieda,

Fujiyuki et al. 2006), *Mastotermes darwiniensis* (Froggatt) (Isoptera: Mastotermitidae) (Li, Frohlich et al. 2003), *Reticulitermes flaviceps* (Kollar) (Isoptera: Rhinotermitidae) (Zhou, Smith et al. 2007), and *Teleogryllus emma* (Ohmachi et Matsuura) (Orthoptera: Gryllidae) (Kim, Choo et al. 2008). Availability of insect genomes has allowed detection of insect cellulases among the *Apis mellifera* (L.) (Kunieda, Fujiyuki et al. 2006) and *Tribolium castaneum* (Herbst) (Consortium 2008) genes. Availability of cloned or sequenced insect cellulases has also allowed classification of these enzymes into glycosyl hydrolase families (GHF) based on sequence homology. Cellulases belonging to GHF9 have predominately been documented in termites (Watanabe and Tokuda 2001) and *T. emma(Kim, Choo et al. 2008)*. Alternatively, enzymes belonging to GHF5 and GHF45 have been documented in coleopteran species, including *Apriona germari* (Lee, Kim et al. 2004; Lee, Lee et al. 2005; Wei, Lee et al. 2006), *Phaedon cochleariae* (Fabricius) (Girard and Jouanin 1999), and *Psacothea hilaris* (Pascoe) (Sugimura, Watanabe et al. 2003). Symbionts found in the digestive system of termite species provide cellulolytic enzymes belonging to GHF7 and GHF45 (Nakashima, Watanabe et al. 2002; Li, Frohlich et al. 2003). It has been hypothesized, at least for termites, that the mixture of numerous GHFs from endogenous and symbiotic sources result in a more efficient degradation of lignocellulosic material (Zhou, Smith et al. 2007).

Even though orthopteran species in the Acrididae family are notorious generalist plant feeders, limited information is available on specific cellulolytic systems in these species. Early literature suggested that orthopteran species lacked significant cellulolytic enzymes and speculated that the energy source was the intercellular fluids and not the degradation the plant cell wall (Davis 1963; Morgan 1975; Martin 1983; Martin 1991). During a screen for insects with cellulolytic activity

against carboxymethyl cellulose (CMC) as substrate, we observed comparatively high activity in the gut of the Carolina grasshopper (*D. carolina*) when measured against that of *R. flaviceps* and *Tenebrio molitor* (L.) (Coleoptera: Tenebrionidae). In this work, we report on the detection and initial characterization of novel cellulolytic enzymes from adult and nymphal life stages of *Dissosteira carolina.*

Materials and Methods:

Grasshopper Collection and Dissections

Dissoteria carolina adults and nymphs were field collected from the IJAMS Mead Quarry site near Knoxville, TN (Lat: 35.95601, Long.:, -83.873255) by hand capture and nets. Individuals were allowed to feed ad libitum on wheat seedlings overnight in plastic cages. Life stage was determined by measuring hind femur length (Pfadt 1994). Individuals were cooled to 4°C before dissection to slow down metabolism and ease of handling and dissections performed on ice. Heads and guts from three individuals in the same developmental stadium were dissected and pooled into separate micro-centrifuge tubes with 100 μl of water. Gut tissue was further separated into foregut, midgut, and hindgut regions. Foregut and midgut regions were separated at the origination of the gastric caeca, and the midgut and hindgut regions were separated at the origination of the Malpighian tubules. Both head and gut samples were minced with micro-scissors and disposable pellet pestles, and homogenized by vortexing to ensure fluid extraction. Samples were cleared of debris by centrifugation at 21,000 x *g* for 3 min. at room temperature. Samples were stored at -80°C until used.

Detection of cellulolytic activity in *D. carolina* fluids

Dissoteira carolina gut and head samples from instars, adults, and gut sections were analyzed using a modified dinitrosalicylic acid (DNSA) assay (Miller 1959). Protein concentrations were first estimated using the Coomassie Protein Assay Reagent (Pierce, Rockford, IL) with BSA as standard. Samples were then mixed with either 2% carboxymethyl cellulose sodium salt (CMC, Sigma-Aldrich, St. Louis, MO) or microcrystalline cellulose (MCC, Acros Organics, Geel, Belgium) suspended in 50 mM sodium citrate buffer, pH 6.0. CMC and MCC assays were conducted with 50 and 150 μg protein, respectively. Due to limited availabililty of digestive fluids, enzymatic activity within the respective gut regions was only assayed using CMC. Samples were incubated for 1 h (CMC) or 2 h (MCC) at 50°C. The DNSA reagent containing Rochelle salt (after Miller 1959) was added to samples to stop enzymatic activity and color was developed at 100°C for 15 min. Samples were cooled and centrifuged at 2000 x *g* for two min. to precipitate any remaining substrate and supernatants transferred to polystyrene microplates. Sample absorbance at 595 nm was determined on a Synergy HT microplate reader (BioTek) using KC4 software (v. 3.1). Background amounts of reducing sugars were corrected for by subtracting initial from final values of the calculated reducing sugars in the sample. One unit of cellulolytic activity is defined as the amount of enzyme that produces 1 μmol of reducing sugar (glucose equivalents) per minute at 50°C and pH 6.0. Specific activities are reported as units mg protein^{-1}. All specific activities reported represent averages from at least three independent biological replicates.

Detection of cellulases by zymography

To detect proteins with cellulolytic activity in *D. carolina* fluids, we used zymograms (Schwarz 1987), with minor modifications as described below. Carboxymethyllcellulose (CMC, 0.2%) was added as substrate to the resolving gel of SDS-10% or 12% PAGE gels before polymerization. CMC was slowly added to the gel mixture to prevent aggregation. Gel polymerization was induced after all CMC was dissolved. Gels were allowed to polymerize overnight at room temperature, then kept at 4°C until used (< two weeks).

D. carolina gut or head fluid extracts were thawed on ice, proteins in samples (100 μg) were made soluble in a half volume (25 μl) of sample buffer (Laemmli 1970) and heated at 70°C for 20 minutes to partially denature enzymes and prevent band smearing due to continuous strong enzymatic activity during electrophoresis. Commercial grade *Aspergillus niger* (van Tieghem) cellulase (Sigma-Aldrich) was used as a positive control (approx. 1 mg per lane). Following heating, samples were briefly centrifuged to collect evaporated solution, then loaded on gels. Samples were loaded in duplicate in each gel for activity and total protein staining. Electrophoretic separation was carried out at 4°C at constant voltage (100 V) for approx. 4 hours, or until the sample buffer dye reached the bottom of the gel.

After electrophoresis, activity bands were excised from the gel and used for activity or total protein staining. Total protein staining was done using Protoblue safe (National Diagnostics) following manufacturer's instructions. For cellulolytic activity staining, gels were washed five times (6 min. each) in 50 ml of wash buffer (0.1 sodium succinate pH 5.8 plus 10mM DTT). A final 30 minute wash in wash buffer without DTT was carried out at 30°C. Gels containing CMC was stained by incubating in a 0.1% Congo Red solution for 10-15 minutes at room

temperature and destained using 1M NaCl until clear activity bands became visible. For increased visualization of activity bands, acetic acid was added (100 μl/50 ml) to the NaCl solution (Waeonukul, Kyu et al. 2007). Following this treatment, the gels turned dark-purple in color with activity bands remaining as clear zones. Images of gels were acquired using a Versadoc 1000 Imager (Bio-Rad), and pictures inverted and enhanced using Adobe Photoshop CS2 software (v. 9.0.2).

Purification of cellulolytic enzymes

We used a two-step liquid chromatography strategy to purify cellulolytic enzymes from *D. carolina* head extracts. Pooled head fluids (45 ml) from adult and nymph *D. carolina* were diluted to 50 ml with 0.01M sodium acetate buffer pH 4.9 (Buffer A). Samples were then cleared by centrifugation (5 min at 21,000 x *g*) then filtered (0.22 μm) before loading into a HiLoad 16/60 Superdex 200 prep grade gel filtration column previously equilibrated with buffer A and connected to an AKTA FPLC system (GE Healthcare). Proteins in the sample were eluted with 240 ml of buffer A at 1 ml/min, collecting 2 ml fractions.

To test for the presence of cellulolytic activity in the FPLC fractions, we used 1% agarose plates containing 0.2% CMC. An aliquot (20 μl) of each fraction was spotted on the plate and allowed to digest the CMC substrate at 30°C for 2 hours. After incubation, plates were stained with 0.1% Congo red and destained with 1 M NaCl. Cleared spots revealed fractions with active cellulolytic protein.

Active fractions as determined by CMC-Agarose plates were pooled and used in anion exchange chromatography with a 5 ml Mono Q 5/50 GL anion exchange column equilibrated with buffer C

(50 mM Bis-Tris, pH 4.8). Prior to anion exchange chromatography, pooled fractions were dialyzed overnight in buffer C. After loading the pooled sample in the column, proteins were eluted using a 50% gradient over five column volumes of buffer D (buffer C plus 1 M NaCl). Fractions were tested for cellulolytic activity using the CMC-agarose plate assay, as described. Fractions containing cellulolytic activity were diluted three fold in buffer A and used for a second anion exchange run using a 50% gradient of buffer D but over two column volumes. Fractions with cellulolytic activity, as determined with the CMC-agarose plate assay, were pooled and concentrated about 100-fold using a speed vacuum (Savant, Farmingdale, NY).

Enzyme sequencing and identification

Zymogram analysis (SDS-8%PAGE, 0.2% CMC) with the purified cellulolytic activity revealed the presence of two separate cellulolytic bands (45 kDa and 43 kDa). The purified cellulase bands were separated by SDS/10%PAGE and transferred to polyvinylidene fluoride (PVDF) membrane overnight. Bands, which were visualized with Coomasie brilliant blue, were excised from the PVDF membrane and submitted for N-terminal protein sequencing (Iowa State Protein Facility, Ames, IA). Returned protein sequences were used for searches of the NCBI and Swiss-Prot database (September 23, 2009) using BLASTp for homology comparison.

Results:

Detection of cellulolytic activity during *D. carolina* development

In a screen for cellulolytic activity in insect digestive fluids (Oppert, Klingeman et al. 2009),enzymes within gut fluids gathered from *D. carolina* adults showed levels of activity

similar to that of *R. speratus* and *T. molitor* (Figure 1). Our first objective was to test for the expression of diverse cellulolytic enzymes during *D. carolina* development. Evaluation was carried out using the DNSA assay to screen gut and head extracts from the different life stages using CMC and MCC as cellulose substrates. Consistent with the screening assay, gut activity was higher than that of head fluids on CMC, but not for MCC (Fig 2A-C). CMC activity assays with gut fluids showed that the 1st nymphal instar had the highest activity levels (Figure 2A). In comparison, head fluid activity on CMC was not significantly different across developmental stadia (Figure 2A left, blue

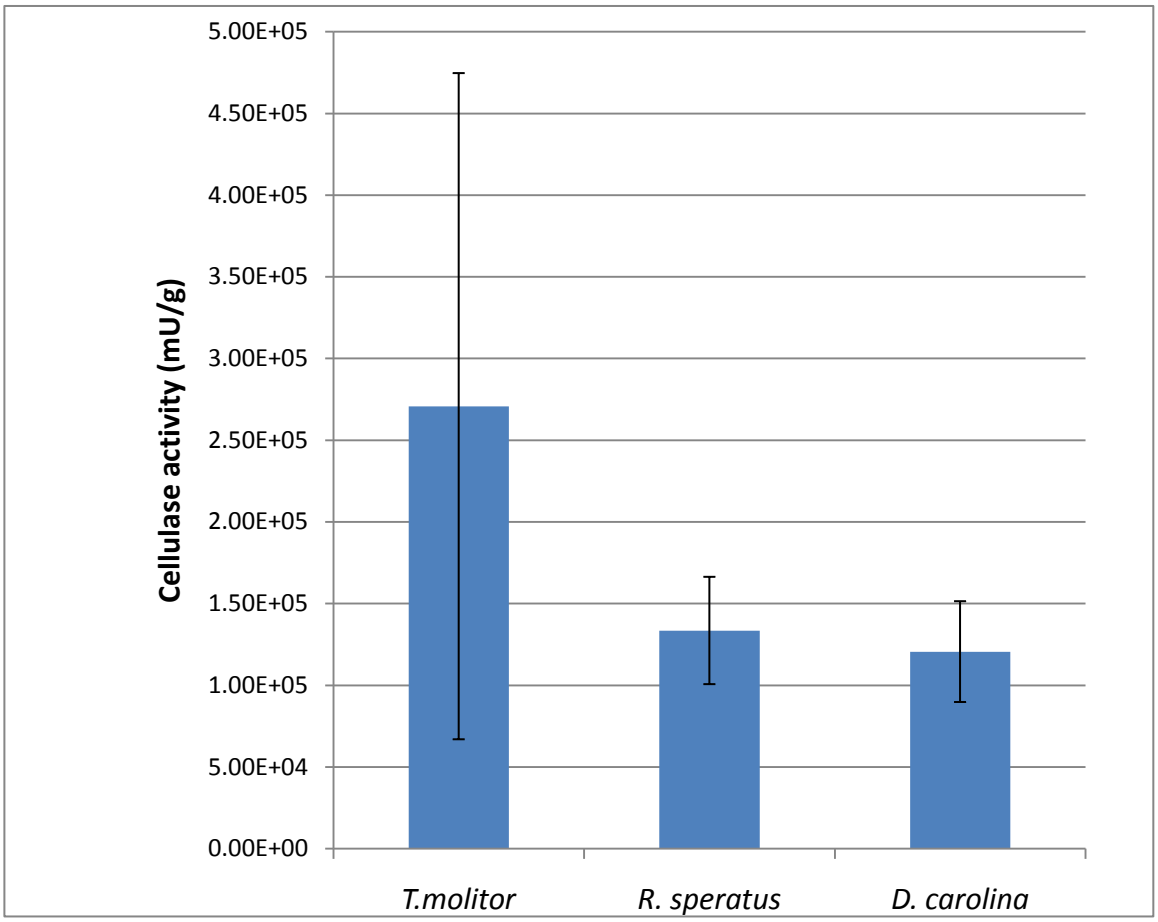
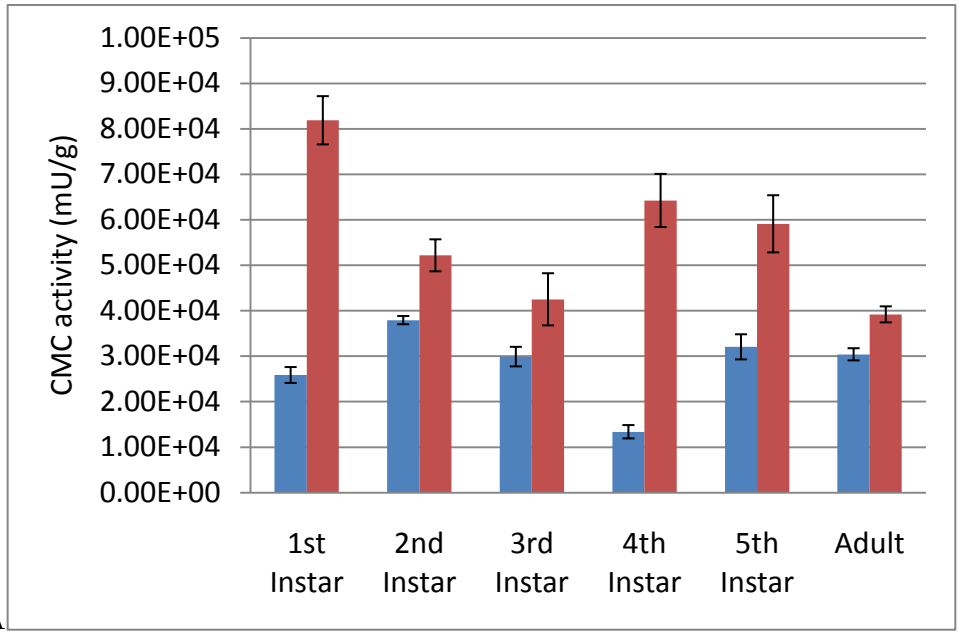
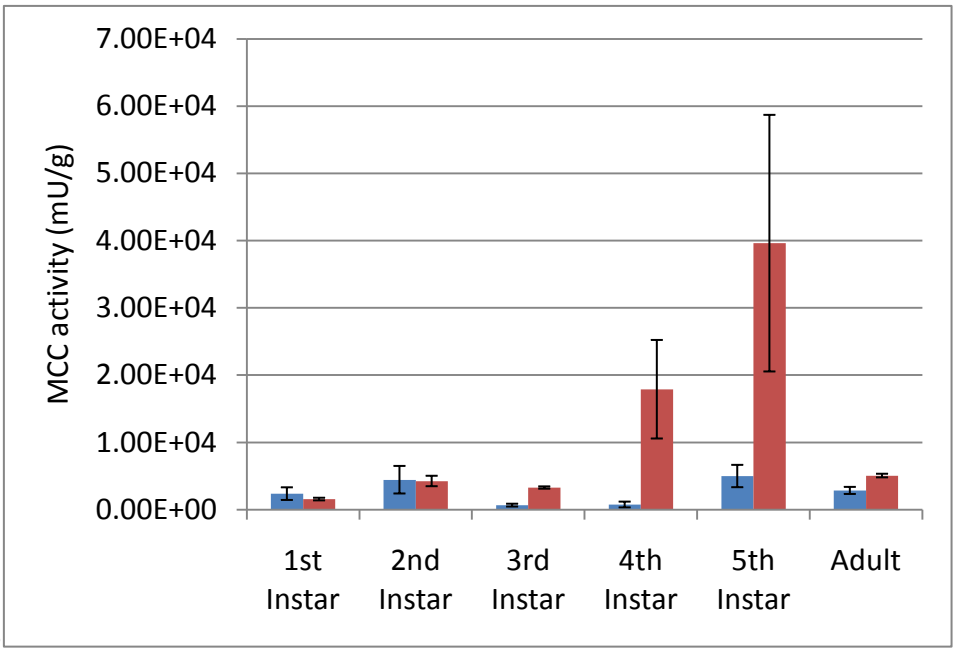


Figure 1: Comparison of cellulolytic activity in gut fluids from *Tenebrio molitor*, *Reticulitermes speratus*, and *Dissosteira carolina* as determined by the DNSA assay. Data for each insect represents the mean and standard error from three independent determinations of CMC degradation by 50 µg of total gut protein for 60 minutes. One-way ANOVA showed no significant difference between the three insects at the 0.05 confidence level. *P*-value 0.40. mU = µmol glucose released per min.

Figure 2: Detection of cellulolytic activity in *Dissosteira carolina* through development. DNSA assay of head (blue) and gut (red) fluids (50 µg of protein per sample) within each *D. carolina* developmental stadium using CMC (A) or MCC (B) as substrate. Shown are the means and standard errors calculated from data from three technical replicates and three biological replicates. Different letters indicate statistically significant differences Duncan's MRT. (C) Zymogram results (12% PAGE + 0.2% CMC) of enzyme activity in gut fluids (100 µg total protein) from different life stages of *D. carolina* development. Arrows indicate the four cellulase activity bands detected.


