Polygalacturonase from Sitophilus oryzae: Possible horizontal transfer of a pectinase gene from fungi to weevils

Authors: Shen, Zhicheng, Denton, Michael, Mutti, Navdeep, Pappan, Kirk, Kanost, Michael R., et al.

Source: Journal of Insect Science, 3(24) : 1-9

Published By: Entomological Society of America

URL: https://doi.org/10.1673/031.003.2401

The BioOne Digital Library (https://bioone.org/) provides worldwide distribution for more than 580 journals and eBooks from BioOne's community of over 150 nonprofit societies, research institutions, and university presses in the biological, ecological, and environmental sciences. The BioOne Digital Library encompasses

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

Shen Z, Denton M, Mutti N, Pappan K, Kanost MR, Reese JC, Reeck GR. 2003. Polygalacturonase from *Sitophilus oryzae*: Possible horizontal transfer of a pectinase gene from fungi to weevils. 9pp. *Journal of Insect Science*, 3:24, Available online: insectscience.org/3.24

Journal of Insect Science

insectscience.org

Polygalacturonase from *Sitophilus oryzae*: Possible horizontal transfer of a pectinase gene from fungi to weevils

Zhicheng Shen,[*1] Michael Denton,[†2] Navdeep Mutti,[*] Kirk Pappan,[†3] Michael R. Kanost,[†] John C. Reese,[*] and Gerald R. Reeck[†]

[*] *Department of Entomology, Kansas State University, Manhattan, Kansas 66506*
[†] *Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506.*
[1] *Current address: Athenix Corporation, 2202 Ellis Rd, Suite B, Durham NC 27703*
[2]*Current address: Northrop Grumman Information Technology, 8111-18th Street, Building 803, Brooks AFB, San Antonio, TX 78235*
[3]*Current address: Department of Pathology, Washington University School of Medicine, St. Louis, MO 63110*
reeck@ksu.edu

Received 14 December 2003, Revised 2 July 2003, Accepted 11 July 2003, Published 12 August 2003

Abstract

Endo-polygalacturonase, one of the group of enzymes known collectively as pectinases, is widely distributed in bacteria, plants and fungi. The enzyme has also been found in several weevil species and a few other insects, such as aphids, but not in *Drosophila melanogaster, Anopheles gambiae,* or *Caenorhabditis elegans* or, as far as is known, in any more primitive animal species. What, then, is the genetic origin of the polygalacturonases in weevils? Since some weevil species harbor symbiotic microorganisms, it has been suggested, reasonably, that the symbionts' genomes of both aphids and weevils, rather than the insects' genomes, could encode polygalacturonase. We report here the cloning of a cDNA that encodes endo-polygalacturonase in the rice weevil, *Sitophilus oryzae* (L.), and investigations based on the cloned cDNA. Our results, which include analysis of genes in antibiotic-treated rice weevils, indicate that the enzyme is, in fact, encoded by the insect genome. Given the apparent absence of the gene in much of the rest of the animal kingdom, it is therefore likely that the rice weevil polygalacturonase gene was incorporated into the weevil's genome by horizontal transfer, possibly from a fungus.

Introduction

Polygalacturonase has been extensively studied in bacteria, plants, and fungi. Well over 200 polygalacturonase and putative polygalacturonase sequences (all homologs) are now in the NCBI-searchable databases. This large family divides neatly into subfamilies, mechanistically and in terms of sequence similarity, containing either exo-polygalacturonases or endo-polygalacturonases (Markovic and Janocek, 2001). Endo-polygalacturonase is not represented in the genomes of *C. elegans, Drosophila melanogaster,* or *Anopheles gambiae* nor in the over 21,000 *Bombyx mori* cDNA sequences presently in the Silkbase EST database (http://www.ab.a.u-tokyo.ac.jp/silkbase).To our knowledge, endo-polygalacturonase activity has never been identified in animal taxa more primitive, evolutionarily, than insects. However an exo-polygalacturonase has been described from a nematode (Jaubert et al. 2002). Among insects, polygalacturonase has been detected in relatively few species (Adams and McAllan, 1956; Laurema and Nuorteva, 1961; Campbell and Dryer 1985). Within the order Coleoptera, strong evidence exists for polygalacturonase in Curculionidae, or weevil family (Campbell 1989; Shen 1996; Doostdar et al. 1997) and in Chrysomelidae (Girard and Jouanin 1999, Reeck et al., unpublished observations) but not in other families.

The question then arises of how an animal species could have acquired an enzyme activity, i.e., endo-polygalacturonase, that is lacking more primitive animal species. One possibility is that the enzyme could be synthesized in symbiotic organisms. Campbell (1989) has raised this possibility as regards the pectinases in *Sitophilus* species, including the rice weevil.

We have previously reported the purification of endo-polygalacturonase from the rice weevil, *Sitophilus oryzae* (Shen et al., 1996) that represented the first purification of this enzyme from any animal source. No attempt was made in that study to identify the encoding genome for the enzyme. Here we report the cloning and sequencing of a cDNA that encodes rice weevil polygalacturonase. Analysis of the cDNA demonstrated that rice weevil polygalacturonase is indeed encoded by the weevil's genome, thus establishing for the first time the existence of an endo-polygalacturonase gene in the animal kingdom. We postulate that the gene was transferred horizontally, perhaps from a fungus, probably before the emergence of the curculionids (weevils). In this regard, the endo-polygalacturonase story in this and perhaps other weevils seems to closely resemble the cellulase genes in the

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

genomes of termites (Watanabe et al., 1998; Tokuda et al.1999).

Materials and Methods

N-terminal amino acid sequencing

We have previously reported a 3200-fold purification of *Sitophilus oryzae* polygalacturonase to apparent homogeneity (Shen et al., 1996). From a sample of that purified polygalacturonase an N-terminal sequence of ATCTVSSYDDVASAIS?CGNINL (where the question mark denotes an unidentified residue) was determined by automated Edman degradation using an Applied Biosystems 473 Protein Sequencing System (home.appliedbiosystems.com) in the Biotechnology Core Facility at Kansas State University.

Total and poly (A)+ RNA isolation

About 5 mg of total RNA was collected from 1 g of live rice weevil adults following the procedure of Chomczynski and Sacchi (1987). Poly (A)+ RNA was isolated by oligo (dT) cellulose column chromatography (Sambrook et al., 1989).

Reverse transcription - polymerase chain reaction (RT-PCR)

Reverse transcription was conducted using rice weevil poly (A)+ RNA and a partially degenerate primer designated as PG4 [5' GACTGACTGANCC(G/A)TGNCC 3']. The primer was based on a GHG motif that is conserved in plant, fungal, and bacterial polygalacturonases (Bussink et al., 1991), where N denotes complete degeneracy in the primer. The first strand of cDNA was synthesized using Moloney Murine Leukemia Virus reverse transcriptase according to the supplier's recommended conditions (Stratagene, www.stratagene.com). This first strand cDNA was used as a template in PCR reactions using PG4 and a primer designated as PG3 [5' GA(T/C)GA(T/C)GTIGCI(A/T)(C/G)NGC 3' where I denotes deoxyinosine] based on the N-terminal amino acid sequence of the rice weevil polygalacturonase. PCR reactions were 100 µl in final volume and contained 2 µl of the first strand cDNA reaction, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgSO_4$, 0.2 mM of each dNTP, 40 pmol of both primers PG3 and PG4, and 1 unit of *Taq* polymerase. Thirty-five cycles of 1 min at 94° C, 1 min at 55° C, and 1.5 min at 72° C were carried out. PCR products were excised from low melting temperature agarose gels, purified using a Wizard miniprep kit and then ligated to pGEM-T vector (both products by Promega). Clones were sequenced using a Sequenase II kit (USB) and T7 or SP6 as primers.

cDNA library construction and screening

An adult rice weevil cDNA library was constructed by directional cloning into the Lambda Zap II vector according to the manufacturer's instructions (Strategene). Around 700,000 initial plaque forming units were obtained before the library was amplified.

Plates with about 20,000 plaques each were prepared according to Sambrook et al. (1989). Plaques were transferred to nitrocellulose membranes, denatured, and then UV-crosslinked to the nitrocellulose. The filters were treated for 4 h at 60° C with hybridization solution (2X SSC, 5X Denhardt's solution, 0.5 % SDS). A partial polygalacturonase cDNA identified by RT-PCR was [^{32}P]-labeled using a Prime-a-Gene kit (Promega, www.promega.com) and used as a probe. The probe was boiled for 5 min and added immediately to the hybridization solution. Filters were hybridized at 60° C for 20 h followed by washing at room temperature with 2X SSC and 0.1 % SDS for 20 min and then 0.1X SSC and 0.1 % SDS for 40 min. The filters were exposed to X-ray film at -70° C with an intensifying screen for 24 h. After autoradiography, PCR screening of putative polygalacturonase clones was performed using the M13 reverse primer and a primer designated as PG40 (5' GCTGCATGTTAAGCTTTCG 3') based on the complement of the RT-PCR polygalacturonase fragment. Promising Lambda ZAP II clones, as identified by PCR screening, were excised and circularized using helper phage according to the manufacturer's instructions (Stratagene). The cDNA clones in pBluescript SK- were sequenced on both strands using T3 and T7 primers as well as primers based on the cDNA sequence. Sequencing was performed by the DNA Sequencing Core Facility, Kansas State University College of Veterinary Medicine, using an ABI 373A DNA Sequencer (home.appliedbiosystems.com).

Rapid amplification of cDNA ends (RACE)

The cDNA clone identified by library screening lacked an initiating methionine, so 5' RACE was performed according to the manufacturer's instructions (GIBCO BRL, www.lifetech.com). Briefly, the first strand of cDNA was synthesized using 0.1 µg of rice weevil poly (A)+ RNA, SuperScript II reverse transcriptase, and PG40 as the extension primer. After transcription the template mRNA was degraded by RNAse H, and the single strand cDNAs were purified with a GlassMAX spin cartridge. The purified cDNAs were poly (dC) tailed by terminal deoxynucleotidyl transferase and then were amplified by PCR using an anchor primer and an internal, gene-specific primer (5' TCGCAGATGCCACATCGTCG 3') complementary to positions 102 to 121 of the cDNA clone. PCR products were cloned into the pGEM-T vector and sequenced as described above.

Genomic DNA isolation and Southern blotting

Legs of approximately 300 adult rice weevils were dissected on ice using micro-scissors. From these we isolated DNA for Southern blotting, given that the legs should be free of symbionts from internal organs. A total of 36 mg of legs were ground in a chilled tissue grinder in homogenization buffer (100 mM Tris-HCl (pH 9.2), 200 mM sucrose, 50 mM EDTA, 0.5 % SDS), and genomic DNA was isolated according to Lis et al. (1983). Genomic DNA was also isolated from whole adult rice weevils according to a method of Bradfield and Wyatt (1983). The genomic DNA from legs and whole insects was digested with various restriction enzymes, separated by agarose gel electrophoresis, and then transferred and cross-linked to a nitrocellulose membrane. The membrane was treated with hybridization solution (25 mM phosphate buffer (pH 7.0), 5X SSC, 5X Denhardt's solution, 50 µg/ml salmon sperm DNA, 50 % formamide) at 37° C for 2 h. The rice weevil polygalacturonase cDNA was [^{32}P]-labeled as described above and, after boiling for 5 min, was added immediately to the hybridization solution. The membrane was hybridized for 15 h at 37° C. Following hybridization the membrane was washed twice for 10 min in 2X SSC and 0.1 % SDS at room temperature and once for 30 min in 0.1X SSC and 0.5 % SDS at 37° C. Finally, the membrane was washed in 0.1X SSC and 0.5 % SDS at 68° C for 30

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

min and then was exposed to X-ray film.