1.00E+05
9.00E+04
8.00E+04
7.00E+04
6.00E+04
5.00E+04
4.00E+04
3.00E+04
2.00E+04
1.00E+04
0.00E+00
CMC activity (mU/g)
1st Instar
2nd Instar
3rd Instar
4th Instar
5th Instar
Adult

A


7.00E+04
6.00E+04
5.00E+04
4.00E+04
3.00E+04
2.00E+04
1.00E+04
0.00E+00
MCC activity (mU/g)
1st Instar
2nd Instar
3rd Instar
4th Instar
5th Instar
Adult

B

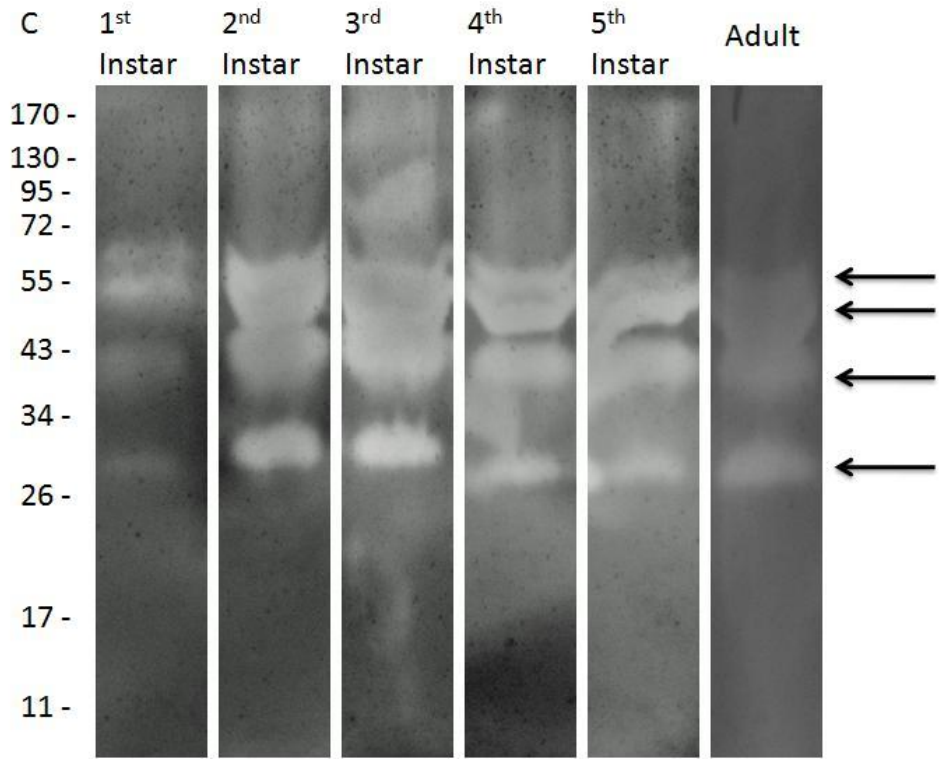


Figure 2 continue

bars). Activity against MCC in gut and head fluids showed the highest activity levels in the fourth and fifth nymphal instars, with activity being much more reduced in other life stages (Fig. 2B). Zymogram analysis (Fig 2C) demonstrated that gut fluids from each instar had the same active cellulase profile of four protein bands of approximately 68, 56, 45, and 27 kDa in size.

Activity against CMC in *D. carolina* is localized to foregut and midgut

Our second objective was to localize cellulolytic activity in the different gut regions (foregut, midgut and hindgut) of *D. carolina*. Based results from our DNSA assay screening through development, and considering the higher amounts of sample available from adults, we used guts dissected from *D. carolina* adults to determine activity against CMC in DNSA assays and zymograms. Caryboxymethylcellulose was chosen instead of MCC due to its more popular use in the literature and the possibility of performing zymograms with this substrate. Three main functional regions were studied: foregut, midgut, and hindgut (Figure 3A). Cellulase activity levels from the DNSA assay revealed that the majority of activity against CMC was localized to the foregut and midgut regions, and significantly reduced in the hindgut (Figure 3B). In agreement with these data, zymograms revealed cellulolytic bands in fluids from foregut and midgut, with no bands detected for fluids from the hindgut (Figure 3C). No differences in number and size of the cellulolytic protein bands were observed between foregut and midgut, suggesting that the same enzymes sustain activity in both regions.

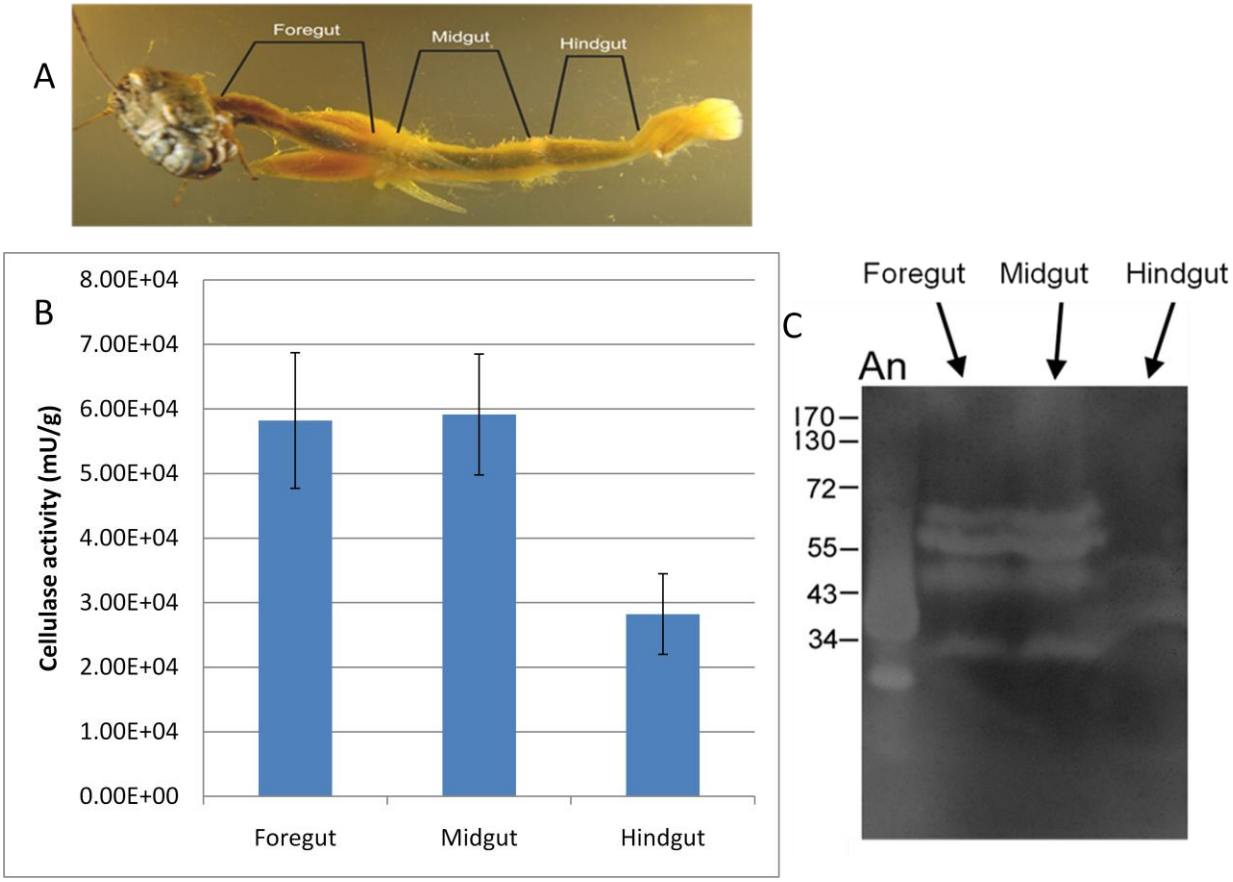


Figure 3: Localization of activity against CMC in gut regions of adult *Dissosteira carolina*. (A) Dissection of *D. carolina* adult showing specific regions the digestive tract that were tested. (B) DNSA assay results for foregut, midgut, and hindgut regions (50 µg of total gut protein evaluated for 60 min incubation with CMC substrate). Shown are means and standard errors from three technical replicates and three biological replicates. Different letters on top of columns represent statistically significant differences (Duncan's MRT $p<0.05$). (C) Zymogram (12% PAGE + 0.2% CMC) depicting active cellulase protein profile in gut regions of adult *D. carolina* (100 µg of total gut protein per lane). A positive cellulolytic control (An) showing activity in *Aspergillus niger* commercial cellulase (8 µg) is also shown.

Purification of a *D. carolina* cellulase:

To isolate the cellulolytic enzymes from the complex protein mixture found in insect fluids, we used fast performance liquid chromatography (FPLC) to separate proteins by size and ionic strength. Based on the relative abundance of cellulase activity, we focused our initial efforts on gut fluids from *D. carolina* adults. Figure 4A shows a chromatograph from the size exclusion purification with the fractions highlighted in

green containing cellulase activity as detected using CMC plates (Figure 4B). Using anion exchange chromatography we were able to purify one cellulolytic band of 45 kDa from gut fluids, as revealed by silver staining and activity assays (Figure 4C). N-terminal sequencing of this 45 kDa gut protein provided a 15 amino acid tag (DPPIARTVGSQL). Database searching using this sequence tag returned a low probability match to a fragment (^{218}VRPIARTVASSP ^{224}) from a beta-glucanase from *Sorangium cellulosum*, a known myxobacterium with cellulolytic activity (Pradella, Hans et al. 2002). Although efforts were made to design low degeneracy primers from the gut protein sequence, the sequence yielded primers too degenerate for successful PCR (data not shown). Focus was then switched to purification of cellulases from the head fluids. After FPLC purification, a single cellulolytic band from head fluids at 45 kDa was detected in zymograms (Fig 4C). Due to limited amounts of protein, silver staining was not performed with purified head bands and instead the protein was bound to a PVDF filter and upon staining with Coomassie brilliant blue two bands of 43 and 45 kDa were detected and excised. Detection of these two bands may be explained by reduced activity of the 43 kDa cellulase on zymograms. Both bands were extracted from the PVDF blot

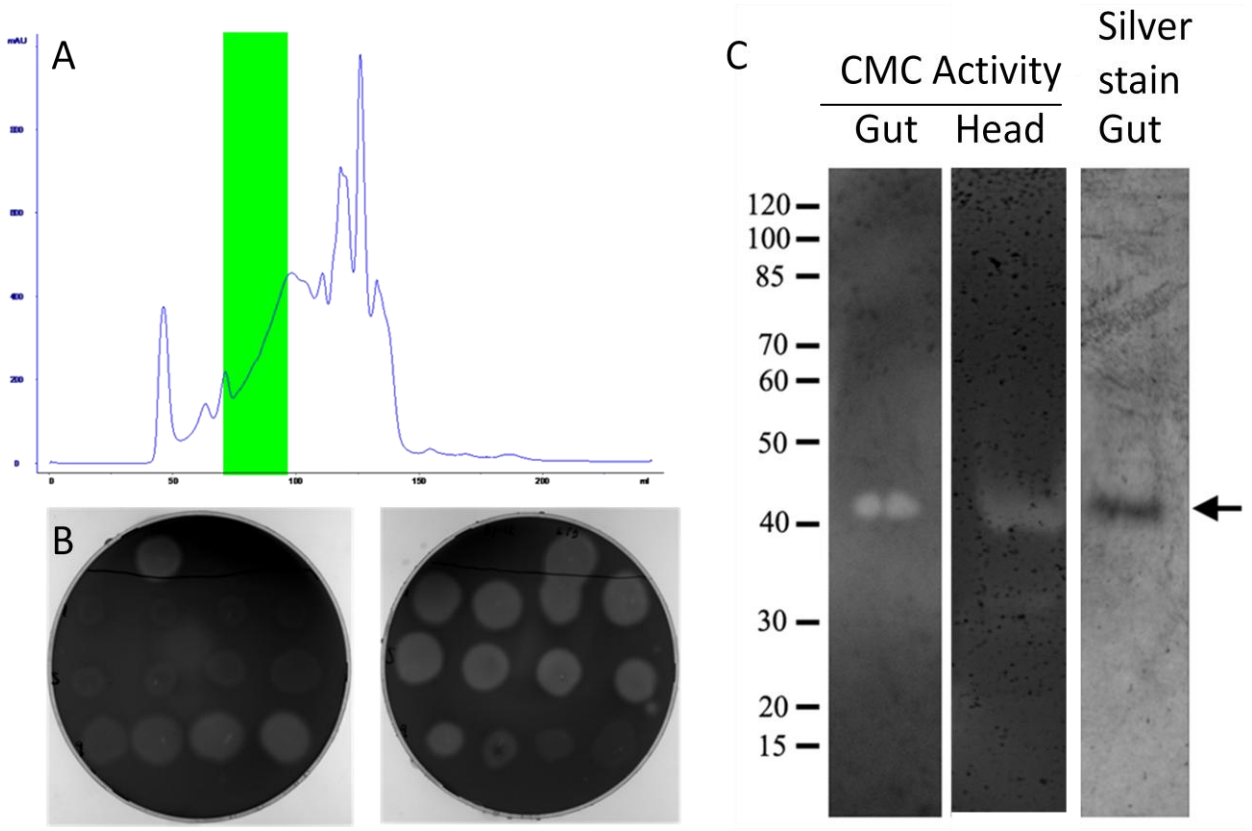


Figure 4: Purification of cellulolytic enzymes from head and gut fluids of adult *Dissosteira carolina.* (A) Chromatograph from size exclusion FPLC purification of cellulolytic activity from *D. carolina* gut fluids. Area highlighted in green corresponds to fractions with cellulase activity as determined by plate CMC zymograms. (B) Plate CMC zymograms (1% Agarose + 0.2% CMC) used to detect cellulase activity in fractions from FPLC run of gut fluid sample. (C) CMC zymogram (8% SDS-PAGE + 0.2% CMC) and silver staining of purified cellulases.

and sent for N-terminal sequencing. From the 43 kDa purified head cellulase we obtained a sequence tag (YEYRDALCKSLLF) that matched with high similarity (75%) to a region in a putative EG from the castor oil plant, *Ricinus communis* (L.) (Malpighiales: Euphorbiaceae) (^{25}EYREALSKSILF^{36}) and an EG from the termite *Mastotermes darwiniensis* (Froggatt) (Isoptera: Mastotermitidae) (^{18}YDYKDVLTKSLLF^{30}) (Li, Frohlich et al. 2003). The N-terminal sequence tag obtained for the 45 kDa protein band from head samples (AKYDYADAIRKSILFYQAQRS) matched with 84% homology to a region of a cellulase from the cricket *T. emma* (^{22}YDYADVIKKSLLFYQAQRS^{40}) (Kim, Choo et al. 2008).

Discussion:

Orthopteran species have traditionally not been studied for cellulolytic activity due to past feeding studies reporting low cellulose degradation in these insects. For example, a fecal evaluation study showed the majority of the cellulose ingested was also excreted by *Chortoicetes terminifera* (Walker) (Orthoptera: Acrididae), leading to a hypothesis that orthopteran species are incapable of digesting cellulose (Clissold, Sanson et al. 2004). CMC has traditionally been utilized as substrate because of its solubility to demonstrate EG activity, while MCC degradation has been attributed to the action of CBH (Clarke 1997). Endoglucanases (EGs) react on the internal portion of cellulose to break glycosidic bonds, and CMC mimics these internal portions because of the extra carboxymethyl groups present that reduce the number of cellulose chain ends (Klesov 1991). In this study, we report on the first putative EG enzyme activity detected in an orthopteran acridid species, *Dissosteira carolina* L.. Previous reports present the case for *Locusta migratoria* (L.) (Orthoptera: Acrididae) having β-glucosidase activity on cellobiose, but

did not investigate the existence of EG activity (Morgan 1975). Recently, an EG belonging to GHF9 was cloned from *T. emma* from a cDNA library generated from whole insects providing the first EG clone from an orthopteran species (Kim, Choo et al. 2008). Evidence of cellulolytic enzymes produced endogenously within the insect suggests that orthopterans can digest cellulose.

In our plate assays, activity against CMC was always higher in fluids from the gut compared to head fluids (including salivary glands), suggesting that EGs are present and active in the gut fluids, rather than being secreted by the salivary glands. Prior reports have suggested an increase in cellulose and cellobiose degradative capabilities throughout larval development with a drop at the pre-pupal stage in *Philosamia ricini* (L.) (Lepidoptera: Saturnidae) (Pant and Ramana 1989). In *D. carolina*, we found that cellulolytic activity does not follow this pattern: relatively similar levels of activity were detected in all developmental stages. The observed discrepancy could be attributed in part to differences in the development of othopteran and lepidopteran insects. Enzymatic activity for MCC degradation was low, although constant, during development, which suggests that initial degradation of cellulosic material is mostly due to mechanical maceration by mandibular action. Alternatively, previous reports have suggested that complete degradation of cellulose in insects may be accomplished via combined EG and β-glucosidase activities (Scrivener, Slaytor et al. 1989; Scrivener 1993).

Zymogram and CMC activity data demonstrate that in *D. carolina* cellulolytic activity is localized to the foregut and midgut regions: consistent with reports of *A. germari* cellulase

activity (AgEgase III)(Wei, Lee et al. 2006) and cellulase EG1 and EG2 activities from *P. cribrata (Pant and Ramana 1989; Scrivener, Slaytor et al. 1989; Watanabe and Tokuda 2001).* Considering that foregut and midgut regions are known to harbor reduced levels of gut flora, our activity localization data suggest that the detected enzymatic activity originates endogenously within the insect. This hypothesis is supported by assays of larval *P. ricini* (Hutt)(Lepidoptera: Saturniidae), that demonstrated highest EG activity levels in the midgut even after long term exposure to antibiotics. These results were interpreted to support an insect-derived origin for these midgut enzymes (Pant and Ramana 1989). A similar antibiotic feeding study on the firebrat species, *Thermobia domestica* (Packard) (Thysanura: Lepismatidae), showed no significant differences in cellulase activity or cellulose assimilation efficiency when insects were fed with antibiotics to reduce intestinal bacteria colonies (Treves 1994). In the cockroach *P. americana,* use of antibiotics to reduce symbionts in the gut resulted in stimulation of cellulase production and cellulolytic activity suggesting that in the event of disrupted symbiosis the cockroach can up-regulate endogenous hydrolase genes (Wharton, Wharton et al. 1965).

Through sequential liquid chromatography we were able to purify two cellulolytic proteins of 43 and 45-kDa size from head fluids with sequence similarity to insect and plant β−1,4 endoglucanases. Endoglucanases of similar size were found in *T. emma, Nasutitermes takasagoensis* (Shiraki) (Isoptera: Termitidae), and *Nasutitermes walkeri* (Hill) (Isoptera: Termitidae) (Tokuda and Watanabe 2007; Kim, Choo et al. 2008). Unfortunately, due to degeneracy of our putative protein sequence, we have been unsuccessful to date in cloning the cDNA encoding these enzymes using PCR primers (data not shown). Current efforts are focused

on using alternative strategies to obtain a full length clone to be able to express these cellulases in heterologous systems and characterize their enzymatic activity against diverse substrates.

Orthopteran species are generalist herbivores making them an attractive source for novel biocatalyst for biofuels, yet there is a lack of genetic or biochemical information on cellulolytic enzymes in these insects. Our data suggest that EG activity is present in fluids from both head and gut tissues and that this activity is conserved throughout development. Localization of the cellulolytic activity to the anterior portion of the gut suggests that the detected enzymes are endogenously produced by the insect. In support of this hypothesis, screening of the digestive system of *Coptotermes formosanus* (Shiraki) (Isoptera: Rhinotermitidae) (revealed a dual digestion system whereas cellulases homologous to endogenous GHF9 termite hydrolases were found only in the salivary, foregut, and midgut regions (Nakashima, Watanabe et al. 2002). This hypothesis is also supported by the purified *D. carolina* enzymes displaying high sequence homology to other endogenous insect cellulases. Characterization of the cellulolytic system in Orthoptera would allow better understanding of the lignocellulose digestion in insects, and provide information helpful for biofuel production. Further research will be necessary to complete the cloning of other cellulolytic enzymes from *D. carolina* and to properly define if endogenous enzymes are responsible for cellulose digestion in Acrididae.

Chapter 3

Cloning, expression, and characterization of a putative cellulase from *Tribolium castaneum* (Coleoptera: Tenebrionidae)

This chapter is a lightly revised version of a paper by the same title to be submitted to the *Insect Biochemistry and Molecular Biology* journal by Jonathan D. Willis, Brenda Oppert, Cris Oppert, William Klingeman, and Juan Luis Jurat-Fuentes. The use of "our" in this chapter refers to my co-author and me. My primary contributions to this paper include (1) the experimental work, (2) the collection and analysis of data, (3) the gathering and interpretation of literature, and (4) most of the writing.

Abstract:

Availability of insect genomes has allowed for discovery and functional characterization of novel genes. In this paper we report on the identification, cloning, expression, and characterization of a putative endo-β-1,4-glucanase from the red flour beetle (*Tribolium castaneum*) we named TcEG1 (*T. castaneum* EndoGlucanase 1). Sequence analysis of the cloned full-length TcEG1 cDNA (1,356 bp) revealed sequence homology to enzymes in glycosyl hydrolase family 9 (GHF9), and verified the presence of a mutation (Gly for Ser).in the conserved catalytic domain for GHF9 cellulases. This cDNA was predicted to encode a 49.5 kDa protein with a calculated pI of 5.39. Heterologous expression of TcEG1 in *Drosophila* DS2 cell cultures resulted in production of a 51-kDa protein that was secreted in the media, as determined by Western blotting. The expressed protein was used to characterize TcEG1 enzymatic activity against diverse cellulose substrates, thermostability, pH optima, and the effect of including a His tag in the cellulase on this cellulolytic activity.

Introduction:

Interest in insect cellulolytic systems has increased in recent years with the need for novel cellulolytic enzymes to serve as biocatalysts for efficient conversion of cellulose to bioethanol. Cellulose is a complex carbohydrate biopolymer structured as numerous linear fibrils intertwined to form crystalline structures. Numerous organisms degrade cellulose to obtain glucose, a process mimicked during preparation of biofuel ethanol from lignocellulosic material (Watanabe and Tokuda 2001; Kumar, Singh et al. 2008). Complete degradation of cellulose to glucose requires synergistic action of endo-β-1,4-glucanases (EG; EC. 3.2.1.4), exo-β-1,4-cellobiohydrolases (CBH; EC. 3.2.1.91), and β-glycosidases (EC. 3.2.1.21) (Clarke 1997). Insect cellulose digestion was traditionally hypothesized to occur through enzymes produced by symbiotic gut microbes (Martin 1983; Martin 1991). The first endogenous insect cellulase genes were found in termites (Watanabe, Noda et al. 1998; Lo, Tokuda et al. 2000), followed by discoveries in beetles (Sugimura, Watanabe et al. 2003; Lee, Kim et al. 2004; Lee, Lee et al. 2005; Wei, Lee et al. 2006), and crickets (Kim, Choo et al. 2008). While discovery of these enzymes is usually challenged by complex purification protocols, several insect cellulases have been identified and cloned using genomic data (EST datasets or cDNA libraries). Data gleaned from sequenced genomes have allowed faster identification and characterization of novel cellulolytic enzyme genes in *Apis mellifera* (L.) (Hymenoptera: Apidae) (Kunieda, Fujiyuki et al. 2006), *Acyrthosiphon pisum* (Johnson) (Hemiptera: Aphididae) (Sabater-Munoz, Legeai et al. 2006), or *Tribolium castaneum*(Herbst) (Coleoptera: Tenebrionidae) (Morris, Lorenzen et al. 2009). In this paper we identify and describe the cloning, expression, and initial characterization of a novel endo-glucanase from *T. castaneum*.

Sequence homology searches using the *T. castaneum* genome (Richards, Gibbs et al. 2008) revealed a predicted cellulase gene with high similarity to a cellulase detected in the *A. mellifera* genome (Kunieda, Fujiyuki et al. 2006). This cellulase was named TcEG1 (*Tribolium castaneum* EndoGlucanase 1) for being the first endo-β-1,4-glucanase reported for *T. castaneum*. Sequence analysis of TcEG1 demonstrated high homology to cellulases belonging to Glycosyl Hydrolase Family 9 (GHF 9). Specific primers were designed based on the *TcEG1* genomic sequence and used to clone a full-length TcEG1 cDNA into *Escherichia coli* (Migula) and insect cell expression systems. Recombinant proteins demonstrated activity on carboxymethyl cellulose (CMC), but not on microcrystalline cellulose (MCC) providing evidence for typical endo-glucanase activity.

Materials and Methods:

Insects

Adults and larvae of *T. castaneum* were obtained from the USDA Grain Marketing and Production Research Center in Manhattan, Kansas. Insects were maintained in an incubator at 26°C with 70% relative humidity and 18L:6D photoperiod and fed plain white flour and oats. Under these conditions it took larvae 2-3 weeks to reach 6^{th} instar: the stadia used for dissections to purify RNA.

Larval head and gut RNA and protein extraction

Sixth instar larvae were collected and anesthetized on ice for 10 min before dissection. Head capsules were separated and placed into micro-centrifuge tubes (10 per tube) with 100 µl of

RNAlater (Ambion, Austin, TX). Dissected guts were placed into a micro-centrifuge tube (10 per tube) containing 100 µl of RNAlater. Tissues were incubated at 4 C° overnight to allow RNAlater to infiltrate and preserve the RNA. Tubes were spun down and RNAlater removed followed by flash freezing and mortar pestle grinding. RNA was extracted using the RNeasy kit (Qiagen, Valencia, CA) following manufacturer's instructions, checked on a 1% agarose gel for purity and stored at -80 C° until used. RNA concentrations were estimated using the Quant-iT RNA assay kit (Invitrogen, Carlsbad, CA) following manufacturer's instructions.

For protein extraction, contents of 15 head capsules or larval guts were pooled and placed in micro-centrifuge tubes containing 100 µl of ultrapure water. Tissues were ground and homogenized with a tissue homogenizer and centrifuged at 5,000 x *g* for 5 min. Supernatants were removed as soluble protein samples and stored at -80 C° until used.

cDNA creation and expression vector construction

RNA from head and gut fluids were used for cDNA synthesis using Superscript III reverse transcriptase (Invitrogen). Primers for TcEG-1 were designed for amplification of the full-length cDNA sequence and to assist with proper insertion into the pET 30a vector using EcoR I and Not I restriction enzyme sites (underlined): TcFwd 5'-GGAATTCATGTTCTACTCATTGTGGGTGCTACTATTT-3' and TcRev 5'-ATAGTTTAGCGGCCGCTCACAATTTATTCTCATTTTCAATATAAAT-3' based on sequence from the *Tribolium* genome database.. Accuprime pfx supermix (Invitrogen) was used for PCR reactions containing 5% DMSO \ 20 mmol of each primer, and 1 µl of cDNA. The protocol for the PCR reaction consisted of an initial 5 minutes at 95°C, followed by 30 rounds of

30 sec at 95°C, 30 sec 60°C, 90 sec at 68°C, followed by a 10 min extension at 68°C. The PCR products were cloned into TOPO TA (Invitrogen). Constructs were transformed into *E. coli* TOP10 competent cells (Invitrogen). The insert sequence was verified for coding frame by sequencing transformants in forward and reverse directions (University of Tennessee Sequencing Facility). The TcEG1 insert in the TOPO plasmid was excised using EcoR I and Not I digestion and gel-purified (Qiagen Gel Extraction Kit). Purified full-length TcEG1 cDNA was ligated in the pET 30a expression vector (previously digested with EcoR I and Not I) using T4 DNA ligase (Invitrogen) to generate the pET30a/TcEG1 expression cassette. Insert sequence was verified for coding frame by sequencing transformants in forward and reverse directions (University of Tennessee Sequencing Facility).

Due to a frame change between pET30a and the pIZT-V5/His vector (Invitrogen), primers designed for correct reading frame and containing EcoRI and NotI restriction sites (underlined) were designed: TcpIZTFwd 5'-GGAATTCGATGTTCTACTCATTGTGGGTGCTACTATTT-3' and TcpIZTRev 5'- ATAGTTTAGCGGCCGCCAATTTATTCTCATTTTCAATATAAAT-3'. PCR reactions were as described above. Amplicons and pIZT-V5/His vector were digested with EcoRI and NotI and ligated using T4 DNA ligase (Invitrogen) to generate the pIZT-V5/His/TcEG1 expression vector. Constructs were transformed into *E.coli* TOP10 cells (Invitrogen). Clones were sequenced in forward and reverse directions (University of Tennessee Sequencing Facility) to ensure the proper reading frames were intact for protein expression.

Sequence alignment and phylogenetic analysis

The TcEG1 sequence was analyzed with comput pI/MW tool on the ExPASy database to predict pI of the protein. O- and N- glycosylation sites were predicted by DictyOGlyc 1.1 (http://www.cbs.dtu.dk/services/DictyOGlyc/) and NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) programs. Signal peptide prediction was performed by SignalP 3.0 (http://www.cbs.dtu.dk).

For the phylogenetic comparison, protein sequences from insect cellulases belonging to the GHF9 family were obtained from the NCBI database and aligned with the predicted TcEG1 protein sequence using ClustalX 2.0.9. After the alignment was performed, an unrooted Phylip tree was generated with the draw tree function in Clustal X 2.0.9, and then exported into NJPLOT to display the tree.

TcEG1 expression and purification in *E. coli* cultures

Purified pET30a/TcEG1 plasmid DNA from transformed TOP10 cells was transformed into *E. coli* strain BL21(DE3). Transformants were grown overnight with constant agitation at 37°C in 5 ml of LB broth containing 30 μg/ml kanamycin. After 12 hr, 5ml of the overnight culture were used to inoculate 500 ml of LB broth plus 30 μg/ml kanamycin in a 1 L flask. Cultures were allowed to grow at 37°C with constant agitation until reaching OD 0.70 (typically about 4-5 hours), and then expression was induced with 1 mM IPTG. Bacterial cultures were pelleted by centrifugation (6,000 x *g* for 6 min) 20 hrs post-induction. Expressed TcEG1 was accumulated as soluble protein inside the bacterial cells. Culture pellets were resuspended in 160 ml of native purification buffer (25mM NaH_2PO_4 pH 8.0, 250 mM NaCl). Cells were lysed by incubation

with lysozyme (0.05 %) for 30 min on ice. Lysates were then drawn through a 20 gauge needle three times to reduce viscosity and cleared by centrifugation (5,000 x *g* for 15 min.). Supernatants containing TcEG1 were stored at -20°C until further purification. TcEG1 was purified from these supernatants affinity chromatography using a His Trap FF 5 ml column pre-equilibrated with buffer A (20 mM phosphate buffer, pH 7.4, 30 mM imidazole 0.5 M NaCl) and connected to an AKTA FPLC system (GE Healthcare). TcEG1 was eluted from the column using a one step gradient of buffer A containing 500 mM imidazole.

Transient expression of TcEG1 in *Drosophila* S2 cell cultures

Cultured *Drosophila* S2 cells were maintained in serum-free insect cell medium (HyQ SFX-Insect, Hyclone) and transfected as previously described (Banks, Hua et al. 2003). The pIZT-V5/His/TcEG1 plasmid DNA used for transfections was prepared using the Qiagen Midiprep kit. Approximately 2×10^6 DS2 cells from a confluent culture were suspended in 2 ml fresh media and allowed to adhere overnight to surface-treated 60 × 15 mm polystyrene dishes (Falcon). Plasmid transfection mixtures were prepared by mixing either 2.5 μg of pIZT, or 5 μg of pIZT-V5/His/TcEG1 plasmids with 1 ml of serum-free insect medium and 20 μl of Cellfectin reagent (Invitrogen). Transfection mixtures were incubated with cells at room temperature for four hours; subsequently fresh media (3 ml final volume) was added to the plates. Cell cultures were incubated at 26°C for three days before collection and centrifugation at room temperature (6,000 x *g* for 3min). Due to the the presence of a predicted secretion signal peptide in the TcEG1 sequence, both cell pellets and supernatant media were kept to test for TcEG1 expression. Supernatant media was concentrated approximately 10 fold using a speedvac (Savant,

Farmingdale, NY), DS2 cell pellets were lysed by resuspension in native purification buffer (25 mM NaH_2PO_4 pH 8.0, 250 mM NaCl) and two cycles of freeze-thawing. After lysis, samples were cleared by centrifugation as above and supernatants collected and stored at -80°C until used.

Detection of TcEG1 expression

Purified TcEG1 protein from *E. coli* cultures (1 μg in 20 μl of Buffer B), DS2 cell pellets (1 μg in 20 μl of native purification buffer), or media supernatant (1 μg in 20 μl), were mixed with 10 μl of 2X sample buffer (Laemmli 1970) and heat denatured at 95°C for for 5 min. Protein concentrations were estimated using the Coomassie Protein Assay Reagent (Pierce, Rockford, IL) following manufacturer's instructions and using BSA as standard. Proteins were separated by electrophoresis on SDS-10%PAGE gels. After electrophoresis, proteins were electrotransferred to polyvinylidene difluoride (PVDF) filters overnight at constant voltage (20 V) at 4°C. Blots were blocked with 3% bovine serum albumin (BSA) in PBST (135 mM NaCl, 2 mM KCl, 10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, pH 7.5, 0.1% Tween-20) for 1 hour at room temperature. Blots were then probed with a 1:10,000 dilution of antisera against 6x His tag conjugated to horse radish peroxidase (HRP) for 1 hour in blocking buffer. After several washes with PBST plus 0.1% BSA for one hour, blots were developed with the Supersignal West Pico enhanced chemiluminescence substrate (Pierce).

Quantification of TcEG1 enzymatic activity

Gut and head digestive fluid samples from *T. castaneum* larvae and recombinant proteins or media from DS2 cell cultures were analyzed for endoglucanase (EG) activity using a modified dinitrosalicylic acid (DNSA) assay (Miller 1959). Samples were mixed with either 2% carboxymethyl cellulose sodium salt (CMC, Sigma-Aldrich, St. Louis, MO) or microcrystalline cellulose (MCC, Acros Organics, Geel, Belgium) suspended in 50 mM sodium citrate buffer, pH 6.0. In-solution CMC and MCC assays were conducted with 50 and 150 μg of protein, respectively. Protein concentrations were estimated using the Coomassie Protein Assay Reagent (Pierce, Rockford, IL) following manufacturer's instructions and using BSA as standard. Samples were incubated for 1 h (CMC) or 2 h (MCC) at 50°C. A modified DNSA reagent containing Rochelle salt was added to samples to stop enzymatic activity and color was developed at 100°C for 15 min. Samples were centrifuged at 2,000 x *g* for two minutes to precipitate any remaining substrate. Supernatants were transferred to polystyrene microplates and the absorbance at 595 nm read on a Synergy HT microplate reader (BioTek) using the KC4 software (v. 3.1). Background amounts of reducing sugars were corrected for by subtracting final from initial values of the calculated reducing sugars in the sample. One unit of cellulolytic activity is defined as the amount of enzyme that produces 1 μmol of reducing sugar (glucose equivalents) min^{-1} at 50°C and pH 6.0. Specific activities are reported as units per mg of protein. Protein concentrations for thermal stability and pH optimization were estimated using the Quant-iT Protein assay kit (Invitrogen) following manufacturer's instructions. Thermal stability of recombinant crude enzyme from DS2 cell cultures was tested by incubating 25 μg of protein at 30°C, 40°C, 50°C, 60°C, or 70°C for 10 minutes. Treated samples were used in assays with

CMC as substrate as described above. The effect of pH on TcEG1 activity was detected by incubating recombinant protein from DS2 cell cultures (10 μg) in 2% CMC dissolved in buffer at pH 2.0 (maleic acid), pH 5.0 (citrate buffer), pH 7.0 (phosphate buffer), or pH 9.0 (glycine buffer), and CMC degradation detected and measured as above. All specific activities and absorbance readings reported represent averages from at least three experiments with material from independent recombinant enzyme or larval batches (biological replicates) performed in triplicate (technical replicates).

Results and Discussion:

Tribolium castaneum is one of the most damaging storage grain pests affecting stored cereal grain crops. Extensive genomic and proteomic resources, including its complete genome sequence, have been generated for this insect, considered one of the main insect models for developmental studies (Consortium 2008). Using these resources we previously described expression of a putative cellulase, we named TcEG1, in the head and gut tissue of *T. castaneum* larvae (Morris, Lorenzen et al. 2009). The full length cDNA from the *TcEG1* gene encodes a predicted 50 kDa protein with a pI of 5.39. Three N- (^{148}NMTR^{151}, ^{184}NQSY^{187}, and ^{335}NISQ^{338}) and one O- (^{304}Thr) predicted glycosylation sites were detected for TcEG1 (Fig. 5).

Protein alignment and phylogenetic analyses revealed that TcEG1 displays high similarity (>80%) with cellulases in the GHF9 family (Figs. 6 and 7). The GHF9 hydrolase family consists of EG, CBH, and EC enzymes from prokaryotic and eukaryotic sources, with a typical 3D structure containing 6 alpha helices (Khademi, Guarino et al. 2002). The catalytic domain is joined to the cellulose binding domain through a linker sequence. The predicted TcEG1 catalytic

domain contains the conserved ^{69}Asp and ^{71}Asp residues ,characteristic of GFH9 enzymes, as well as a glutamine residue (^{427}Gln) acting as proton acceptor at the C-terminus (Fig. 5). A conserved serine residue at position 70 among GHF9 members is substituted by a Gly residue in TcEG1 (Fig. 6). Further research would be necessary to examine the implications of this substitution for catalytic activity.

The GHF9 family contains the majority of identified and cloned insect EGs, including enzymes from termite (Watanabe and Tokuda 2001), cricket (Kim, Choo et al. 2008), and honeybee (Kunieda, Fujiyuki et al. 2006) species. An unrooted phylogenetic tree (Fig. 7)

```
TcEG-1 (1)   MFYSLWVLLFITFGSAYDYRHALKLSLLFYEAQRSGPLPASNRITWRSDS  (50)
TcEG-1 (51)   ALDDHGLNNEDLTGGYYDASDYVKFSFTMAFTTTILTWGMLSFRHAYNEA (100)
TcEG-1 (101) GQYQHGLDAIKWATDYFIKCHASKYEFYGQIGDFSLDHAFWGRPEEMNMT (150)
TcEG-1 (151) RPAYKIDIEHPGSDLAGEAVAALAAASLLFENFNQSYSDELLRHATELYD (200)
TcEG-1 (201) FATMYRGLYHDSIPGAKVYYESTGYGDELTWAAIWLYKKTKDVKYIEQAE (250)
TcEG-1 (251) TFYTKFRIKDRPNEFFYNKKVAGIQLLLAELTLRPEYISTIKNFCDYTIK (300)
TcEG-1 (301) EQIRTPKGLIFIDKSGTLSHAANVAFICLHAGLTLNISQAAYVSFAKEQI (350)
TcEG-1 (351) NYMLGSTGQSFVVGYGQNYPKQPHHSASSCPNLPEPCGWKQFTWKGPNPQ (400)
TcEG-1 (401) ILYGALVSGPDQNDHYEDVREEFLYNEVTLDYNAGFQSTLAGLIYIENENKL(452)
```

Figure 5: Tc-EG1 sequence with gray highlight showing predicted signal peptide for secretion. Predicted N-glycosylation sites are underlined and O-glycosylation sites highlighted in black. Square region denotes proton donor catalytic region. Proton acceptor is Glu highlighted in yellow.