Analysis of several genes in tetracycline-treated rice weevils

One hundred adult rice weevils were reared on 100 g of wheat flour, with or without (as control) tetracycline (0.5 µg/g flour) to kill symbionts (Heddi et al. 1998). Over a period of several weeks, DNA was extracted from groups of five rice weevils, at each of several time points, following the procedure of Bender et al. (1983). PCR primers, shown in Table 1, were used for the amplification of several target genes – the 18S-rRNA gene from *S. oryza*e, the 16S-rRNA genes from *Wolbachia*, the *Sitophilus orzyae* principal endosymbiont, SOPE (Heddi et al. 1998), and the rice weevil polygalacturonase gene. Purified genomic DNA (30 ng) was used as a template for PCR for which we used primers at 0.3 µM and PCR master mix from Promega (www.promega.com) in a final volume of 25 µl. Briefly, PCR conditions were 94^0 C for 5 min, followed by 30 cycles of 94^0 C for 30 sec, the annealing temperature for 45 sec and 72^0 C for 45 sec. Annealing temperatures were 53^0 C for 18S-rRNA, 56^0 C for polygalacturonase and SOPE genes, and 57^0 C for *Wolbachia* (for which 35 cycles were used). PCR products were examined by electrophoresis on 2% agarose gels.

In situ hybridization

A digoxigenin-labeled polygalacturonase RNA probe was synthesized using *Bam* HI-linearized polygalacturonase cDNA and a DIG RNA labeling kit (Boehringer Mannhiem, www.roche.com). Adult rice weevil digestive tracts were dissected on glass slides in ice-cold fixing solution (1X PBS, 0.6 % Triton X-100, 5 % [w/v] paraformaldehyde) and were fixed in the same solution for 1 h at room temperature. Fixing, incubation, and hybridization steps were conducted in a humid chamber to keep the digestive tracts from desiccating. After fixing, the gut tracts were washed 4 times in PBST (1X PBS and 0.6 % Triton X-100) for 10 min at room temperature. The tracts were then treated with hybridization solution (5X SSC, 2 % (w/v) blocking reagent (Boehringer Mannheim), 0.1 % (w/v) N-laurylsarcosine, 0.02 % (w/v) SDS, 50 % formamide) for 1 h at 60° C. The solution was aspirated from the slides and replaced with either hybridization solution alone or containing digoxigenin-labeled RNA probe. A cover slip was placed on the slides and they were hybridized for 18 h at 60° C. The guts were washed 3 times for 30 min with hybridization solution at 55° C followed by 4 washes in PBST for 10 min at room temperature. The gut tracts were next incubated with anti-digoxigenin antibody conjugated to alkaline phosphatase (diluted 1:100 in PBST) at 4° C for 18 h. The tracts were rinsed 4 times with PBST followed by 3 rinses with developing buffer (100 mM Tris-HCl (pH 9.5), 100 mM NaCl, 50 mM $MgCl_2$). The alkaline phosphate color reaction was initiated by adding NBT and X-phosphate according to the manufacturer's instructions (Bio-Rad, www.bio-rad.com) and the slides were allowed to develop in the dark. The color reactions were quenched by rinsing the gut tracts several times with PBST containing 50 mM EDTA. The guts were then viewed and photographed on a compound microscope.

Table 1. Primers used for PCR amplification of several genes.

Gene	Primer sequence	Accession number	PCR product size
Rice weevil 18S- rRNA	F- 5' TACAACGACGACGACGAAC 3'	AF276517	412 bp
	R-5' TAGATACGACCCTCAGCCAG 3'		
PGase	F- 5' TGCGATTTCAAGTTGCACCA 3'	AF207068	164 bp
	R- 5' GTTTTCCCGACACTTGAACCA 3'		
SOPE 16S-rRNA	F- 5' TAATAGCGCCATCGATTGAC 3'	AF005235	568 bp
	R- 5' CCGAAGGCACCAAGGCAT 3'		
Wolbachia 16S-rRNA	F- 5' CCCGAGGCTTAACCTTGGAA 3'	AF035160	213 bp
	R- 5' TCCATCGTTTACAGCGTGGA 3'		

Sequence alignments

We aligned a conserved region, as identified by Bussink et al. (1991), of sequences of polygalacturonases from different kingdoms using ClustalW (1.8). Similarity of various polygalacturonase sequences compared to the rice weevil polygalacturonase sequence were determined using the "BLAST 2 sequences" at the NCBI website (www.ncbi.nlm.nih.gov/BLAST).

Results

Isolation of cDNA clone for rice weevil polygalacturonase

Degenerate primers for PCR were designed based on the N-terminal sequence of rice weevil polygalacturonase and an internal sequence that is highly conserved (Bussink et al., 1991) in fungal, plant, and bacterial polygalacturonases. RT-PCR using these primers produced a band of ~600 bp (data not shown). This band was ligated to pGEM-T, and two clones were selected for partial sequencing from both directions. Both clones were identical and contained a continuous open reading frame that included the deduced amino acid sequence DDVASAISSCTTINL corresponding to the N-terminal sequence of the purified rice weevil polygalacturonase. A partial cDNA was identified using the cloned RT-PCR product as a probe to screen an adult rice weevil cDNA library. The 5'-end of the cDNA sequence was completed by 5'-RACE. Figure 1 shows the 1236-bp nucleotide sequence of the full-length cDNA. An open reading frame stretching from nucleotide 16 to 1110 encodes an inferred amino acid sequence of 365 residues. The underlined sequence observed near the N-terminus of this putative polypeptide corresponds to the N-terminal sequence of the rice weevil polygalacturonase. This suggests that the amino acid sequence encoded by the isolated cDNA clone codes for the endo-polygalacturonase protein previously purified from rice weevil extracts (Shen et al., 1996). The rice weevil polygalacturonase sequence is clearly similar to those of polygalacturonases from fungi, plants, and bacteria. An alignment of a central 120-residue region of high similarity among several polygalacturonases (Bussink et al., 1991) shows that rice weevil polygalacturonase possesses a number of amino acid residues that are strictly conserved across the different kingdoms (Figure 2). We conclude that the isolated cDNA clone encodes the polygalacturonase protein that was previously purified from adult rice weevils.

The N-terminal sequence of the isolated protein is preceded by a presumed signal peptide comprised of the first 17 residues (Figure 1). It seems likely that there is some trimming at the N-terminus after cleavage of the signal peptide since the most likely position of cleavage of the signal peptide (von Heijne, 1986) is several residues before the N-terminus of the enzyme as previously isolated. The crystal structure of a polygalacturonase from the

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

```
AAACAAGTCCTTATAATGAACGCACTGCTTTACAACATCATATTTCTTGTTTCTTTGGCA   60
               M  N  A  L  L  Y  N  I  I  F  L  V  S  L  A     15

TTAGCCAAACCTTCAAACAACTCCTTAGGGGCCACTTGTACAGTAAGTTCATACGACGAT   120
L  A  K  P  S  N  N  S  L  G  A  T  C  T  V  S  S  Y  D  D     35
      ↑

GTGGCATCTGCGATTTCAAGTTGCACCACCATCAACCTGGGCTCTTTTACAGTGCCGGCC   180
V  A  S  A  I  S  S  C  T  T  I  N  L  G  S  F  T  V  P  A     55

GGTACAACTTTAGCTTTAAAATTAAAAACCGGCACCACTGTCAATATCCAAGGTACTATC   240
G  T  T  L  A  L  K  L  K  T  G  T  T  V  N  I  Q  G  T  I     75

AAATTCGGATACCAAGAATGGAAAGGACCTTTGGTTCAAGTGTCGGGAAAACAAGTTACA   300
K  F  G  Y  Q  E  W  K  G  P  L  V  Q  V  S  G  K  Q  V  T     95

ATTAATGGTGCTAGTGGAAAATTTGATGGAGAAGGTGCTAACTACTGGGACAGTCAAGGT   360
I  N  G  A  S  G  K  F  D  G  E  G  A  N  Y  W  D  S  Q  G     115

GACAAAGGAAAAACAAGGCCAAAGTTTATCAAAATATCTACTACATCAAGTTCTACTTTC   420
D  K  G  K  T  R  P  K  F  I  K  I  S  T  T  S  S  S  T  F     135

AATGATATTACTCTGAAAAATTGTCCTCATCAATGCGTCTCTATTAGCAGCGCTTCAGAT   480
N  D  I  T  L  K  N  C  P  H  Q  C  V  S  I  S  S  A  S  D     155

ACTACTTTATCAGGTTGGAACATTGACGTATCAGCAGGAGATAGCGCAGGTGGACACAAC   540
T  T  L  S  G  W  N  I  D  V  S  A  G  D  S  A  G  G  H  N     175

ACTGACGGCTTCGATATCTCCGGCTCGACCGGCATCACCATTAAAAACTCCGTTCCCAAG   600
T  D  G  F  D  I  S  G  S  T  G  I  T  I  K  N  S  V  P  K     195

AACCAAGACGATTGCGTGGCTATTAACCAAGGTTCAAACTTGGTTTTCGAAAGCTTAACA   660
N  Q  D  D  C  V  A  I  N  Q  G  S  N  L  V  F  E  S  L  T     215

TGCAGCGGTGGACACGGGCTCAGTCTCTCAGTAGGACAAAGTACTGAGAATGGGGATGCA   720
C  S  G  G  H  G  L  S  L  S  V  G  Q  S  T  E  N  G  D  A     235

AACACAGTTAAAAACGTCGTTTTTAGCGATTGTACCGTAAAGAGTTCTCGAAATGGCATC   780
N  T  V  K  N  V  V  F  S  D  C  T  V  K  S  S  R  N  G  I     255

CACATAAAGACGCACAAAGATGCCGGTACCGGTGCAATCAGTGATGTGACTTACGAAAAC   840
H  I  K  T  H  K  D  A  G  T  G  A  I  S  D  V  T  Y  E  N     275

ATTAAACTTTCTGGAATTACCAATTATGGCATTAACGTTCAAGAGATTTATGCAAATGGT   900
I  K  L  S  G  I  T  N  Y  G  I  N  V  Q  E  I  Y  A  N  G     295

GGCGATAGTGTTGATCCGTTAGGTAATATACCTACTACAAATTTGAATCTCCAATCGGTG   960
G  D  S  V  D  P  L  G  N  I  P  T  T  N  L  N  L  Q  S  V     315

ACAGGCTCCATGTCTGGAGGATCTAGCAGCATGTCGGTATACATTTTATGCGGCGATGGA   1020
T  G  S  M  S  G  G  S  S  S  M  S  V  Y  I  L  C  G  D  G     335

GGATGCAGTAATTGGGCATGGTCCGGTGTTTCTATTTCCGGCGCGAAAAAAGCAAACAGC   1080
G  C  S  N  W  A  W  S  G  V  S  I  S  G  A  K  K  A  N  S     355

TGTAACTTTACGCCCTGTGGATTCACTTGTTAAACGTTTAAGAATTTTAATAAATAGAAC   1140
C  N  F  T  P  C  G  F  T  C   365

ATAGCTTTAGTAATATATTATATTATGGCCAAATATTTTTTTATACTTATAATGAATAAA   1200
ATATACACTGTAAAAGAAAAAAAGAAAAAAAAAAAA   1236
```

Figure 1. Nucleotide and inferred amino acid sequences of the rice weevil endopolygalacturonase cDNA clone. The underlined region of the amino acid sequence corresponds to the N-terminal sequence from the purified rice weevil polygalacturonase as determined by Edman degradation. The underlined nucleotides near the 3' end correspond to a eukaryotic polyadenylation signal. The arrow after Ala-17 represents a predicted signal peptide cleavage site. The first seventeen nucleotides were determined by 5' RACE. The accession number for this sequence is: AF207068.