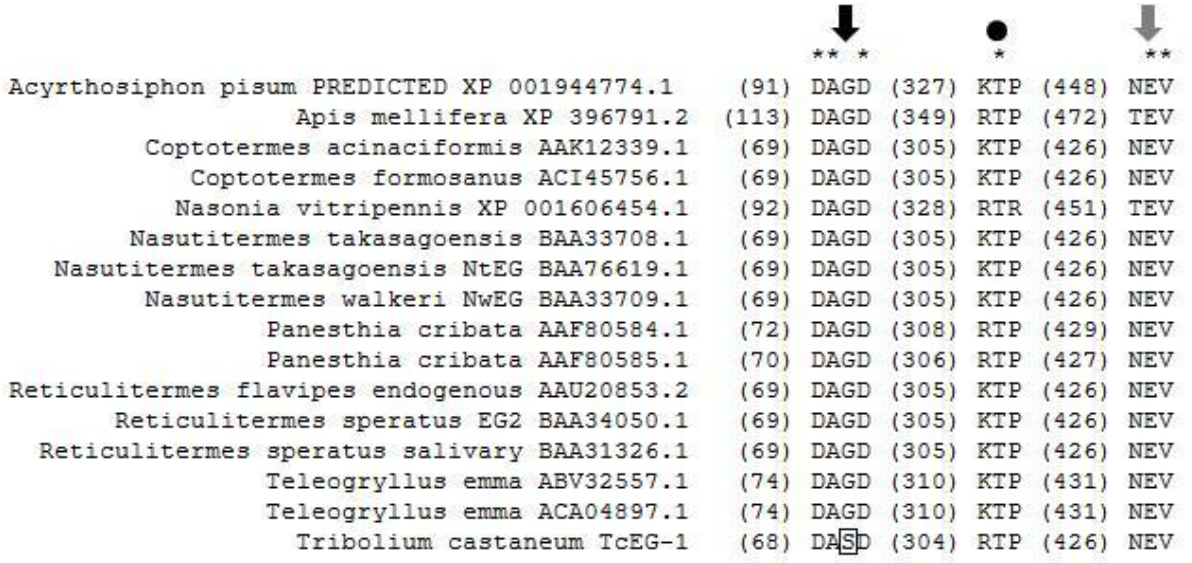


Figure 6: Sequence alignment of insect cellulases belonging to the GHF 9 hydrolase family. The black arrow denotes a proton donor catalytic region. * denotes conservation amongst all sequences. Square region denotes Gly mutation for TcEG1 . Conserved O-glycosylation Thr denoted by black circle. Proton acceptor is Glu denoted by the grey arrow.

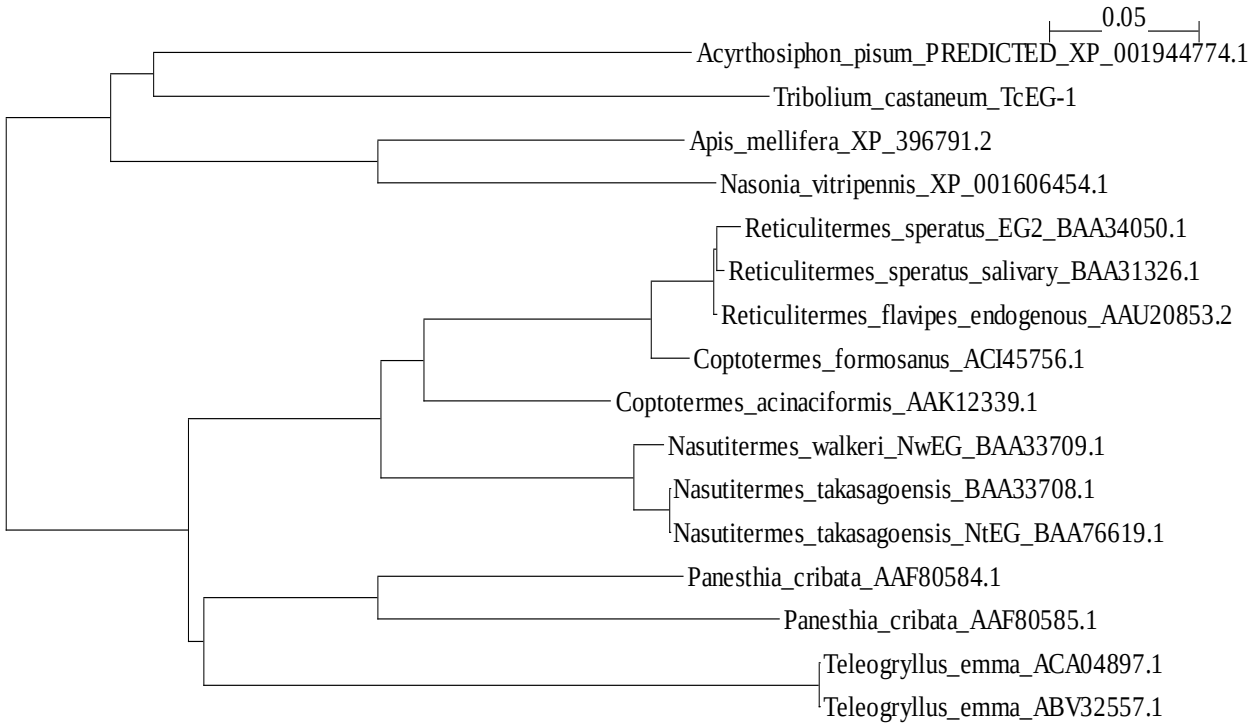


Figure 7: Phylogenetic tree obtained by using complete protein sequences of selected and predicted endoglucanases belonging to GHF9 from insects/symbionts with ascension numbers from the NCBI database. Vector NTI AlignX program (Invitrogen) was used to generate a basic sequence alignment and unrooted tree to demonstrate homology among protein sequences.

suggests high similarity of TcEG1 to a predicted EG from the pea aphid *Acyrthosiphon pisum* (Johnson) (Hemiptera: Aphididae) (Sabater-Munoz, Legeai et al. 2006), *Apis mellifera* (L.) (Hymenoptera: Apidae) (Kunieda, Fujiyuki et al. 2006), and *Nasonia vitripennis* (Ashmead) (Hymenoptera: Pteromalidae) all of which were also originally found by genome annotation.

Primers for PCR were designed based on the genomic TcEG1 sequence and used to amplify a 1,356 bp product from cDNA made from larval *T. castaneum* gut tissue (Figure 8 A). Amplicons were cloned into the pET30a and pIZT-V5/His vectors for heterologous protein expression in bacterial and insect cell culture systems respectively. Expression of TcEG1 protein in *E. coli* cultures, or secreted in the media in DS2 cell cultures was detected using antisera to an N- terminal (*E. coli*) or a C-terminal (DS2 cells) His tag (Figure 8 B). In agreement with the predicted molecular weight for TcEG1, expressed protein in DS2 insect cells was 50 kDa in size, while a 35 kDa protein was detected in *E. coli* cultures. This discrepancy in size is due to a premature stop codon being introduced via a PCR amplification error in the pET30a/TcEG1 expression construct. An adenosine was deleted from DNA position 809 in the cDNA, resulting in a frame shift and generation of a stop codon at amino acid 277, thus removing the proton acceptor site for catalytic activity. Lack of full length TcEG1 expression in *E. coli* prevented us from comparing TcEG1 expressed in bacterial and insect systems to test the importance of glycosylation for TcEG1 activity. However, previous documentation with AgEGase 1 and AgEGase 1 belonging to GHF45 from *Apriona germari* demonstrated that N-glycosylation is

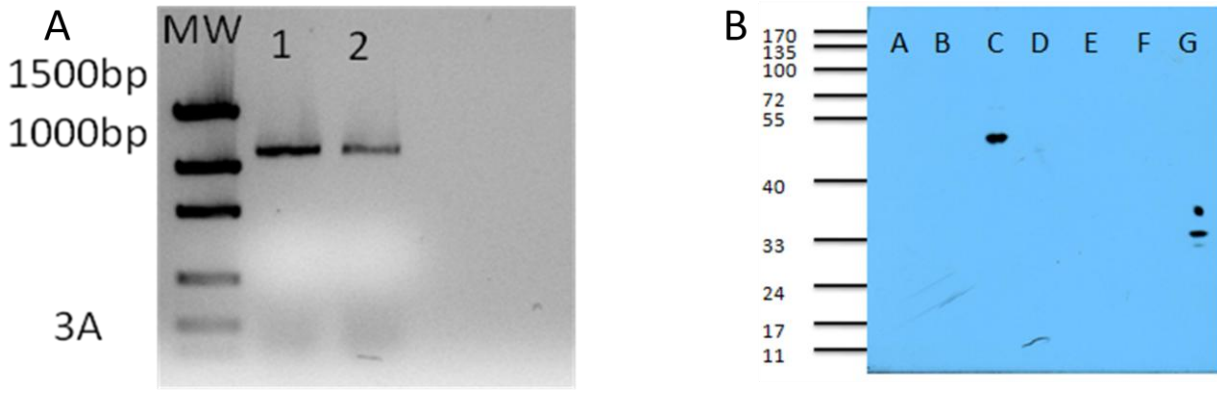


Figure 8: Cloning or TcEG1 and expression in DS2 cell and E. coli cultures. **(A)** 1% Agarose gel showing TcEG1 amplicon (1,356 bp) amplified from *T. castaneum* larval head (lane 1) or gut tissue (lane 2). **(B)** Detection of TcEG1 expressed in *E. coli* or DS2 cell cultures by Western blotting. Proteins (X μg)in supernatant (A, B, and C) or pellet (D, E, and F) from DS2 cell cultures transfected with cellfectin only (A and D), pIZT (B and E), or pIZT-TcEG1 (C and F) were separated by electrophoresis and blotted on PVDF filters. Expressed TcEG1 was detected by probing with antisera against His tag. TcEG1 produced and purified from *E. coli* is shown in (G) for comparison.

important for activity in the conserved catalytic site NSTF (Wei, Lee et al. 2005; Wei, Lee et al.2006). Further work would be necessary to establish the importance of glycosylation for TcEG1 activity.

Enzymes within gut fluids from .*T. castaneum* larvae and recombinant TcEG1 were tested for cellulolytic activity with CMC and MCC as substrates. Activity against CMC is typically used as proxy for EG activity (Xiao, Storms et al. 2005), while activity against MCC is considered a substrate for estimating CBH activity (Chundawat, Balan et al. 2008). Activity against MCC was not detectable for any of the tested larval and recombinant samples (data not shown). In contrast, all TcEG1 and larval samples displayed activity against CMC (Fig.9), demonstrating that TcEG1 is an EG. *T. castaneum* gut fluids displayed detectable activity, evidence for the expression of EG cellulolytic enzymes in this insect. The highest activity against CMC in this study was from the secreted TcEG1 from DS2 insect cell cultures (385.1 U/mg). This level of activity is lower than that reported for TeEG-1 from *T. emma* expressed in Sf9 cells (948.1 U/mg) (Kim, Choo et al. 2008). Activity from a recombinant cellulase from *Coptotermes formasanus* (Shiraki) (Isoptera: Rhinotermitidae) expressed in *E. coli* was at 19 and 16 U/mg for native and His-tagged proteins, respectively (Zhang, Lax et al. 2009). The relatively low activity for the *C. formasanus* enzyme compared to *T. emma* cellulase may be due to lack of glycosylation, as reported for EG cellulases from *A. germari* (Wei, Lee et al. 2005; Wei, Lee et al. 2006). As expected from the lack of the catalytic site, our truncated TcEG1 protein expressed in *E. coli* showed no activity against CMC.

We used two forms of TcEG-1 (native and containing a His tag) expressed in DS2 cells to test for thermal stability and optimum pH conditions for activity. In these assays,

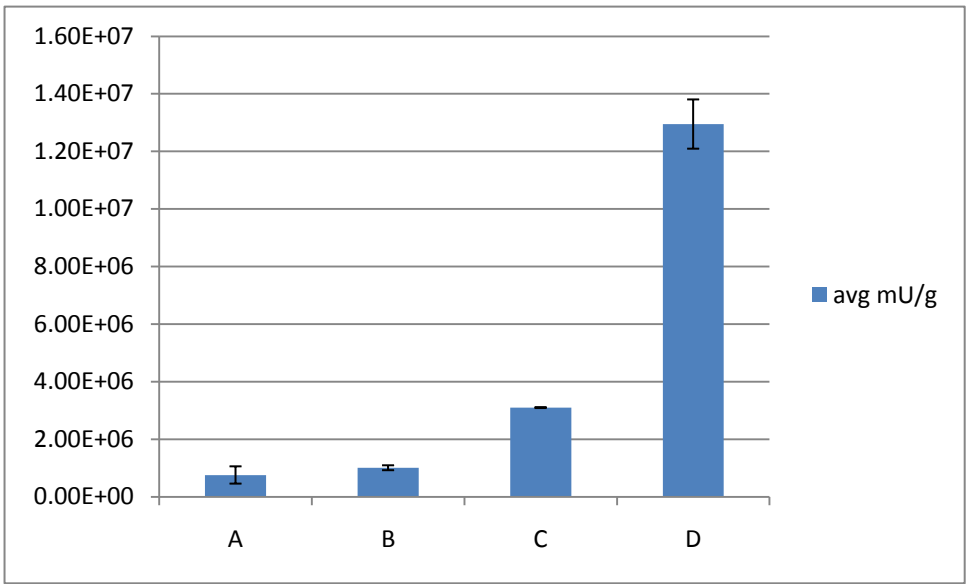
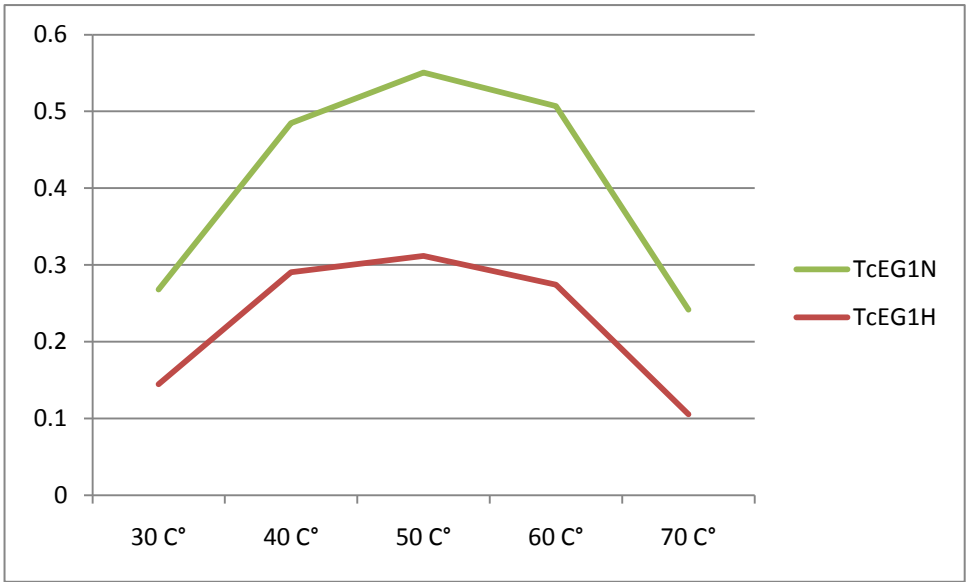


Figure 9: Comparison of cellulase activity in gut fluids from *Tribolium castaneum* larvae and recombinant TcEG-1 protein. (A) *T. castaneum* larval gut fluids, (B) Media from DS2 cells transfected with cellfectin (negative control), (C) media from DS2 cells transfected with pIZT, or (D) media from DS2 cells transfected with pIZT/V5/TcEG1. Data for each experiment represents the mean and standard deviation from three independent determinations of CMC degradation from three independent samples.

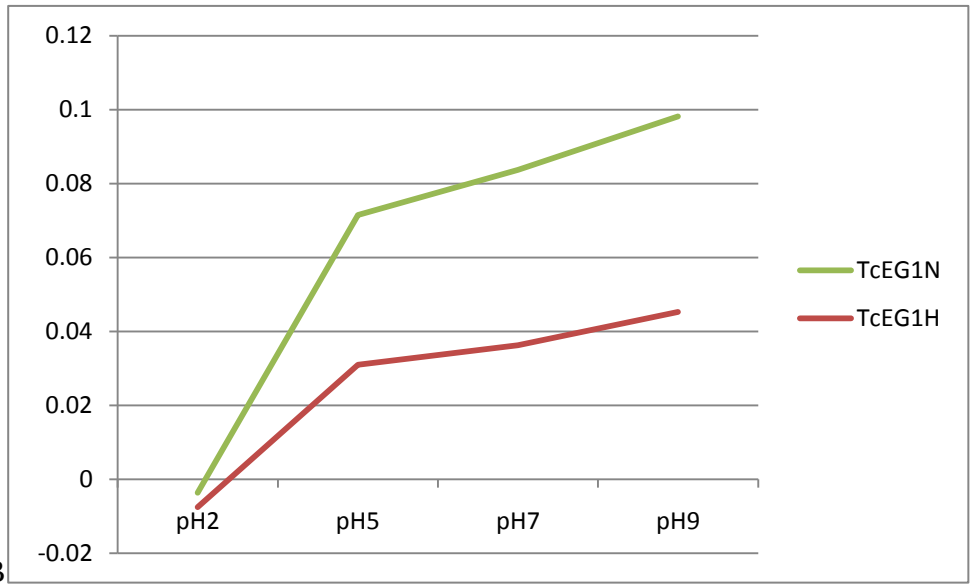
TcEG-1 displayed stable activity against CMC up to 50C°, while native TcEG-1 displayed similar thermostability but higher levels of activity, suggesting a negative influence of His-tagging on activity (Fig. 10A). There are previous reports that present evidence on the detrimental effect of addition of a His tag for insect cellulase activity. For example, a C-terminal tagged EG from *C. formosanus* (CfEG3a) was showed to display reduced activity when compared to native CfEG3a (Zhang, Lax et al. 2009). Possible explanations for this activity decline include difficulties with protein folding or reducing the structural integrity of the protein (Terpe 2003). The thermostability data for either form of TcEG1 suggests that this enzyme is not as thermostable as other insect cellulases, like cellulases from *Apriona germari* (Hope) (Coleoptera: Cerambycidae) showing activity at 60°C (Wei, Lee et al. 2006). Optimal pH was found to be at pH 9 for both tagged and non-tagged enzymes, although again the non-tagged form displayed higher activity (Fig. 10B). This high pH optimum for TcEG1 activity does not agree to values typically found for alternative insect cellulases which have optima at around pH 4-5 (Watanabe and Tokuda 2001). An exception to this observation is cellulolytic activity in the beetle *Aulacophora foveicollis* (L.) (Hemiptera: Aphididae), which was found to have an optimal pH of 7.8 (Sami 2008).

Evidence presented here represents the first functional evidence for an EG from *T. castaneum*. Future research for TcEG1 is needed to determine localization of activity for the protein within the *T. castaneum* digestive tract, as well as changes in expression through developmental stadia. If TcEG1 activity is crucial for larval survival, it could be exploited as a potential insecticide target using RNAi, as previously demonstrated for the

A



B



B

Figure 10: Testing thermostability and pH optimiza for soluble TcEG-1 activity against CMC substrate in non-tagged TcEG-1 (TcS) or His tagged TcEG1 (TcH). Absorbances were corrected for media in extracts.