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

```
                 •  •                 ••    •                     •
S.oryzae 173 GHNTDGFDISGSTGITIKNSVPKNQDDCVAINQGS------NLVFESLTCSGGHGLSLSV 226
A.niger  182 GHNTDGFDISESTGVYISGATVKNQDDCIAINSGE------SISFTGGTCSGGHGLSIGS 235
S.cerev. 175 GHNTDGFDVGSSSNVLIQGCTVYNQDDCIAVNSGS------TIKFMNNYCYNGHGISVGS 228
maize    204 SPNTDGIHMGDSSGITITNTVIGVGDDCISIGPGTS-----KVNITGVTCGPGHGISIGS 258
tomato   245 SPNTDGVHVSNTQYIQISDTIIGTGDDCISIVSGSQ-----NVQATNITCGPGHGISIGS 299
E.carot. 198 ARNTDGIDPMSSKNITIAYSNIATGDDNVAIKAYKGRAETRNISILHNDFGTGHGMSIGS 257
P.solan. 280 VKNTDGFDPGQSTNVVLAYSYINTGDDHVAVKASSG--PTRNLLFAHNHFYYGHGLSIGS 337
               ****     :  : :        ** :::                    ***:*:
                                                •
S.oryzae 227 GQSTENGDANTVKNVVFSDCTVKSSRNGIHIKTHKDAG-TGAISDVTYENIKLSGITNYG 285
A.niger  236 VGGR---DDNTVKNVTISDSTVSNSANGVRIKTIYKET--GDVSEITYSNIQLSGITDYG 290
S.cerev. 229 VGGR---SDNTVNGFWAENNHVINSDNGLRIKTVEGAT--GTVTNVNFISNKISGIKSYG 283
maize    259 LGRYK--DEKDVTDINVKDCTLKKTMFGVRIKAYEDAASVLTVSKIHYENIKMEDSANP- 315
tomato   300 LGSGN--SEAYVSNVTVNEAKIIGAENGVRIKTWQGGS--GQASNIKFLNVEMQDVKYP- 354
E.carot. 258 ET-MG--VYNVTVDDLKMNG----TTNGLRIKSDKSAA--GVVNGVRYSNVVMKNVAKP- 307
P.solan. 338 ETNTG--VSNMLVTDLTMDGNDSSAGNGLRIKSDASRG--GKVTNIVYDGICMRNVKEP- 392
                                :     :  *::**:            : :    :

S.oryzae 286 INVQEIYA 293
A.niger  291 IVIEQDYE 298
S.cerer. 284 IVIEGDYL 291
maize    316 IFIDMKYC 323
tomato   355 IIIDQNYC 362
E.carot. 308 IVIDTVYE 315
P.solan. 393 LVFDPFYS 400
             :  :  *
```

Figure 2. Multiple alignment of the conserved domain of polygalacturonases from various kingdoms. (*) denotes identical residues, (:)denotes highly conserved residues, and (·) indicates active site residues identified by Pickersgill et al. (1998). Dashes represent gaps that were inserted to preserve the alignments. S.oryzae, *Sitophilus oryzae* (accession number AF207068); A.niger, *Aspergillus niger* (P26213); S.cerev., *Saccharomyces cerevisiae* (P47180); maize, *Zea mays* (CAA40910); tomato, *Lycopersicon esculentum* (M37304); E.carot., *Erwinia carotovora* (3891426); P.solan., *Pseudomonas solanacearum* (M33692)

bacterium *Erwinia carotovora* (Pickersgill et al., 1998) identified Asp-202, Asp-205, Asp-223, and Asp-224 of the *Erwinia* sequence as the likely acid and base donors of the catalytic mechanism. The rice weevil polygalacturonase contains these putative catalytic residues and, overall, it contains six out of seven residues identified in the active site cleft of the *Erwinia* polygalacturonase (Figure 2). Over its full length, the rice weevil polygalacturonase exhibits 45 % positional identity with the polygalacturonase from *Aspergillus niger*, 24 % positional identity with maize polygalacturonase, and 26 % positional identity with the *Erwinia* polygalacturonase. Similarity indexes for the same polygalacturonases are 60 %, 42 %, and 41 %, respectively.

Southern analysis with whole-body and leg DNAs

Southern analysis was conducted using the cloned rice weevil polygalacturonase cDNA as a probe of endonuclease digested DNA from either entire weevils or from legs carefully dissected from weevils (Figure 3). The DNA from legs should be free of the *S. oryzae* principal endosymbiont (SOPE), a prokaryotic organism which is located in specialized cells called mycetocytes, or bacteriocytes, associated with internal organs (Campbell, 1989; Heddi et al., 1998), although DNA from *Wolbachia* may be present in the legs (Heddi et al., 1998). The Southern blot patterns obtained with whole-body and leg DNAs were indistinguishable for each enzyme digestion examined (*Eco*RI, *Hind*III, and *Xba*I) and are consistent with one gene (or locus) for rice weevil polygalacturonase. We conclude that the gene that encodes the cloned polygalacturonase sequence occurs in the rice weevil's legs. This indicates, then, that the gene for rice weevil polygalacturonase is not encoded by the SOPE genome, since one would not expect SOPE to occur in legs.

In situ hybridization with rice weevil polygalacturonase mRNA

In situ hybridization carried out on guts of adult rice weevils revealed that the polygalacturonase transcript occurred in the main portion of the midgut epithelium. Little if any were found in the finger-like caecae that project from the gut (Figure 4). This is of interest since the caecae have been considered the main place of residence of SOPE endosymbionts in the gut (Musgrave and Miller, 1953; Campbell, 1989).

RT-PCR analysis of polygalacturonase gene in antibiotic-treated rice weevils

SOPE, the principal endosymbiont of the rice weevil and the *Wolbachia* symbiont can be eliminated by extended treatment of weevils with tetracycline (Heddi et al., 1998). We used RT-PCR to monitor the presence of genes for 18S-rRNA of *S. oryzae*, rRNA genes of SOPE and *Wolbachia*, and the rice weevil polygalacturonase gene in tetracycline-treated adult rice weevils. As is shown in Figure 5, the symbiont genes disappeared after several weeks of antibiotic treatment, whereas the polygalacturonase gene remained, along with the 18S-rRNA gene of the rice weevil, throughout the tetracycline treatment.

Discussion

The possibility has been previously raised that rice weevil

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

pectinases might be encoded by endosymbiotic bacteria (Campbell, 1989). This was an eminently reasonable hypothesis given that pectinases were widespread in bacteria (and in fungi and in plants) but unknown in animal species other than insects. Furthermore, the hypothesis seemed to correspond to a view that persisted into the 1980s that cellulases in termites and cockroaches are encoded by symbiotic microbes — protozoa, fungi or bacteria.