Cell-1 cellulase gene from *Reticulitermes flavipes* (Kollar) (Isoptera: Rhinotermitidae) (Zhou, Wheeler et al. 2008), or inhibitors, as reported for *A. foveicollis* cellulase (Pant and Ramana 1989). Optimization of TcEG1 through genetic engineering (Zhang, Lax et al. 2009) could enhance activity and thermostability for incorporation of TcEG1 in enzyme mixtures in bioreactors to degrade lignocellulosic material for production of ethanol.

References:

Araujo, R. A., R. N. C. Guedes, et al. (2008). "Enhanced proteolytic and cellulolytic activity in insecticide-resistant strains of the maize weevil, *Sitophilus zeamais*." Journal of Stored Products Research **44**: 354-359.

Banks, D. J., G. Hua, et al. (2003). "Cloning of a *Heliothis virescens* 110 kDa aminopeptidase N and expression in Drosophila S2 cells." Insect Biochem Mol Biol **33**(5): 499-508.

Bayer, E. A., H. Chanzy, et al. (1998). "Cellulose, cellulases and cellulosomes." Curr Opin Struct Biol **8**(5): 548-57.

Bhatia, Y., S. Mishra, et al. (2002). "Microbial β -glucosidase: cloning, properties, and applications. ." Critical Reviews in Biotechnology **22**(4): 375-407.

Bourne, Y. and B. Henrissat (2001). "Glycoside hydrolases and glycosyltransferases: families and functional modules." Curr Opin Struct Biol **11**(5): 593-600.

Bravo, A., S. S. Gill, et al. (2007). "Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control." Toxicon **49**(4): 423-435.

Cazemier, A. E., J. M. Huub, et al. (1997.). "Fibre digestion in arthropods." Comparative Biochemistry and Physiology **118A**(1): 101-109.

Chundawat, S. P., V. Balan, et al. (2008). "High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass." Biotechnol Bioeng **99**(6): 1281-94.

Clarke, A. J. (1997). Biodegradation of cellulose: enzymology and biotechnology. . Lancaster, PA., Technomic Publishing Company.

Cleveland, L. R. (1924). "The physiology and symbiotic relationships between the intestinal protozoa of termites and their host, with special reference to *Reticulitermes flavipes*. ." The Biological Bulletin **46**(4): 117-227.

Clissold, F. J., G. D. Sanson, et al. (2004). "Indigestibility of plant cell wall by Australian plague locust, *Chortoicetes terminifera*. ." Entomologia experimentalis et applicata **112**(3): 159-168.

Consortium, *T.* G. S. (2008). "The genome of the model beetle and pest *Tribolium castaneum*." Nature **452**(24): 949-955.

Davies, G. and B. Henrissat (1995). "Structures and mechanisms of glycosyl hydrolases." Structure **3**(9): 853-9.

Davis, G. R. F. (1963). "Carbohydrase activity of alimentary canal homogenates of five grasshopper species (Orthoptera: Acrididae)." Archives Internationales de Physiologie et de Biochimie **71**(2): 166-174.

Delalibera, I. J., J. Handelsman, et al. (2005). "Contrasts in Cellulolytic Activities of Gut Microorganisms Between the Wood Borer, *Saperda vestita* (Coleoptera: Cerambycidae), and the Bark Beetles, *Ips pini* and *Dendroctonus frontalis* (Coleoptera: Curculionidae)." Environmental Entomology **34**(3): 541-547.

Demirbas, A. (2007). "Progress and recent trends in biofuels." Progress in Energy and Combustion Science **33**: 1-18.

Ducros, V., M. Czjzek, et al. (1995). "Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5." Structure **3**(9): 939-49.

Ferreira, A. H., S. R. Marana, et al. (2001). "Purification, molecular cloning, and properties of a beta-glycosidase isolated from midgut lumen of *Tenebrio molitor* (Coleoptera) larvae." Insect Biochem Mol Biol **31**(11): 1065-76.

Ferreira, C. and W. R. Terra (1983). "Physical and kinetic properties of a plasma-membrane-bound beta-D-glucosidase (cellobiase) from midgut cells of an insect (*Rhynchosciara americana* larva)." Biochem J **213**(1): 43-51.

Gijzen, H. J., C. van der Drift, et al. (1994). "Effect of Host Diet and Hindgut Microbial Composition on Cellulolytic Activity in the Hindgut of the American Cockroach, *Periplaneta americana*." Appl Environ Microbiol **60**(6): 1822-1826.

Girard, C. and L. Jouanin (1999). "Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, *Phaedon cochleariae*." Insect Biochem Mol Biol **29**(12): 1129-42.

Gruno, M., P. Valjamae, et al. (2004). "Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate." Biotechnol Bioeng **86**(5): 503-11.

Henrissat, B. (1991). "A classification of glycosyl hydrolases based on amino acid sequence similarities." Biochem J **280 (Pt 2)**: 309-16.

Henrissat, B. and A. Bairoch (1993). "New families in the classification of glycosyl hydrolases based on amino acid sequence similarities." Biochem J **293 (Pt 3)**: 781-8.

Henrissat, B. and G. Davies (1997). "Structural and sequence-based classification of glycoside hydrolases." Curr Opin Struct Biol **7**(5): 637-44.

Holtzapple, M., M. Cognata, et al. (1990). "Inhibition of *Trichoderma reesei* cellulase by sugars and solvents." Biotechnol Bioeng **36**(3): 275-87.

Itoh, M., S. Takeda, et al. (1991). "Cloning and sequence analysis of membrane-bound alkaline phosphatase cDNA of the silkworm, *Bombyx mori*." Biochim Biophys Acta **1129**(1): 135-138.

Jacobson, R. L. and Y. Schlein (1997). "Cellulase activity of *Leishmania major* in the sandfly vector and in culture." J Eukaryot Microbiol **44**(3): 216-9.

Juge, N. (2006). "Plant protein inhibitors of cell wall degrading enzymes." Trends Plant Sci **11**(7): 359-67.

Khademi, S., L. A. Guarino, et al. (2002). "Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*." Acta Crystallogr D Biol Crystallogr **58**(Pt 4): 653-9.

Kim, N., Y. M. Choo, et al. (2008). "Molecular cloning and characterization of a glycosyl hydrolase family 9 cellulase distributed throughout the digestive tract of the cricket *Teleogryllus emma*." Comp Biochem Physiol B Biochem Mol Biol **150**(4): 368-76.

Klesov, A. A. (1991). "Biochemistry and enzymology of cellulose hydrolysis." Biokhimiya **55**: 1295-1318.

Kumar, R., S. Singh, et al. (2008). "Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives." J Ind Microbiol Biotechnol **35**(5): 377-91.

Kunieda, T., T. Fujiyuki, et al. (2006). "Carbohydrate metabolism genes and pathways in insects: insights from the honey bee genome." Insect Mol Biol **15**(5): 563-76.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-5.

Lee, S. J., S. R. Kim, et al. (2004). "cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*." Comp Biochem Physiol B Biochem Mol Biol **139**(1): 107-16.

Lee, S. J., K. S. Lee, et al. (2005). "A novel cellulase gene from the mulberry longicorn beetle, *Apriona germari*: gene structure, expression, and enzymatic activity." Comp Biochem Physiol B Biochem Mol Biol **140**(4): 551-60.
Lehane, M. J. and P. F. Billingsley (1996). Biology of the insect midgut. London, Chapman and Hall.
Li, L., J. Frohlich, et al. (2003). "Termite gut symbiotic archaezoa are becoming living metabolic fossils." Eukaryot Cell **2**(5): 1091-8.
Lo, N., G. Tokuda, et al. (2000). "Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. ." Current Biology **10**(13): 801-804.
Lynd, L. R., M. S. Laser, et al. (2008). "How biotech can transform biofuels." Nat Biotechnol **26**(2): 169-72.
Marana, S. R., W. R. Terra, et al. (1995). "Midgut B-D-Glucosidases from *Abracis flavolineata* (Orthoptera: Acrididae). Physical properties, substrate specificities and function." Insect Biochem Molecular Biology **25**(7): 835-843.
Martin, M. M. (1983). "Cellulose digestion in insects." Comparative Biochemistry and Physiology **75A**(3): 313-324.
Martin, M. M. (1991). "The Evolution of Cellulose Digestion in Insects." Philosophical Transactions: Biological Sciences **33**(1267): 281-288.
Mekala, N. K., R. R. Singhania, et al. (2008). "Cellulase production under solid-state fermentation by *Trichoderma reesei* RUT C30: statistical optimization of process parameters." Appl Biochem Biotechnol **151**(2-3): 122-31.
Miller, G. L. (1959). "Use of the dinitrosalicylic acid reagent for the determination of reducing sugars. ." Analytical Chemistry **31**(3): 426-428.
Mo, J., T. Yang, et al. (2004). "Cellulase activity in five species of important termites in China." Applied Entomological Zoology **39**(4): 635-641.
Morgan, M. R. J. (1975). "Relationship between gut cellobiose, lactase, aryl B-glucosidase, and aryl B-galactosidase activities of *Locusta migratoria*." Insect Biochemistry **5**: 609-617.
Morris, K., M. D. Lorenzen, et al. (2009). "The *Tribolium castaneum* larval gut transcriptome and proteome: A resource for the study of the coleopteran gut." J Proteome Res.
Mousdale, D. M. (2008). Biofuels: biotechnology, chemistry, and sustainable development. Boca, Raton Florida, CRC Press.
Nakashima, K., H. Watanabe, et al. (2002). "Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki." Insect Biochem Mol Biol **32**(7): 777-84.
Nakonieczny, M., K. Michalczyk, et al. (2006). "Midgut glycosidases activities in monophagous larvae of Apollo butterfly, *Parnassius apollo ssp. frankenbergeri*." C R Biol **329**(10): 765-74.
Ni, J., G. Tokuda, et al. (2007). "Heterologous expression and enzymatic characterization of B-glucosidase form the drywood-eating termite, *Neotermes koshunensis*." Applied Entomological Zoology **42**(3): 457-463.
Oppert, C., W. E. Klingeman, et al. (2009). "Prospecting for cellulolytic activity in insect digestive fluids." Comp Biochem Physiol B Biochem Mol Biol.
Pant, R. and D. Ramana (1989). "Cellulolytic activity in a phytophagous lepidopteran insect *Philosamia ricini*: the orgin of the enzymes." Insect Biochemistry **19**(3): 269-276.

Pfadt, R. E. (1994). "Field Guide to Common Western Grasshoppers." Wyoming Agriculture Experiment Station Bulletin 912 Retrieved June 9th, 2008, from http://www.sidney.ars.usda.gov/grasshopper/extrnlpg/ghwywest/fieldgde.htm.

Pradella, S., A. Hans, et al. (2002). "Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56." Arch Microbiol **178**(6): 484-92.

Richards, S., R. A. Gibbs, et al. (2008). "The genome of the model beetle and pest *Tribolium castaneum*." Nature **452**(7190): 949-55.

Rodrigo-Simon, A., S. Caccia, et al. (2008). "*Bacillus thuringiensis* Cry1Ac toxin-binding and pore-forming activity in brush border membrane vesicles prepared from anterior and posterior midgut regions of lepidopteran larvae." Appl. Environ. Microbiol. **74**(6): 1710-1716.

Sabater-Munoz, B., F. Legeai, et al. (2006). "Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera)." Genome Biol **7**(3): R21.

Sami, A. J., Sharkoori, A.R. (2008). "Biochemical characterization of endo-1,4-B-D-glucanase activity of a green insect pest *Aulacophora foveicollis* (Lucas)." Life Science Journal **5**(2): 30-36.

Scharf, M. E., Tartar, A. (2008). "Termite digestomes as sources for novel lignocellulases." Biofuels,Bioproducts & Biorefining.

Schiott, M., H. H. De Fine Licht, et al. (2008). "Towards a molecular understanding of symbiont function: identification of a fungal gene for the degradation of xylan in the fungus gardens of leaf-cutting ants." BMC Microbiol **8**: 40.

Schwarz, W. H., K. Bronnenmeier, et al. (1987). "Activity Staining of Cellulases in Polyacrylamide Gels Containing Mixed Linkage B-Glucans " Analytical Biochemistry **164**(1): 72-77.

Scrivener, A. M., M. Slaytor, et al. (1989). "Symbiont-independent digestion of cellulose and starch in *Panesthia cribrata* Saussure, an australian wood-eating cockroach. ." Journal of Insect Physiology **35**(12): 935-941.

Scrivener, A. M., Slaytor, M. (1993). "Properties of the endogenous cellulas from *Panesthia cribrata* Sassure and purification of major endo-β-1,4-glucanase components." Insect Biochem Molecular Biology **24**(3): 223-231.

Somerville, A. C. a. C. (2008). "Cellulosic Biofuels." Annual review of Plant Biology **60**: 165-182.

Sowadski, J. M., M. D. Handschumacher, et al. (1985). "Refined structure of alkaline phosphatase from *Escherichia coli* at 2.8 A resolution." J Mol Biol **186**(2): 417-433.

Sugimura, M., H. Watanabe, et al. (2003). "Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, *Psacothea hilaris*." Eur J Biochem **270**(16): 3455-60.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." Appl Microbiol Biotechnol **60**(5): 523-33.

Tokuda, G., N. Lo, et al. (2005). "Marked variations in patterns of cellulase activity against crystalline- vs carbyoxymethayl-cellulose in the digetive systems of diverse, wood-feeding termites." Physiological Entomology **30**: 372-380.

Tokuda, G. and H. Watanabe (2007). "Hidden cellulases in termites: revision of an old hypothesis." Biol Lett **3**(3): 336-9.

Tokuda, G., H. Watanabe, et al. (1997). "Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis* (Shiraki): distribution of cellulases and properties of endo-beta-1,4-glucanase." Zoolog Sci **14**(1): 83-93.

Treves, D. S., and Martin. M.M. (1994). "Cellulose digestion in primitive hexaopods: effect of ingested antibiotics on gut microbial populations and gut cellulase levels in the firerbrat, *Thermobia domestica* (Zygentoma, Lepismatidae)." Journal of Chemical Ecology **20**(8): 2003-2020.

U.S.DOE (2006). Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda. U. S. D. o. E. O. o. S. a. O. o. E. E. a. R. Energy. Rockville, Maryland.

Van Frankenhuyzen, K. and C. Nystrom (2002). "The *Bacillus thuringiensis* toxin specificity database." http://www.glfc.cfs.nrcan.gc.ca/bacillus

Vasanthakumar, A., J. Handelsman, et al. (2008). "Gut microbiota of an invasive subcortical beetle, *Agrilus planipennis* Fairmaire, across various life stages." Environ Entomol **37**(5): 1344-53.

Waeonukul, R., K. L. Kyu, et al. (2007). "Multiple Cellulases and Xylanases from *Bacillus circulans* B6 During Growth on Avicel under and Aerobic Condition." Thai Journal of Biotechnology **8**(1): 27-32.

Walters, K. H., Smock, L.A. (1991). "Cellulase activity of leaf litter and stream-dwelling, shredder macroinvertebrates." Hydrobiologia **220**: 29-35.

Watanabe, H., H. Noda, et al. (1998). "A cellulase gene of termite origin." Nature **394**(6691): 330-1.

Watanabe, H. and G. Tokuda (2001). "Animal cellulases." Cell Mol Life Sci **58**(9): 1167-78.

Wei, Y. D., K. S. Lee, et al. (2006). "N-linked glycosylation of a beetle (*Apriona germari*) cellulase Ag-EGase II is necessary for enzymatic activity." Insect Biochem Mol Biol **36**(6): 435-41.

Wei, Y. D., K. S. Lee, et al. (2006). "Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, *Apriona germari*." Comp Biochem Physiol B Biochem Mol Biol **145**(2): 220-9.

Wei, Y. D., S. J. Lee, et al. (2005). "N-glycosylation is necessary for enzymatic activity of a beetle (*Apriona germari*) cellulase." Biochem Biophys Res Commun **329**(1): 331-6.

Wharton, D. R., M. L. Wharton, et al. (1965). "Cellulase in the cockroach, with special reference to *Periplaneta americana* (L)." J Insect Physiol **11**(7): 947-59.

Wolf, B., P. Zwick, et al. (1997). "Feeding ecology of the freshwater detritivore *Ptychoptera paludosa* (Diptera, Nematocera)." Freshwater Biology **38**(375-386).

Wyman, C. E., B. E. Dale, et al. (2005). "Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover." Bioresour Technol **96**(18): 2026-32.

Xiao-juan, L. I., Y. Xiong-fei, et al. (2008). "Cellulase in *Anoplophora glabripennis* adults emerging from different host tree species." Forest Studies of China **10**(1): 27-31.

Xiao, Z., R. Storms, et al. (2005). "Microplate-based carboxymethylcellulose assay for endoglucanase activity." Anal Biochem **342**(1): 176-8.

Yang, T., M. J., et al. (2003). "Purification and some properties of cellulase from *Odontotermes formosanus* (Isoptera: Termitidae) " Insect Science **11**(1): 1-10.

Yapi, D., D. Gnakri, et al. (2009). "Purification and biochemical characterization of a specific β"-glucosidase from the digestive fluid of larvae of the palm weevil, *Rhynchophorus palmarum*." Journal of Insect Science **9**(4).

Yu, S. J. (1989). "β-Glucosidase in four phytophagous lepidoptera." Insect Biochemistry **19**(1): 103-108.

Zhang, D., A. R. Lax, et al. (2009). "Differential cellulolytic activity of native-form and C-terminal tagged-form cellulase derived from *Coptotermes formosanus* and expressed in E. coli." Insect Biochem Mol Biol.

Zhou, X., J. A. Smith, et al. (2007). "Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*." Gene **395**(1-2): 29-39.

Zhou, X., M. M. Wheeler, et al. (2008). "RNA interference in the termite *Reticulitermes flavipes* through ingestion of double-stranded RNA." Insect Biochem Mol Biol **38**(8): 805-15.

Zverlov, V. V., W. Holl, et al. (2003). "Enzymes for digestion of cellulose and other polysaccharides in the gut of longhorn beetle larvae, *Rhagium inquisitor* L. (Col., Cerambycidae)." International Biodeterioration & Biodegradation **51**: 175-179.

Appendix

This Appendix is a published version of a paper by the same title to accepted by the journal *Insect Biochemistry and Molecular Biology* by Omaththage P. Pera, Jonathan D. Willis, Michael J. Adang, and Juan Luis Jurat-Fuentes. The use of "our" in this chapter refers to my co-author and me. My primary contributions to this paper include (1) the experimental work for cloning and heterologous protein expression, (2) the collection and analysis of data from my experimental work, (3) part of the writing for materials and methods, (4) and review of the final manuscript.

Cloning and characterization of the Cry1Ac-binding alkaline phosphatase (HvALP) from *Heliothis virescens*

Omaththage P. Perera[a], Jonathan D. Willis[b], Michael J. Adang[c], Juan L. Jurat-Fuentes[b]*

[a]Southern Insect Management Rsearch Unit, USDA-ARS, Stoneville, MS 38776, USA.

[b]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA

[c]Departments of Entomology, and Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30605, USA.