Just as the view of the genetic origin of termite cellulases has been revised (Slaytor, 1992; Watanabe et al., 1998; Tokuda et al. 1999), we believe that an endosymbiotic origin of weevil pectinases should be rejected. The data we have presented in this paper strongly indicate that the rice weevil polygalacturonase is encoded by the insect's genome. The cDNA has a poly(A)+ tail downstream of an AAUAA polyadenylation signal characteristic of eukaryotic transcripts. This indicates that the transcript is not of prokaryotic origin. The gene is detected by Southern analysis in DNA isolated both from insect legs as well as in DNA isolated the entire insects. Clearly, the gene is not encoded by a symbiont harbored in internal organs. Polygalacturonase activity in the rice weevil can be accounted for entirely by activity in the gut (results not shown) and the transcripts for the enzyme are located in gut as indicated by our *in situ* hybridization results. Within the anterior midgut, the transcripts that encode polygalacturonase occur in the gut epithelium itself and not in the projections, or caecae, that have long been cited as the regions of the gut where endosymbionts are present (Musgrave and Miller, 1953; Campbell, 1989). Heddi et al. (1998) found that SOPE is limited to the larval bacteriome, the ovarian bacteriome, and the female germ cells of *S. oryzae* and that *Wolbachia* is highly abundant in the germ cells, but is present at low density in other tissues such as muscle and cuticle. Sequences similar to polygalacturonase are absent from available *Wolbachia* genome sequences (http://www.tigr.org/tdb/mdb/mdbinprogress.html). Thus, several lines of evidence point toward the presence of the polygalacturonase gene in the weevil genome, rather than in an endosymbiont. Finally, the presence of the polygalacturonase gene in antibiotic-treated rice weevils that have lost SOPE and *Wolbachia* genes clearly indicates that antibiotic-sensitive symbiotic organisms are not the genetic origin of rice weevil polygalacturonase. We conclude that the rice weevil endo-polygalacturonase appears to be a gut digestive enzyme encoded by the insect itself.

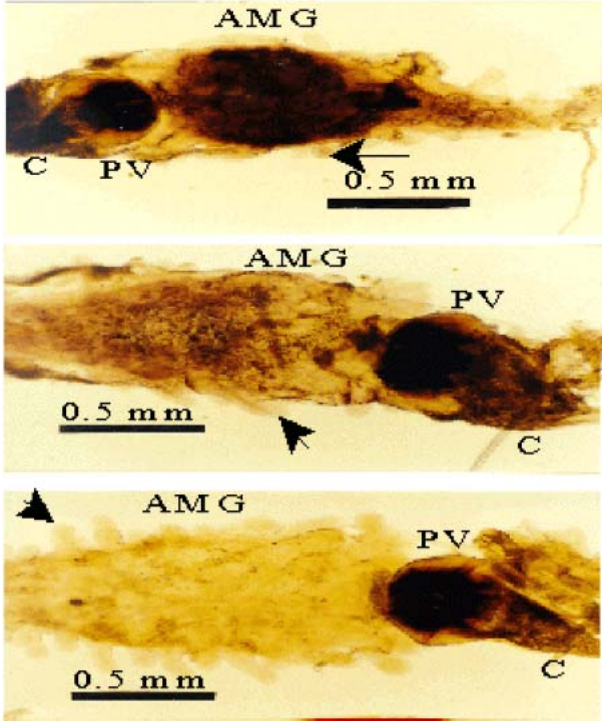

Figure 4. Detection of polygalacturonase mRNA in adult rice weevil digestive tracts by *in situ* hybridization. Paraformaldehyde-fixed digestive tracts were treated with a digoxigenin (DIG)-labeled RNA probe followed by incubation with an anti-DIG antibody (top panel); with hybridization solution lacking the RNA probe followed by incubation with the anti-DIG antibody (middle panel); or with the DIG-labeled RNA probe but not with the anti-DIG antibody (bottom panel). The middle and bottom panels, then, are controls for the top panel. The guts were photographed under a compound microscope using brightfield illumination. Arrows indicate the finger-like caecae. C, crop; PV, periventriculus; AMG, anterior midgut.

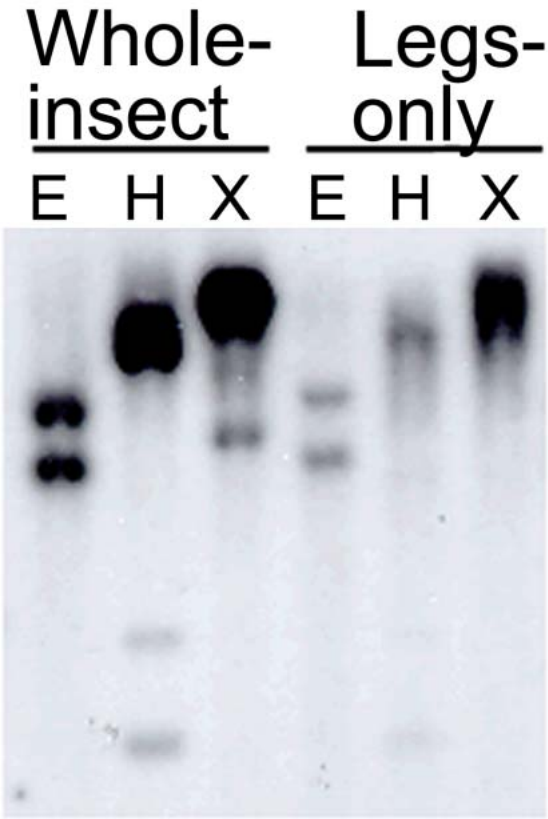

Figure 3. Southern blot analysis of the rice weevil polygalacturonase gene. Genomic DNA was extracted from either whole adult insects (left panels) or legs carefully dissected from insects (right panels) and was then digested with the indicated restriction endonucleases. The cDNA shown in Figure 1 was radiolabeled and used as the probe. E, *Eco* RI; H, *Hind* III; X, *Xba* I

Polygalacturonase enzyme has been isolated from midguts of the sugarcane rootstalk borer weevil, *Diaprepes abbreviatus*, which, like the rice weevil is in the family Curculionidae (Doostdar et al. 1997). Girard and Jouanin (1999) obtained a sequence for a putative polygalacturonase in sequencing randomly selected clones in a gut cDNA library from the mustard beetle (*Phaedon*