*To whom correspondence should be addressed

E-mail: jurat@utk.edu

Tel: (865) 974-5931

Fax: (865) 946-1769

ABSTRACT

Membrane bound alkaline phosphatases (mALPs, EC 3.1.3.1) in the insect midgut have been reported as functional receptors for Cry toxins from the bacterium *Bacillus thuringiensis*. We previously reported the identification of HvALP in the midgut of *Heliothis virescens* larvae as a Cry1Ac binding protein that is down-regulated in Cry1Ac-resistant insects. To further characterize HvALP, we localized mALP protein to foregut and midgut tissues using anti-mALP serum and then cloned five mALPs from *H. virescens* larval midgut. All five clones displayed high levels of sequence identity (above 90%), suggesting that they may represent allelic variants, and grouped with other lepidopteran mALPs in sequence alignments. All these cloned ALPs were predicted to contain a glycosylphosphatidylinositol (GPI) anchor and were named HvmALP1 to 5. We expressed two of the most diverse HvmALPs in a heterologous system to test binding of Cry1Ac and recognition by HvALP-cross reacting antiserum. Our data highlight the importance of glycosylation for Cry1Ac binding to HvALP and suggest that, depending on glycosylation, all the identified HvmALPs may be synonymous with HvALP, the Cry1Ac-binding phosphatase identified in *H. virescens* midgut epithelium.

Keywords: Alkaline phosphatase, Cry1Ac binding, *Heliothis virescens, Bacillus thuringiensis,* Lepidoptera

INTRODUCTION

The insecticidal Cry1Ac toxin from the bacterium *Bacillus thuringiensis* (Bt) is expressed in transgenic Bt cotton cultivars to effectively control *Heliothis virescens* larvae (Perlak et al. 2001). The current model for Cry1A intoxication in susceptible larvae includes toxin solubilization and activation in the insect midgut fluids, followed by binding to cadherin-like proteins. According to the model of Bravo et al (Bravo et al. 2007), binding of activated toxin monomers to cadherin results in additional toxin processing by an unidentified protease. This additional cleavage results in formation of toxin oligomers, which in the case of Cry1Ac display high affinity binding to N-acetylgalactosamine (GalNAc) residues on N-aminopeptidases (APNs) (Pardo-Lopez et al. 2006), and possibly alkaline phosphatases (ALPs) (Jurat-Fuentes and Adang 2004). This second binding step is conducive to toxin insertion into the cell membrane and formation of a pore that leads to cell death by osmotic cell lysis. Alternatively, it has also been proposed that binding of Cry1Ac monomers to cadherin activates an intracellular cell death pathway (Zhang et al. 2006).

We previously reported the identification of a membrane-bound form of alkaline phosphatase (HvALP) as a Cry1Ac-binding protein in the *H. virescens* midgut brush border membrane (Jurat-Fuentes and Adang 2004; Krishnamoorthy et al. 2007). HvALP is attached to the cell membrane by a glycosylphosphatidylinositol (GPI) anchor and contains a terminal N-linked GalNAc residue recognized by Cry1Ac (Jurat-Fuentes and Adang 2004). Levels of HvALP are reduced in larvae of the YHD2-B strain of *H. virescens*, which are highly resistant to Cry1Ac (Jurat-Fuentes and Adang 2004). Furthermore, HvALP levels directly correlate with

susceptibility in backcrosses (Jurat-Fuentes and Adang 2007), suggesting a correlation between Cry1Ac binding to HvALP and toxicity.

The goal of this project was to further characterize HvALP and its interaction with Cry1Ac. HvALP was detected only in the foregut and midgut epithelia. Based on this information, we used mRNA purified from *H. virescens* larval gut and rapid amplification of cDNA ends (RACE) to obtain five clones encoding mALPs. Using heterologous expression of the two most diverse clones in *Escherichia coli*, we confirmed the importance of glycosylation for Cry1Ac binding and identified all mALP clones as synonymous with HvALP.

MATERIALS AND METHODS

Insects

Heliothis virescens larvae from a laboratory colony maintained at the Southern Insect Management Research Unit, USDA-ARS, (Stoneville, MS) were used as the source of genetic and protein materials for this work. Larvae were reared on artificial diet (modified from Shaver and Raulston, 1971) and fourth instar larvae used for dissection to obtain specific tissues.

Cry1Ac toxin purification and labeling.

B. thuringiensis strain HD-73 obtained from the *Bacillus* Genetic Stock Center (Ohio, USA) was used to produce Cry1Ac crystals. Crystalline inclusions were solubilized by incubation overnight at 30°C in 50 mM Na_2CO_3 pH 9.8, 100 mM NaCl 0.1% 2-β-mercaptoethanol. After solubilization, samples were clarified by centrifugation at 30,000 x g for 20 min. at 4°C. Cry1Ac protoxin in the supernatant was activated by treatment with 0.1% (v/v) of midgut fluids obtained from actively feeding 4th instar *H. virescens* larvae. After incubation for 1 h. at room temperature, the activated toxin sample was loaded on

a HiTrap HP Q 5ml anion exchange column (GE Healthcare) pre-equilibrated with 50 mM Na_2CO_3 pH 9.8 (buffer A). Activated Cry1Ac toxin was eluted from the column using a linear gradient of 50 mM Na_2CO_3 pH 9.8, 1 M NaCl (buffer B). Purified toxin samples (verified by reducing SDS-8% PAGE) were pooled, protein concentration determined using the method of Bradford (Bradford 1976) with BSA as standard, and stored at –80°C until used.

Purified Cry1Ac toxin (1 mg) was biotinylated using a 1:30 molar ratio of EZ link NHS-LC-biotin (Pierce) following manufacturer's instructions. After biotinylation, labeled toxin samples were extensively dialyzed in 20 mM Na_2CO_3 pH 9.8, 150 mM NaCl at 4°C. Labeled toxins were quantified as above before use in ligand blotting.

Two dimensional electrophoresis (2DE) separation of brush border membrane vesicle (BBMV) proteins

Brush border membrane vesicles (BBMV) were isolated as described elsewhere (Wolfersberger et al. 1987). BBMV proteins were extracted by precipitation and then solubilized as previously described (Krishnamoorthy et al. 2007). Solubilized proteins were quantified using the 2D Quant Kit (GE Healthcare) following manufacturer's instructions and kept at -80°C until used.

BBMV protein samples (50 μg) were used to rehydrate 7 cm, pH 4-7 Immobiline DryStrips (GE Healthcare) overnight in rehydration buffer (solubilization buffer plus 0.018 M dithiothreitol [DTT], and 0.5% ampholytes). Following rehydration, strips were subjected to isoelectric focusing (IEF) using a Multiphor II unit (GE Healthcare) following manufacturer's recommendations. Temperature was maintained at 20°C throughout focusing. Focused strips

were equilibrated for 15 minutes in equilibration buffer (6 M urea [Plus-One; GE Healthcare], 2% SDS, 30% glycerol, 0.05 M Tris [pH 8.8], 0.002% bromophenol blue) containing 1% DTT followed by a second equilibration for 15 minutes in equilibration buffer plus 4% iodoacetamide. For second dimension separation we used an EttanDalt six electrophoresis system (GE Healthcare) and SDS-8% PAGE gels. Separated proteins were transferred to PVDF filters overnight and processed as described for Western and ligand blots.

Localization of HvALP expression in larval tissues

For tissue extract preparation, 4th instar *H. virescens* larvae were carefully dissected and the specific tissues (foregut, midgut, hindgut, muscle, Malphigian tubules, fat body, hemolymph, and gut fluid) were collected in centrifuge tubes containing 0.5 ml of PBS (135 mM NaCl, 2 mM KCl, 10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, pH 7.5). Samples were homogenized using disposable pellet pestles and then 50 μl of 10% SDS was added to each sample. Protein extracts were clarified by centrifugation for 5 min. at 21,000 x *g* at room temperature. Supernatants were collected and protein concentrations quantified using the 2D-quantKit (GE Healthcare) with BSA as standard.

Tissue extracts (20 μg) or BBMV (15 μg) proteins were mixed with loading buffer (Laemmli 1970), heat denatured or loaded directly and resolved using SDS-8%PAGE gels. After electrophoretic separation, gels were used for alkaline phosphatase (ALP) activity staining (Jurat-Fuentes and Adang 2004), total protein staining with ProtoBlue™ safe stain (National Diagnostics, GA), or transferred overnight to polyvinylidene difluoride (PVDF) filters at 4°C with 20 V constant voltage.

For ALP activity staining, samples were not heat denatured before gel loading. Following electrophoresis gels were washed with ALP buffer (100 mM Tris/HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$) for 15 minutes at room temperature. After addition of 330 μg/ml of *p*-nitro blue tetrazolium chloride (NBT) and 165 μg/ml of 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) to the ALP buffer, ALP activity was visualized by the formation of a purple-red precipitate. Reactions were stopped by incubation of gels in 50 ml of PBS pH 7.5 containing 200 μl of 500 mM EDTA pH 8.0.

For Western blotting, following electrotransfer filters were blocked for one hour in PBST (135 mM NaCl, 2 mM KCl, 10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, pH 7.5, 0.1% Tween-20) plus 3% BSA. After blocking, all incubations and washes were in PBST plus 0.1% BSA. For detection of HvALP, blocked filters were probed with a 1:10,000 dilution of rabbit antiserum developed against the full length *Anopheles gambiae* membrane–bound alkaline phosphatase (mALP) expressed in *E. coli*, generously provided by Dr. Gang Hua (University of Georgia, Athens, GA). Goat anti-rabbit serum conjugated with horseradish peroxidase (HRP) was used as secondary antibodies and blots developed using Supersignal West Pico enhanced chemiluminescence substrate (Pierce).

For ligand blots 0.1 μg/ml of biotinylated Cry1Ac was used to probe blocked filters for one hour at room temperature. After washing, blots were probed with streptavidin-HRP conjugate (Invitrogen) at a 1:30,000 dilution for one hour at room temperature. Proteins binding Cry1Ac were detected with enhanced chemiluminescence substrate as above.

RNA extraction and cDNA synthesis

Midguts of fourth instar *H. virescens* larvae were dissected and food boli were carefully removed before snap freezing in a 1.5 ml micro-centrifuge tube kept on dry ice. Pools of twenty midguts were homogenized in 1 ml of Trizol reagent (Invitrogen) followed by addition of 200 µl of chloroform and vortexing for 10 sec. The mixture was incubated at room temperature for 15 minutes and centrifuged at 16,000 x *g* for 10 min. at room temperature. The top aqueous phase was carefully transferred to a new 1.5 ml micro-centrifuge tube containing 600 µl of isopropanol and total RNA was precipitated by centrifuging at 16,000 g for 15 min. in a refrigerated centrifuge. RNA pellets were re-suspended in 200 µl of 10 mM Tris-HCl, pH 8.0.

Messenger RNA was recovered from the total RNA preparation using the poly(A)$^{+}$-Tract mRNA purification system (Promega) following manufacturer's instructions. Briefly, approximately 100 µg of total RNA was hybridized with 2 µl of 10 µM Biotin labeled oligo(dT) in a buffer containing 6 X SSC (900 mM NaCl, 90 mM Sodium Citrate, pH 7.0) and 0.1% SDS. The biotinylated oligo(dT) annealed to mRNA was added to streptavidin coated paramagnetic beads pre-washed three times with 0.5X SSC, followed by incubation at room temperature for 10 min. with intermittent mixing by inversion to facilitate binding of biotin-oligo(dT) to paramagnetic beads. The magnetic beads were captured using a magnetic tube stand and the liquid was aspirated out carefully. Three washes with 0.1 X SSC were performed to remove contaminants and the mRNA was released from the magnetic beads by adding 50 µl of RNAse free water followed by magnetic capture of particles. The aqueous phase containing mRNA was transferred to a new tube and was used immediately in the synthesis of full length cDNA using the GeneRacer™ RLM-RACE kit (Invitrogen) following manufacturer's instructions. Briefly, approximately 2 µg of mRNA was treated with calf intestinal phosphatase to remove 5'-

phosphate in non-capped/degraded mRNA, followed by phenol extraction and precipitation. Dephosphorylated mRNA was treated with tobacco acid phosphatase to remove the cap structure from the full length mRNA. Resulting full-length mRNA with a 5'-phosphate molecule was extracted with phenol and precipitated prior to ligation with the RNA oligonucleotide provided with the GeneRacer™ kit. Reverse transcription of the mRNA with a T7-Oligo(dT) primer yielded cDNA containing the T7 promoter sequence at the 3'-end and the synthetic GeneRacer™ 5'-primer sequence at the 5'-end. Double stranded cDNA production was carried out by amplification of the first-strand cDNA for 10 rounds using the GeneRacer™ 5'-primer and T7 primer in a 100 µl reaction containing 20 µl of the reverse transcription reaction, 10 µl of 10X PCR buffer, 4 µl of $MgCl_2$, 2 µl of 10 mM dNTP mix, 30 pmoles each of 5'-RACE primer and T7 Primer, and 2 units of Platinum Taq DNA polymerase (Invitrogen).

Cloning of membrane-bound alkaline phosphatase from H. virescens larval gut tissue

Conceptually translated sequences for *Anopheles gambiae* and *Bombyx mori* membrane bound ALP sequences from the National Center for Biotechnology Information (NCBI) Entrez server (accession numbers XM 308522 and D90454, respectively) were aligned using AlignX module of Vector NTI Suite version 10.3 (Invitrogen). A degenerate primer ALP1-F (5'-GARACGCACGGYGGMGAYGAYGT-3') and its reverse complement ALP1-R (5'-ACRTCRTCKCCRCCGTGCGTYTC-3'), were designed to a conserved amino acid block and the corresponding area of the DNA sequence alignment. ALP1-R primer and GeneRacer 5'-primer were used to amplify the 5'-end of the *H. virescens* ALP gene from midgut cDNA. Amplified cDNA was cloned into PCR 2.1-TOPO TA cloning vector and Top10-MachT1® cells

were transformed using 3 µl of the amplification reaction. Transformations were plated on Luria Bertani agar plates containing 50 µg/ml kanamycin and 40 µg/ml X-Gal. DNA sequences of recombinant clones were obtained by sequencing with the Big-Dye reagent system and ABI 3730xl instrument (Applied Biosystems). All sequence analyses were carried out at the USDA-ARS mid-South Area Genomics Laboratory, Stoneville, MS.

Primers designed using the sequence information from 5' clones positive for ALP were used with the GeneRacer 3' primer to amplify the 3'-end of the *H. virescens* ALP. Finally, all nucleotide sequences were assembled to generate full-length cDNA sequences. All nucleotide and amino acid analyses were carried out using Vector NTI Suite version 10.3. The PSORT program (http://psort.nibb.ac.jp/) was used to detect protein sorting signals and predict protein localization, the OGPET v1.0 (http://ogpet.utep.edu/OGPET/) and NetNGlyc v1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) programs were used to determine potential O- and N-glycosylation sites, and the big-PI predictor server (http://mendel.imp.ac.at/sat/gpi/gpi_server.html) was used to predict glycosylphosphatidylinositol (GPI) anchoring sites. For sequence alignments we used the AlignX software included in the Vector NTI Advance 10 package (Invitrogen).

Heterologous expression of selected alkaline phosphatase cDNA clones

Membrane-bound *Heliothis virescens* alkaline phosphatase (HvmALP) clones displaying the highest sequence diversity (HvmALP1 and HvmALP2), were subcloned into the pET30a expression vector (Novagen) using PCR primers designed with *EcoRI* and *NotI* restriction sites at the 5' and 3' ends, respectively. To ensure proper protein expression, coding frame in

transformants was verified by sequencing in forward and reverse directions (University of Tennessee Sequencing Facility).

For protein expression, purified vectors were transformed into *E. coli* strain BL21(DE3). Transformants were grown overnight with constant agitation at 37°C in 5 ml of LB broth containing 30 μg/ml kanamycin. The following morning, 100 μl of this overnight culture were used to inoculate 100 ml of LB broth plus 30 μg/ml kanamycin in a 250 ml flask. This culture was allowed to grow at 37°C with constant agitation until reaching OD 0.70 (typically about 4-5 hours). After bacterial cultures reached an OD between 0.7-1, expression was induced with 1 mM IPTG. Aliquots (1 ml) were obtained at a series of time points (0, 2, 4, 6, and 20 hours after induction) to determine optimal expression. This pilot study suggested that peak expression for both clones was observed 2-4 hours after induction (data not shown). Consequently, in subsequent experiments, bacterial cultures were collected for analysis three hours post-induction. After this time, bacterial cultures were collected by centrifugation. Pellets were solubilized in sample buffer (Laemmli 1970) and 15 μl aliquots separated on a 10%SDS-PAGE gel. Expression of HvALP clones was detected using Western blotting as described above.

RESULTS

Tissue expression of HvALP in H. virescens larvae

We previously reported on the identification of HvALP, a Cry1Ac-binding mALP in brush border membrane vesicles (BBMV) from guts of *H. virescens* larvae (Jurat-Fuentes and Adang 2004). Figure 1 shows the detection of *H. virescens* BBMV proteins cross-reacting with antiserum against mALP from *A. gambiae* or biotinylated Cry1Ac on 2DE blots. In those blots, mALP antiserum and Cry1Ac toxin bound to the same chain of BBMV protein spots. The lack

of cross-reactivity with other BBMV proteins in 1D (data not shown) or 2D blots validated the specificity of the anti-mALP serum.

To localize expression of HvALP in *H. virescens* larvae, we screened protein extracts from several larval tissues for both ALP activity and protein detected by anti-mALP serum (Fig. 2). In ALP activity gels phosphatase activity bands of about 68-kDa in size were detected in all tested tissue samples, except for hemolymph and hindgut. These results suggest a broad distribution of alkaline phosphatases in *H. virescens* larval tissues. In comparison, mALP antiserum only detected 68-kDa proteins in tissue extracts prepared from foregut tissue, midgut tissue and BBMV (Fig. 2). The anti-mALP serum also detected a 68-kDa protein in gut fluid, although the signal was weaker than in tissues or BBMV. These results suggest that even though a number of alkaline phosphatase isoforms are present in various *H. virescens* larval tissues, expression of the 68-kDa Cry1Ac-binding HvALP is localized to the foregut and midgut, the latter being the target tissue for Cry1Ac intoxication (Bravo, Gill et al. 2007).

Cloning of full-length HvALP cDNA

Based on HvALP expression being exclusive to gut tissue, we designed a RT-PCR strategy to clone HvmALP cDNA from a larval midgut cDNA library. RT-PCR amplification with degenerate primers to HvALP sequence tags (Krishnamoorthy et al. 2007) and protein regions conserved among insect membrane-bound alkaline phosphatases yielded a partial ALP sequence that was used to design specific primers for Rapid Amplification of cDNA Ends (RACE). After extensive RACE cloning, we obtained clones representing a total of five

different ALP sequences. This multiplicity of ALP isoforms was not the result of PCR errors, since all the clones were repeatedly amplified in independent PCR experiments.

All five sequences had high amino acid identity (more than 94%), which suggested that some of the clones represent allelic variants of the same gene. Most sequence divergence was observed in the 3' UTR region of the clones. Protein sequence alignments (Fig. 3) also denoted this high identity, with most variability observed after the first 190 residues on the N-termini. Sequence alignments revealed that all the clones contained a predicted GPI-anchor site at ^{517}A. Based on this information, we named these clones predicted to be tethered to the cell membrane by a GPI anchor as HvmALP (*H. virescens* membrane-bound alkaline phosphatase) 1 to 5 (accession numbers **FJ416470**, **FJ416471**, **FJ416472**, and **FJ416473** for HvmALP1 to HvmALP4 respectively, and **EF531619** for HvmALP5).

All HvmALP cDNAs are predicted to encode full length proteins of about 59 kDa in molecular size (Fig. 3), suggesting that glycosylation and other posttranslational modifications may make up for the mass difference between these mALP proteins and HvALP (estimated to be 68 kDa) detected in larval tissues (Fig. 2). All five mALP clones contain a signal peptide for localization to the cell surface, a predicted phosphatase domain and active site (^{122}IADSACTAT^{130}, with ^{125}S being the active site), and two potential N-glycosylation sites at residues ^{200}NRTW^{203} and ^{279}NVSH^{282}, while no potential O-glycosylation patterns were detected.

When comparing HvmALP protein sequences with other eukaryotic ALPs (Fig. 4), all clones grouped together with ALPs from *Helicoverpa armigera*, with up to 90% sequence identity (for *H. armigera* AFC40807). Sequence identity to membrane-bound forms of ALP from *Bombyx mori* and *Bombyx mandarina* was lower (61%), and other insect ALPs were less similar to HvmALP sequences. Considering the grouping of the HvmALP clones in these

sequence alignments (Fig. 4), HvmALP1 and HvmALP3 are probably allelic variants of one gene, while HvmALP2, HvmALP4, and HvmALP5 would represent allelic variants of an alternative ALP gene. All HvmALP protein sequences contained the predicted essential amino acids for alkaline phosphatase activity (Sowadski, Handschumacher et al. 1985) in highly conserved protein sequence regions (Fig. 5). Interestingly, the two predicted N-glycosylation sites found in all isoforms of HvmALP (^{200}NRT^{203} and ^{279}NVS^{281}) were conserved in *H. armigera* ALP proteins, but did not align with N-glycosylation sequences in other eukaryotic ALPs, including *Bombyx* sequences.

Heterologous expression of mALP clones and identification of HvALP.

Due to the number of HvmALP clones and the high sequence identity among them, we hypothesized that all of them would be recognized by the antisera to mALP from *A. gambiae*, and thus be synonymous with HvALP. To test this hypothesis, we selected the most sequence-diverse clones, HvmALP1 and HvmALP2 sharing 94% identity, and expressed them in *Escherichia coli* cultures. Protein bands of the expected mass for both isoforms were detected in stained gels and by probing Western blots with antisera to the 6 x histidine tag added by the *E. coli* expression vector (Fig. 6). In ALP activity gels, alkaline phosphatase activity was not detected for these proteins (data not shown), suggesting incorrect folding or posttranslational modifications of the expressed phosphatases. As predicted, both expressed HvmALP proteins were recognized by antiserum to the *A. gambiae* mALP (Fig. 6). Analysis of the stained gels suggests that the HvmALP1 clone was expressed more efficiently, explaining the higher level of signal for this protein in the immunoblots. These results suggest that both isoforms are

synonymous with HvALP. As expected from the reported binding of Cry1Ac to GalNAc residues on HvALP (Jurat-Fuentes and Adang 2004), no binding of biotinylated Cry1Ac to any of the expressed HvmALP clones was detected in ligand blots (Fig. 6). We did observe Cry1Ac binding to small *E. coli* proteins, although these proteins were detected even for control samples, indicating that they do not represent truncated or degraded expressed mALP proteins. As a positive control, Cry1Ac bound to HvALP in *H. virescens* BBMV (Fig. 6, lane 5).

DISCUSSION

Alkaline phosphatases (EC 3.1.3.1) are common abundant enzymes that are mainly involved in removing phosphate groups from organic molecules. In insects, ALPs have been reported to be involved in several biological processes (Eguchi 1995, Chang et al. 1993). Physiological levels of midgut alkaline phosphatase activity have been reported to change during stress or pathogenesis in insects (Kucera and Weiser 1974; Miao 2002; Sukhanova et al. 1996), in some cases related to the low energy state associated with infection (Sujak et al. 1978). We have previously described the identification of HvALP, a membrane-bound alkaline phosphatase (mALP) that binds Cry1Ac toxin in midgut brush border membrane from *H. virescens* (Jurat-Fuentes and Adang 2004). In addition to toxin binding, down-regulation of HvALP in midguts of larvae from Cry1Ac-resistant *H. virescens* larvae (Jurat-Fuentes and Adang 2004; Jurat-Fuentes and Adang 2007), suggests a role for HvALP in Cry1Ac toxicity. Towards the characterization of the role of HvALP in Cry1Ac intoxication, we report the cloning of mALP isoforms from the *H. virescens* midgut through cloning and expression of two diverse mALP clones in a heterologous system, in an attempt to identify HvALP among the ALP isoforms.

In agreement with the reported pattern of localization for *M. sexta* mALP (Chen et al. 2005), HvALP expression was localized to the foregut and midgut regions of the *H. virescens* larval gut. The midgut region is where Cry1Ac toxin mode of action takes place (reviewed in Bravo et al. 2007). The posterior region of the larval midgut has been previously reported to display higher levels of Cry1Ac-induced pore formation than the anterior part of the midgut (Carroll et al. 1997; (Rodrigo-Simon, Caccia et al. 2008). Considering that the *H. virescens* cadherin (HevCaLP) is mostly localized to the anterior part of the gut (Aimanova et al. 2006), localization of HvALP throughout the midgut would favor the sequential participation of multiple receptors in Cry1Ac intoxication as proposed by the model of Bravo et al. (Bravo et al. 2007). After toxin activation and binding to HevCaLP in the anterior midgut region, resulting Cry1Ac toxin oligomers would interact with GalNAc residues on GPI-anchored proteins on lipid rafts (aminopeptidases and alkaline phosphatases) and insert on the cell membrane to create a pore that would lead to osmotic cell lysis. This hypothesis does not exclude the possibility that binding to cadherin may also result in midgut cell death per se as proposed by an alternative model (Zhang et al. 2006).

In contrast to the case of *B. mori* (Takesue et al. 1989), we did not found a soluble (sALP) alkaline phosphatase form in our cloning efforts. It is possible that the protein detected in midgut fluid by anti-mALP serum is a soluble isoform of ALP. Alternatively, it is also possible that mALP isoforms are released into the gut lumen by cleavage of the GPI anchors as is the case in *B. mori* (Takesue et al. 1989).

The high sequence identity observed between HvmALP isoforms suggests that they may represent allelic variants of a limited number of genes. In *B. mori*, one mALP and one sALP

cDNA have been cloned (Itoh et al. 1999; Itoh et al. 1991). In this insect, the single *mALP* and *sALP* genes are arranged in tandem and seem to have been derived from duplication (Itoh et al. 2003). Sequence alignments clearly suggest that the five HvmALP sequences we obtained represent allelic variants of at least two mALP genes. Thus, the HvmALP1 and HvmALP3 proteins would be isoforms of HvmALP2, HvmALP4, and HvmALP5. Interestingly, the *sALP* gene from *B. mori* (Itoh et al. 1999) and the human intestinal ALP gene (Henthorn et al. 1988) were reported to be transcribed into a minimum of two mRNAs of different size. In contrast, we found multiple cDNAs of the same size but with slightly differences in sequence in *H. virescens* midgut. Further research is necessary to determine the number of mALP genes present in *H. virescens*.

When expressed in a heterologous system, clones displaying the highest diversity among the HvmALP clones were detected by antisera cross-reacting with HvALP in immunoblots. Considering the high sequence identity among all the HvmALP clones and the presence of the sequence tag previously obtained for HvALP (GFFLFVENR) (Krishnamoorthy et al. 2007), we propose that all HvmALP isoforms are synonymous with HvALP. This observation may explain the long chain of protein spots detected by antiserum to mALP in 2D blots, with each isoform representing a number of spots. Considering the importance of N-glycosylation for Cry1Ac binding to HvALP (Jurat-Fuentes and Adang 2004), and that the antiserum used to detect HvALP was developed against a non-glycosylated protein, it is possible that although all HvmALP proteins may be detected by the antisera, they may not contain the specific sugars necessary for Cry1Ac binding. In this regard, it is interesting to note that the N-glycosylation sites predicted to exist in HvmALP isoforms were preserved in *H. armigera* ALP sequences, but

did not align to corresponding N-glycosylation sites in other insect ALPs (including *B. mori*). In *B. mori* mALP a single N-glycosylation site was predicted, and localized more proximal to the N terminus of the protein (Itoh, Takeda et al. 1991). Although speculative, this observation may suggest that localization of these N-linked sugars on ALP may be crucial for Cry1Ac binding and may explain higher Cry1Ac susceptibility in *H. virescens* and *H. armigera* when compared to *B. mori* larvae (Van Frankenhuyzen and Nystrom 2002). In contrast, *B. mori* larvae are more susceptible than *H. virescens* or *H. armigera* to other Cry1A toxins whose binding to gut proteins is not inhibited by N-acetylgalactosamine.

Previous reports have demonstrated interactions between Cry1Ac toxin and alkaline phosphatases in Lepidoptera resulting in decreased enzymatic activity (English and Readdy 1989; Sangadala et al. 1994). We have previously reported on decreased HvALP levels in midguts of larvae from Cry1Ac-resistant strains of *H. virescens* (Jurat-Fuentes and Adang 2004; Jurat-Fuentes and Adang 2007). In *Aedes aegypti* larvae, a mALP has been reported as a functional Cry11Aa toxin receptor (Fernandez et al. 2006). Taken together, this information suggests a direct role of HvALP in Cry1Ac intoxication, yet further research is needed to characterize the specific physiological functions of the HvmALP isoforms and their putative role as functional Cry1Ac receptors.

ACKNOWLEDGEMENTS: This research was partially supported by USDA-NRI grant number 2004-35607-14936 to M.J. Adang and J.L-Jurat-Fuentes. The authors express their gratitude to Dr. Brenda Oppert and Dr. Juan Ferre for reviewing this manuscript. Disclaimer: The use of trade, firm, or corporation names in this publication is for the information and

convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

REFERENCES

Aimanova, K. G., Zhuang, M., Gill, S. S., 2006. Expression of Cry1Ac cadherin receptors in insect midgut and cell lines. J. Invertebr. Pathol. **92**: 178-187.

Bradford, M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**: 248-254.

Bravo, A., Gill, S. S., Soberon, M., 2007. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. Toxicon **49**: 423-435.

Carroll, J., Wolfersberger, M. G., Ellar, D. J., 1997. The *Bacillus thuringiensis* Cry1Ac toxin-induced permeability change in *Manduca sexta* midgut brush border membrane vesicles proceeds by more than one mechanism. J. Cell Sci. **110**: 3099-3104.

Chang, W., Zachow, K., Bentley, D., 1993. Expression of epithelial alkaline phosphatase in segmentally iterated bands during grasshopper limb morphogenesis. Development **118**: 651-663.

Chen, J., Brown, M. R., Hua, G., Adang, M. J., 2005. Comparison of the localization of *Bacillus thuringiensis* Cry1A delta-endotoxins and their binding proteins in larval midgut of tobacco hornworm, *Manduca sexta*. Cell Tissue Res. **321**: 123-129.

Eguchi, M., 1995. Alkaline phosphatase isozymes in insects and comparison with mammalian enzyme. Comp.Biochem.Physiol. **111B**: 151-162.

English, L., Readdy, T. L. 1989. Delta endotoxin inhibits a phosphatase in midgut epithelial membranes of *Heliothis virescens*. Insect Biochem. **19**: 145-152.

Fernandez, L. E., Aimanova, K. G., Gill, S. S., Bravo, A., Soberon, M., 2006. A GPI-anchored alkaline phosphatase is a functional midgut receptor of Cry11Aa toxin in *Aedes aegypti* larvae. Biochem. J. **394**: 77-84.

Henthorn, P. S., Raducha, M., Kadesch, T., Weiss, M. J., Harris, H., 1988. Sequence and characterization of the human intestinal alkaline phosphatase gene. J. Biol. Chem. **263**: 12011-12019.

Itoh, M., Inoue, T., Kanamori, Y., Nishida, S., Yamaguchi, M., 2003. Tandem duplication of alkaline phosphatase genes and polymorphism in the intergenic sequence in *Bombyx mori*. Mol. Genet. Genomics **270**: 114-120.

Itoh, M., Kanamori, Y., 1999. Cloning of soluble alkaline phosphatase cDNA and molecular basis of the polymorphic nature in alkaline phosphatase isozymes of *Bombyx mori* midgut. Insect Biochem. Molec. Biol. **29**: 121-129.

Itoh, M., Takeda, S., Yamamoto, H., Izumi, S., Tomino, S., Eguchi, M., 1991. Cloning and sequence analysis of membrane-bound alkaline phosphatase cDNA of the silkworm, *Bombyx Mori*. Biochim. Biophys. Acta, **1129**: 135-138.

Jurat-Fuentes, J. L., Adang, M. J., 2004. Characterization of a Cry1Ac-receptor alkaline phosphatase in susceptible and resistant *Heliothis virescens* larvae. Eur. J. Biochem. **271**: 3127-3135.

Jurat-Fuentes, J. L. Adang, M. J., 2007. A proteomic approach to study resistance to *Bacillus thuringiensis* Cry toxins in *Heliothis virescens* larvae. J. Invertebr. Pathol. **95**: 187-191.

Krishnamoorthy, M., Jurat-Fuentes, J. L., McNall, R. J., Andacht, T., Adang, M., 2007. Identification of novel Cry1Ac binding proteins in midgut membranes from *Heliothis virescens* using proteomic analyses. Insect Biochem. Molec. Biol. **37**:189-201.

Kucera, M., Weiser, J., 1974. Alkaline-Phosphatase in Last Larval Instar of *Barathra-Brassicae* (Lepidoptera) Infected by *Nosema Plodiae*. Acta Entomol. Bohem. **71**: 289-293.

Laemmli, U. K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**: 680-685.

Miao, Y.-G. 2002. Studies on the activity of the alkaline phosphatase in the midgut of infected silkworm, *Bombyx mor*i L. J. Appl. Ent. **126**: 138-142.

Pardo-Lopez, L., Gomez, I., 2006. Structural changes of the Cry1Ac oligomeric pre-pore from *Bacillus thuringiensis* induced by N-acetylgalactosamine facilitates toxin membrane insertion. Biochemistry **45**: 10329-10336.

Perlak, F. J., Oppenhuizen, M., Gustafson, K., Voth, R., Sivasupramaniam, S., Heering, D., Carey, B., Ihrig, R. A., Roberts, J. K., 2001. Development and commercial use of Bollgard cotton in the USA - early promises versus today's reality. Plant J. **27**: 489-501.

Rodrigo-Simon, A, Caccia, S and Ferre, J (2008) *Bacillus thuringiensis* Cry1Ac toxin-binding and pore-forming activity in brush border membrane vesicles prepared from anterior and posterior midgut regions of lepidopteran larvae. *Appl. Environ. Microbiol.* **74**: 1710-1716.

Sangadala, S., Walters, F. S., English, L. H., Adang, M. J., 1994. A mixture of *Manduca sexta* aminopeptidase and phosphatase enhances *Bacillus thuringiensis* insecticidal CryIA(c) toxin binding and $^{86}Rb^{(+)}$-K^+ efflux *in vitro*. J. Biol. Chem. **269**: 10088-10092.

Shaver, T. N., Raulston J. R. 1971. A soybean-wheat germ diet for rearing tobacco budworm. A. Entomol. Soc. Am. 64: 1077-1079.

Sowadski, JM, Handschumacher, MD, Murthy, HM, Foster, BA and Wyckoff, HW (1985) Refined structure of alkaline phosphatase from *Escherichia coli* at 2.8 A resolution. *J Mol Biol* **186**: 417-433.

Sujak, P., Ziemnicki, K., Ziemnicka, J., Lipa, J. J., Obuchowicz, L., 1978. Acid and alkaline-phosphatase activity in fat-body and midgut of beet armyworm, *Spodoptera exigua* (Lepidoptera: Noctuidae), infected with nuclear polyhedrosis virus. J. Invertebr. Pathol. **31**: 4-9.

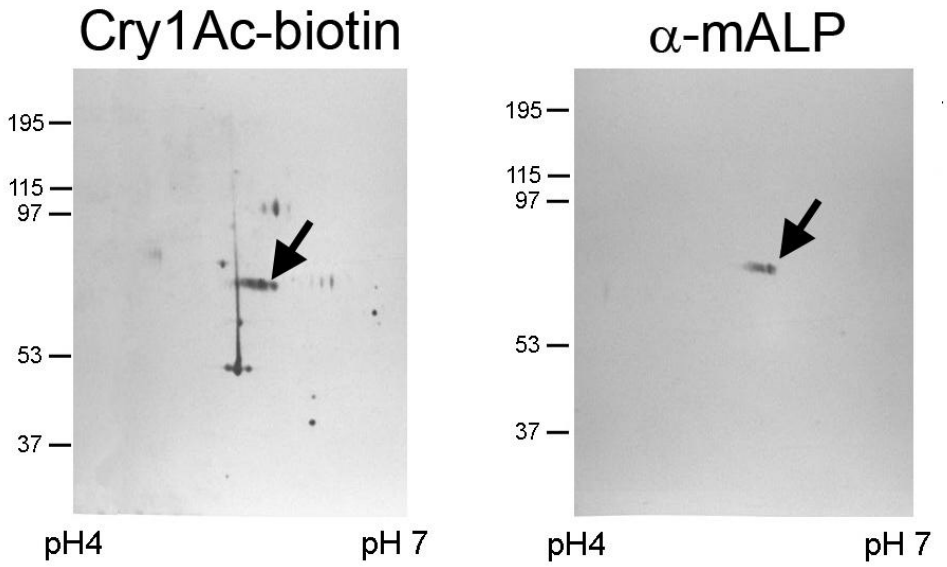
Sukhanova, M. Z., Grenback, L. G., Gruntenko, N. E., Khlebodarova, T. M. Rauschenbach, I. Y., 1996. Alkaline phosphatase in *Drosophila* under heat stress. J. Insect Physiol. **42**: 161-165.

Takesue, Y., Yokota, K., Miyajima, K., Taguchi, R., Ikezawa, H., 1989. Membrane anchors of alkaline phosphatase and trehalase associated with the plasma membrane of larval midgut epithelial cells of the silkworm, *Bombyx mori*. J. Biochem. **105**: 998-1001.

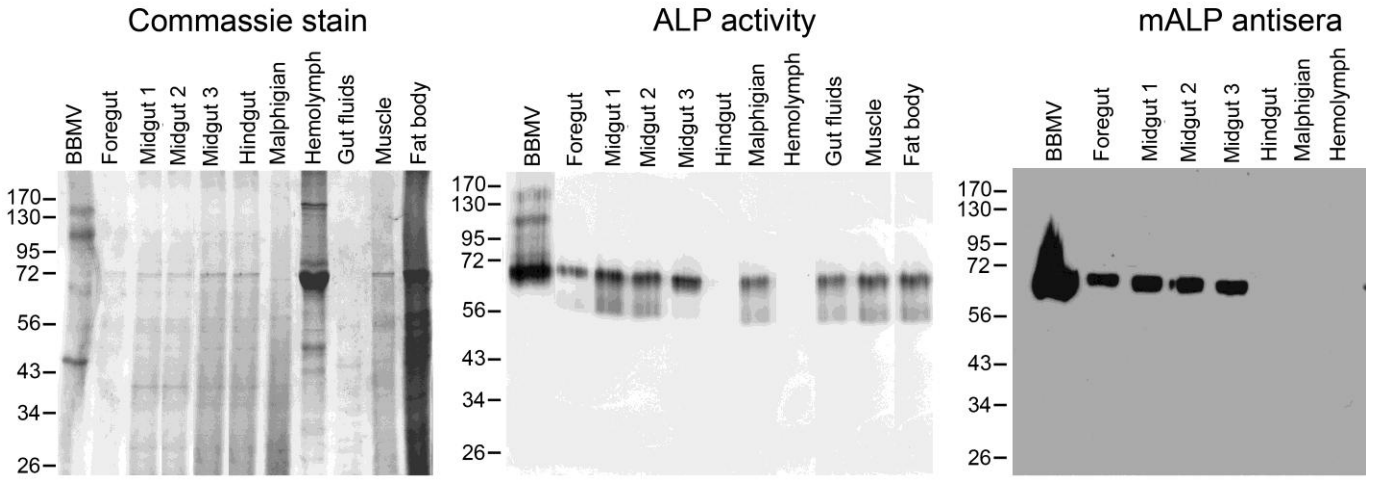
Van Frankenhuyzen, K and Nystrom, C (2002) The *Bacillus thuringiensis* toxin specificity database. *http://www.glfc.cfs.nrcan.gc.ca/bacillus* (September 2008).