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

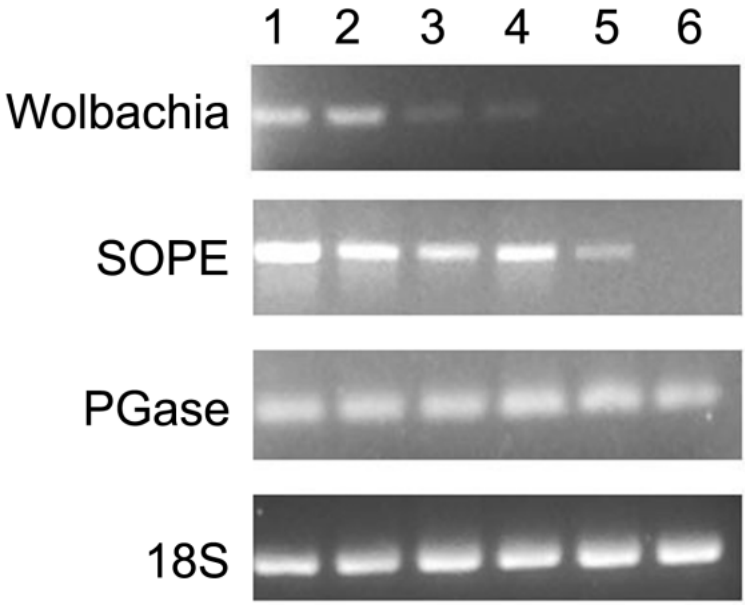

Figure 5. PCR analysis of several genes in tetracycline-treated adult rice weevils. At several times of rearing rice weevils on a diet containing tetracycline (see Materials and Methods), PCR was used to detect 4 genes: the 16S-rRNA genes in *Wolbachia* and in the principal endosymbiont (SOPE), rice weevil polygalacturonase, and the 18S-rRNA gene of *S. oyrzae*. Lane 1 is a control from insects reared on wheat flour not supplemented with tetracycline. Lanes 2, 3, 4, 5 and 6 correspond to samples taken after 2, 4, 5, 6 and 7 weeks, respectively, of feeding on tetracycline-supplemented wheat flour.

cochleariae), which is from another coleopteran family, Chrysomelidae. Neither of these studies examined the genetic origin of the polygalacturonase, but the results do indicate the presence of this enzyme in curculionids and chrysomelids. In this regard, we have found high pectinase activities in guts of the Southern corn rootworm, *Diabrotica undecimpunctata howardi*, another chrysomelid (results not shown).

Thus, a picture is emerging of polygalacturonase as a gut digestive enzyme in curculionids and chrysomelids. Although only a few species have been examined at this point, the results of all the relevant studies suggest that pectinases are found routinely within curculionids and chrysomelids and not in other coleopteran families. Database searching does not reveal endo-polygalacturonase-related sequences in other invertebrates, including three (*C. elegans, D. melanogaster, and An. gambiae*) for which the entire genomes have been sequenced. This suggests, then, that the polygalacturonase gene in the rice weevil (and presumably in other curculionids and in chrysomelids) arose by horizontal transfer, in an ancient species that was a common ancestor to Curculionoidea and Chrysomelidoidea, sister superfamilies within the order Coleoptera (Farrell, 1998; Beutel and Haas, 2000). Since the rice weevil polygalacturonase sequence is most closely related to fungal polygalacturonases, we hypothesize that the donor species was a fungus. A phylogenetic study of a large number of polygalacturonase sequences (Markovic and Janecek, 2001) placed the *Phaedon* polygalacturonase (as the only known animal sequence) within the fungal group, which is also consistent with a fungal origin for the beetle polygalacturonase genes. Acquisition of a polygalacturonase could provide a selective advantage to phytophagous insects, as the ability to digest pectin should increase the availability of nutrients released from leaves or other plant tissues high in pectin.

The possibility of genes having been transferred horizontally from microorganisms to higher organisms has received a considerable stimulus recently as genes found in the human genome have been postulated to be bacterial in origin and not found in other animal species (International Human Genome Sequencing Consortium, 2001; Salzberg et al. 2001). This idea has been challenged, and it has been pointed out that loss of unneeded genes from some taxa may also explain the presence of certain genes in disparate groups of organisms (Stanhope et al., 2001). The invertebrate genomes that have been sequenced so far are from species that do not feed on plants. If, with sequencing of other insect genomes, polygalacturonase genes are identified in species outside of the Curculionidae and Chrysomelidae, it might be concluded that the gene was lost from *C. elegans*, *D. melanogaster*, and *An. gambiae*, due the lack of selection pressure in the face of a non-plant diet. On this point, we would note that we have searched unsuccessfully for polygalacturonase activity and genes in *Tenebrio* and *Tribolium* (Reeck et al., unpublished observations). Phylogenetic analysis, to test whether the degree of similarity of polygalacturonase genes between weevils and fungi is greater than that of other homologous genes that have been diverging since the last common ancestor of insects and fungi, would help to clarify the probability of horizontal gene transfer in this case. However, the polygalacturonase sequence is the only nuclear-encoded *Sitophilus* protein sequence that has a homolog in fungi, and thus such analysis must await sequencing of additional *Sitophilus* genes.

In a sense, the rice weevil polygalacturonase story may parallel the situation with cellulase genes in termites and cockroaches. Classically, these have been thought of as being encoded exclusively by the genomes of symbiotic protozoa. This view was undermined by biochemical studies of Slaytor and his colleagues, who reported and characterized cellulases from symbiont-free insects or organs (for example, Slaytor, 1992). Slaytor's evidence for endogenous celullases in insects and cockroaches has been supported in recent years by investigations at the molecular genetic level (Watanabe et al., 1998; Tokuda et al.1999). One difference from the rice weevil polygalacturonase situation is that there is no evidence now that any genome other than the rice weevil's can encode this enzyme, whereas with at least one species of termite, a symbiotic organism does encode cellulase (Veivers et al., 1983).

Jaubert et al. (2002) have isolated an exo-polygalacturonase (and its encoding cDNA) from a nematode, *Meloidogyne incognita*. The authors propose that the encoding gene is from the animal's genome rather than from a symbiont's genome. The nematode exo-polygalacruatonase strongly resembles bacterial polygalacturonases, whereas the rice weevil's endo-polygalacturonase most strongly resembles fungal polygalacturonases. The nematode and rice weevil enzymes have only a low level of sequence similarity to each other. These data suggest the possibility of two independent horizontal transfers of polygalacturonase genes (one exo-polygalacturonase gene and one endo-polygalacturonase gene) into animal species. We would note that Kondo et al. (2002) have presented strong evidence for the incorporation of a portion of a *Wolbachia* genome

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

into the genome of its host insect, the adzuki bean beetle, *Callosobruchus chinesis*, thus providing a particularly clear-cut case of horizontal gene transfer into a weevil's genome.