Wolfersberger, M. G., Luthy, P., Maurer, A., Parenti, P., Sacchi, V. F., Giordana, B., Hanozet, G. M., 1987. Preparation and partial characterization of amino acid transporting brush border membrane vesicles from the larval midgut of the cabbage butterfly (*Pieris brassicae*). Comp. Biochem. Physiol. **86A**: 301-308.

Zhang, X., Candas, M., Griko, N. B., Taussig, R., Bulla, L. A. Jr., 2006. A mechanism of cell death involving an adenylyl cyclase/PKA signaling pathway is induced by the Cry1Ab toxin of *Bacillus thuringiensis*. Proc. Natl. Acad. Sci. USA **103**: 9897-902.



Appendix 4.-Detection of Cry1Ac binding proteins (left panel) and HvALP (right panel) in blots of 2DE-resolved H. virescens BBMV proteins. Proteins (50 μg) were separated by isoelectrofocusing (IEF) on pH 4-7 Immobiline DryStrips in the first dimension and SDS-8% PAGE gels for the second dimension. 5 nM of biotinylated Cry1Ac and antiserum to *A. gambiae* mALP were used to probe blots. Arrows point to HvALP spots detected by both Cry1Ac and anti-mALP serum. Masses of molecular weight markers (kDA) are indicated at the left of each blot.



Appendix 5.-Detection of ALP activity and HvALP protein in 4th instar H. virescens larval tissues. Tissue protein extracts were prepared as described in materials and methodsExtracts (20 μg) or BBMVproteins (10 μg protein) were separated by SDS-10% PAGE, and stained for total protein with Coomassie Blue, stained for ALP activity with NBT-BCIP, or transferred to PVDF filters and probed with anti-*A. gambiae* mALP serum, as indicated. Samples are as indicated at the top of each lane.

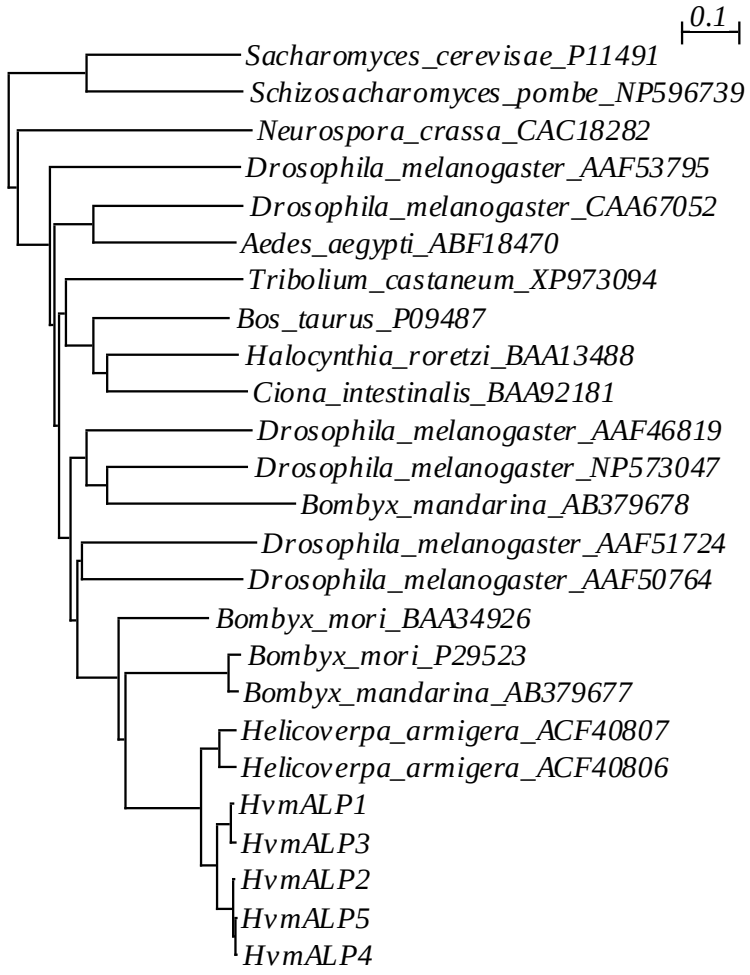
Appendix 6.- Protein sequence alignment of membrane-bound midgut alkaline phosphatases from H. virescens larvaeFigure The presence of complete sequence identity (*), conservative changes (:), and non-conservative changes (.) is denoted below each residue. The signal peptide for expression on the cell surface is highlighted in gray. The predicted phosphatase domain is included in a square, with the predicted active site in bold. Putative N-glycosylation sites are underlined. Predicted GPI-anchor sites are highlighted in black.

```
          1                                                  50
HvmALP1   MMSLYQCLLA VLCCAACARA HWFHPAATAG RAAATTRVET SANYWVQDAQ
HvmALP2   MMSLYQCLLA VLCCAACARA HWFHPAATAG RAAATTRVET SANYWVQDAQ
HvmALP3   MMSLYQCLLA VLCCAACARA HWFHPAATAG RAAATTRVET SANYWVQDAQ
HvmALP4   MMSLYQCLLA VLCCAACARA HWFHPAATAG RAAATTRVET SANYLVQDAQ
HvmALP5   MMSLYQCLLA VLCCAACARA HWFHPAATAG RAAATTRVET SANYWVQDAQ
          ********** ********** ********** ********** ****.*****
          51                                                100
HvmALP1   AAIDARLAQV ESVKKARNVI MFLGDGMSVP TLAAARTLLG QRQGNTGEET
HvmALP2   AAIDARLAQV ESVKKARNVI MFLGDGMSVP TLAAARTLLG QRQGNTGEET
HvmALP3   AAIDARLAQV ESVKKARNVI MFLGDGMSVP TLAAARTLLG QRQGNTGEET
HvmALP4   AAIDARLAQV ESVKKARNVI MFLGDGMSVP TLAAARTLLG QRQGNTGEET
HvmALP5   AAIDARLAQV ESVKKARNVI MFLGDGMSVP TLAAARTLLG QRQGNTGEET
          ********** ********** ********** ********** **********
          101                                               150
HvmALP1   KLHFETFPTI GLVKTYCVDA QIADSACTAT AYLCGVKNNY GAIGVDATVR
HvmALP2   KLHFETFPTI GLVKTYCVDA QIADSACTAT AYLCGVKNNY GAIGVDATVR
HvmALP3   KLHFETFPTI GLVKTYCVDA QIADSACTAT AYLCGVKNNY GAIGVDATVR
HvmALP4   KLHFETFPTI GLVKTYCVDA QIADSACTAT AYLCGVKNNY GAIGVDATVR
HvmALP5   KLHFETFPTI GLVKTYCVDA QIADSACTAT AYLCGVKNNY GAIGVDATVR
          ********** ********** ********** ********** **********
          151                                               200
HvmALP1   RGDCQTASNT ATHVESIAEW ALADGRDVGI VTTTRITHAS PAGTYAKTAN
HvmALP2   RGDCQTASNT ATHVESIAEW ALADGRDVGI VTTTRITHAS PAGTFAKTAN
HvmALP3   RGDCQTASNT ATHVESIAEW ALADGRDVGI VTTTRITHAS PAGTYAKTAN
HvmALP4   RGDCQTASNT ATHVESIAEW ALADGRDVGI VTTTRITHAS PAGTFAKTAN
HvmALP5   RGDCQTASNT ATHVESIAEW ALADGRDVGI VTTTRITHAS PAGTFAKTAN
          ********** ********** ********** ********** ****:*****
          201                                               250
HvmALP1   RTWENDGEVS QMGFDAKDCP DIAHQLVHHH PGNKFKVILG GGRRAFLPNT
HvmALP2   RTWENDGEVS QMGLNAKDCP DIAHQLVHHH PGNKFKVIFG GGKRAFLPNT
HvmALP3   RTWENDGEVS QMGFDAKDCP DIAHQLVHHH PGNKFKVILG GGRRAFLPNT
HvmALP4   RTWENDGEVS QMGLNAKDCP DIAHQLVHHH PGNKFKVIFG GGKRAFLPNT
HvmALP5   RTWENDGEVS QMGLNAKDCP DIAHQLVHHH PGNKFKVIFG GGKRAFLPNT
          ********** ***.****** ********** ********.* **:*******
          251                                               300
HvmALP1   VQDDEGSYGR RIDNRDLIKE WEDDKVARNV SHQYVWNREQ LMSLNDDLPE
HvmALP2   EQDEKGSYGR RLDNRNLIKE WENDKVSRNV SHQYVWNREQ LMSLNDDLPE
HvmALP3   VQDDEGSYGR RIDNRDLIKE WEDDKVARNV SHQYVWNREQ LMSLNDDLPE
HvmALP4   EQDEKGSYGR RLDNRNLIKE WENDKVSRNV SHQYVWNREQ LMSLNDDLPE
HvmALP5   EQDEKGSYGR RLDNRNLIKE WENDKVSRNV SHQYVWNREQ LMSLNDDLPE
          .**:.***** *:***.**** **.***:*** ********** **********
          301                                               350
HvmALP1   YMLGLFGSSH MKYHMKSNPQ SDPTLAELTE VAIRSLRRNE KGFFLFVEGG
HvmALP2   YMLGLFGSSH MTYHMKSDPQ SDPTLAELTE VAIRSLRRNE KGFFLFVEGG
HvmALP3   YMLGLFGSSH MKYHMKSNPQ SDPTLAELTE VAIRSLRRNE KGFFLFVEGG
HvmALP4   YMLGLFESSH MTYHMKSDPQ SEPTLAELTE LAIRSLRRNE KGFFLFVEGG
HvmALP5   YMLGLFESSH MTYHMKSDPQ SEPTLAELTE LAIRSLRRNE KGFFLFVEGG
          ********** *.*****.** *:******** :********* **********
          351                                               400
HvmALP1   RIDHAHHDNL VELALDETLE MDKAVATATQ LLSEDDSLIV VTADHAHVMT
HvmALP2   RIDHAHHDNL VELALDETLE MDKAVATATQ LLSEDDSLIV VTADHAHVMT
HvmALP3   RIDHAHHDNL VELAPDETLE MDKAVATATQ LLSEDDSLIV VTADHAHVMT
```

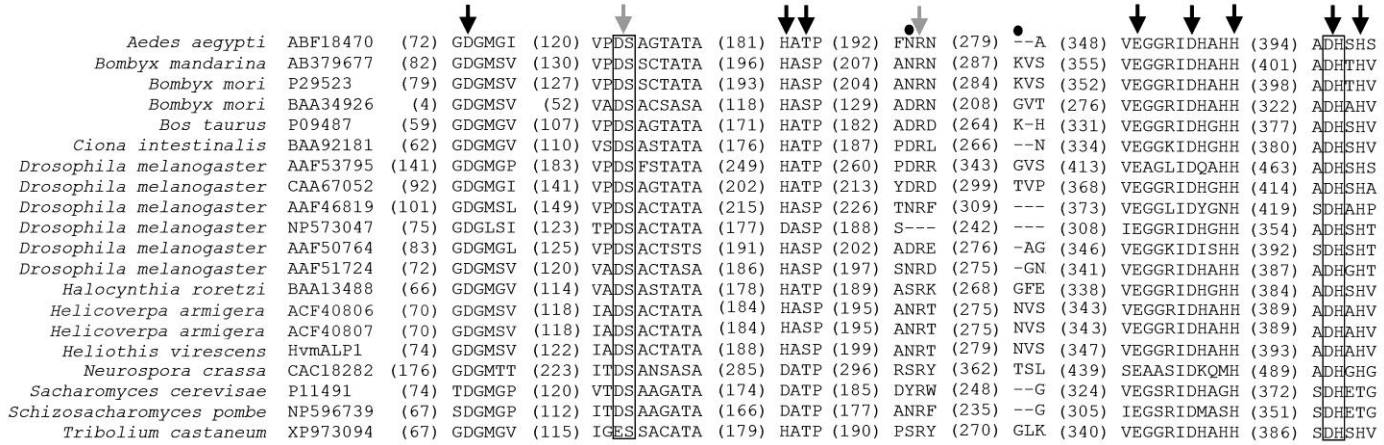
```
HvmALP4  RIDHAHHDNL VELALDETLE MGKAVATATQ LLSEDDSLIV VTADHAHVMT
HvmALP5  RIDHAHHDNL VELALDETLE MDKAVATATQ LLSEDDSLIV VTADHAHVMT
         ********** ****.***** *.******** ********** **********
         401                                                 450
HvmALP1  INGYSGRGND ILGPSRDLGR DSMPYMTLSY TNGPGFRPHV NGIRPDVTAE
HvmALP2  FNGYSNRGRD ILGPSRDVDL DNVPYMTLTY ANGPGFRSHV NDIRPDVTAE
HvmALP3  INGSSGRGND ILGPSRDLGR DSMPYMTLSY TNGPGFRPHV NGIRPDVTAE
HvmALP4  FNGYSNRGRD TLGPSRDVDL DNVPYMTLTY ANGPGFRSHV NDIRPDVTAE
HvmALP5  FNGYSNRGRD ILGPSRDVDL DNVPYMTLTY ANGPGFRSHV NDIRPDVTAE
         .**.*.**.* .******:.. *..*****:* .******.** *.********

         451                                                 500
HvmALP1  PNFRTLDWES HVDVPLEDET HGGDDVAVFA RGPHHSMFTG LYEQSQLPHL
HvmALP2  SNYRSLDWES HVDVPLEDET HGGDDVAVFA RGPHHSMFTG LYEQSQLPHL
HvmALP3  PNFRTLDWES HVDVPLGDET HGGDDVAVFA RGPHHSMFTG LYEQSQLPHL
HvmALP4  SNYRSLDWES HVDVPLEDET HGGDDVAVFA RGPHHSMFTG LYEQSQLPHL
HvmALP5  SNYRSLDWES HVDVPLEDET HGGDDVAVFA RGPRHSMFTG LYEQSQLPHL
         .*:*:***** ******.*** ********** ***.****** *:********
         501                                      539
HvmALP1  MAYAACIGPG RHACVSAAHL PTAHFFIALF ALFTPILLK
HvmALP2  MAYAACIGPG RHACVSAAHL PTAHFFIALF ALFTPILLK
HvmALP3  MAYAACIGPG RHACVSAAHL PSAHFLIALL ALFTSILLR
HvmALP4  MAYAACIGPG RHACVSAAHL PTAHFFIALF ALFTPILLK
HvmALP5  MAYAACIGPG RHACVSAAHL PTAHFFIALF ALFTPILLK
         ********** ********** *:***.***. ****.***:
```

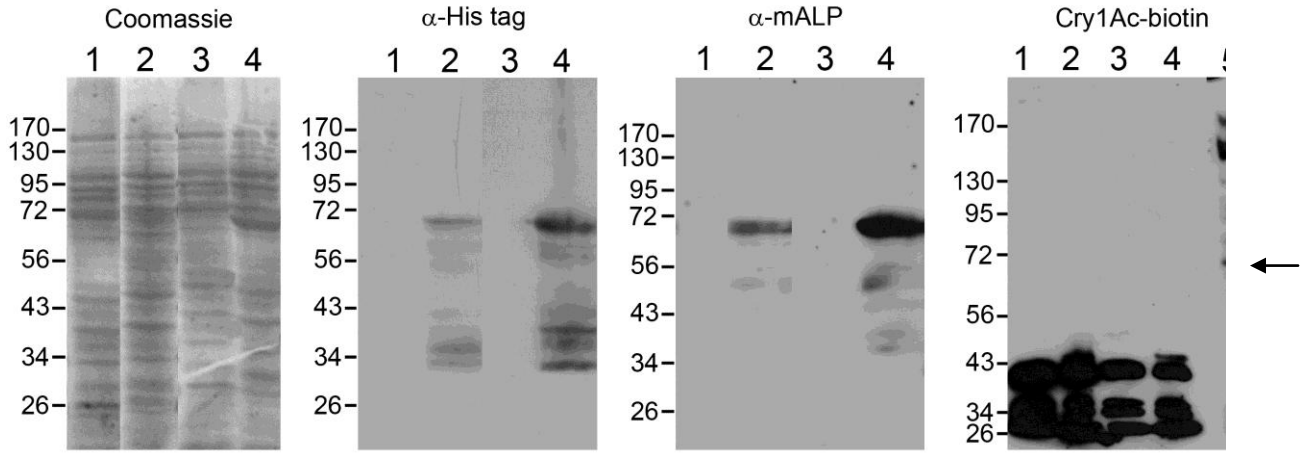
Figure A3 continue



Appendix 7.- Phylogenetic tree obtained by using complete protein sequences of selected and predicted alkaline phosphatases. ClustalX 2.0.9 was used to generate a basic sequence alignment to search for homology among protein sequences. After the general alignment was performed an unrooted Phylip tree was generated with the draw tree function in Clustal X 2.0.9, and then exported into NJPLOT to display the tree.



Appendix 8.- Alignment of partial amino acid sequences from reported and predicted alkaline phosphatases (species and accession numbers as indicated). Specific amino acid residues involved in substrate binding (black arrows), binding to metal ligand (gray arrows), and predicted to be N-glycosylated (black circles) are indicated. Boxes represent an amino acid pair with the same assigned function. The amino acid number in each sequence is indicated in parentheses.



Appendix 9.- Cry1Ac does not interact with HvmALP cDNAs expressed in Escherichia coli. Aliquots of E. coli cultures carrying pET vectors encoding specific full length cDNA sequence were separated by SDS- 10% PAGE. Aliquots of *E. coli* cultures carrying pET vectors encoding specific full length cDNA sequence were separated by SDS- 10% PAGE. After electrophoresis, proteins were stained (Coomassie), or transferred to PVDF filters and probed with antiserum to His tag (α-His tag), antiserum against mALP of *Am. gambiae* (α-mALP), or biotinylated Cry1Ac toxin (Cry1Ac-biotin). Cross-reacting proteins and bound toxin were detected using enhanced chemiluminescence. Lane1: HvmALP2 before induction, lane 2: HvmALP2 2h. after induction, lane 3: HvmALP1 before induction, lane 4: HvmALP1 2h. After induction, lane 5: BBMV proteins (20 μg). Arrow indicates position of HvALP in BBMV lane.

Vita

Jonathan Duran Willis was born August, 23rd 1984 to the parents of James and Pam Willis in Watkinsville, GA. He graduated from Oconee County High School with Honors in 2002. Taking minimal time off between schools he quickly started his education at the University of Georgia during the summer semester. He began working the following fall in Dr. Michael Scanlon's lab in the Plant Biology department as a student worker evaluating plant development genes in maize. After three years he moved to serve as a Student Services Contractor as a lab manager for Dr. Marirosa Molina at the Environmental Protection Agency examining fecal bacteria detection techniques for water systems. Jonathan completed his BS in Applied Biotechnology and moved to Knoxville, Tennessee to begin his graduate career pursuing a MS in Entomology.