Another rice weevil pectinase, pectin methylesterase, is under active investigation in our laboratory. The enzyme itself has been isolated (Shen et al., 1999) and we have cloned its encoding cDNA. Through our studies on these two pectinases we hope to better understand the roles of these enzymes in the rice weevil and, eventually, in other insect species.

Acknowledgements

This work was supported by grants from the National Science Foundation (MCB-9974805), United States Department of Agriculture (95-37302-1837), and the Kansas State University Special Group Incentive Award Program. This work was also supported by the Kansas Agricultural Experiment Station (contribution number 00-188-J). Voucher specimens, No. 052, are deposited in the KSU Museum of Entomological and Prairie Arthropod Research, Manhattan, KS.

References

Adams JB, McAllan JV. 1956. Pectinase in the saliva of *Myzus persicae* (Sulz.) (Homoptera: Aphididae) *Canadia Journal of. Zoology* 34: 541-543.

Bender, W, Spierer, P, Hogness, DS. 1983. Chromosomal walking and jumping to isolate DNA from the ace and rosy loci and the bithorax complex in *Drosophila melanogaster. Journal of Molecular Biology* 168: 17-33.

Beutel RG, Haas F. 2000. Phylogenetic relationships of the suborders of coleoptera (Insecta) *Cladistics* 16: 103-141.

Bradfield JY, Wyatt GR. 1983. X-linkage of a vitellogenin gene in *Locusta migratoria. Chromosoma* 88: 190-193.

Bussink HJD, Buxton FP, Visser J. 1991. Expression and sequence comparison of the *Aspergillus niger* and *Aspergillus tubigensis* genes encoding polygalacturonase II. *Current Genetics* 19: 467-474.

Campbell BC, Dreyer DL. 1985. Host-plant resistance of sorghum: differential hydrolysis of sorghum pectic substances by polysaccharases of greenbug biotypes (*Schizaphis graminum*, Homoptera: Aphididae). *Archives of Insect Biochemistry and Physiology* 2: 203-215.

Campbell BC. 1989. On the role of microbial symbiotes in herbivorous insects. In: *Insect-Plant Interactions* Bernays EA, editor. 1-44. Boca Raton: CRC Press.

Chomczynski P, Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* 162: 156-159.

Doostdar H, McCollum TG, Mayer RT. 1997. Purification and characterization of an endo-polygalacturonase from the gut of West Indies sugarcane rootstalk borer weevil (*Diaprepes abbreviatus* L.) larvae. *Comparative Biochemistry and Physiology* 118B: 861-867.

Farrell BD. 1998. "Inordinate fondness" explained: Why are there so many beetles? *Science* 281: 555-559

Girard C, Jouanin L. 1999. Molecular cloning of cDNAs encoding a range of digestive enzymes from the phytophagous beetle, *Phaedon cochleariae*. *Insect Biochemistry and Molecular Biol*ogy 29: 1129-1142.

Heddi A, Grenier AM., Khatchadourian C, Charles H, Nardon P. 1998. Four intracellular genomes direct weevil biology: Nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. *Proceedings of the National Academy of Sciences (USA)* 96: 6814-6819.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Jaubert S, Laffaire J-B, Abad, P, Rosso M-N. 2002. A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidgyne incognita*. *FEBS Letters* 522: 109-112.

Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T. 2002. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proceedings of the National Academy of Science*. 99: 14280 – 14285.

Laurema S, Nuorteva P. 1961. On the occurrence of pectin polygalacturonase in the salivary glands of Heteroptera and Homoptera Auchenorrhyncha. *Annales Entomologici Fennici.* 27: 89-93.

Lis JT, Simon JA, Sutton, CA. 1983. New heat shock puffs and beta-galactosidase activity resulting from transformation of *Drosophila* with an HSP70-lacZ hybrid gene. *Cell* 35: 403-410.

Markovic O, Janecek S. 2001. Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. *Protein Engineering* 14: 615-631

Musgrave AJ, Miller JJ. 1953. Some microorganisms associated with the weevils *Sitophilus granarius* (L) and *Sitophilus oryza* (L). I. Distribution and description of the organisms. *Canadian. Entomology* 85: 387-390.

Pickersgill R, Smith D, Worboys K, Jenkins J. 1998. Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. carotovora. *Journal of Biological Chemistry* 273: 24660-24664.

Salzberg SL, White O, Peterson J, Eisen JA. 2001. Microbial genes in the human genome: Lateral transfer or gene loss? *Science* 292: 1903 – 1906.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual*. 2nd edition. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Shen ZC, Manning G, Reese JC, Reeck GR. 1999. Pectin methylesterase from the rice weevil, *Sitophilus oryzae* (L.) (Coleoptera: Curculionidae): Purification and characterization. *Insect Biochemistry and Molecular Biology* 29: 209-214.

Shen ZC, Reese JC, Reeck GR. 1996. Purification and characterization of polygalacturonase from the rice weevil, *Sitophilus oryzae* (Coleoptera: Curculionidae). *Insect Biochemistry and Molecular Biology* 26: 427-433.

Slaytor M. 1992. Cellulose digestion in termites and cockroaches: what role do symbionts play? *Comparative Biochemistry*

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use

and Physiology 103B: 775-784.

Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411: 940-941.

Tokuda G, Lo N, Watanabe H, Slaytor M, Matsumoto T, Noda H. 1999. Metazoan cellulase genes from termites: intron/exon structures and sites of expression. *Biochimica et Biophysica Acta* 1447: 146-159.

Veivers PC, O'Brien RW, Slaytor M. 1983. Selective defaunation of *Mastotermes darwiniensis* and its effect on cellulose and starch metabolism in termites. *Insect Biochemistry* 28: 95-101.

von Heijne G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research.* 14: 4683-4690.

Watanabe H, Noda H, Tokuda G, Lo N. 1998. A cellulase gene of termite origin. *Nature* 394: 330-331.

Downloaded From: https://staging.bioone.org/journals/Journal-of-Insect-Science on 22 Mar 2025
Terms of Use: https://staging.bioone.org/terms-of-use