A GENOMIC ANALYSIS OF THE INSECT PEST POPULATIONS OF COWPEA IN WEST AFRICA

BY

TOLULOPE ADEBIMPE AGUNBIADE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Entomology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Barry R. Pittendrigh, Chair, Director of Research
Associate Professor George Czapar
Associate Professor Bettina Francis
Assistant Professor Brian Allan

ABSTRACT

Cowpea [*Vigna unguiculata* (L.) Walp (Fabaceae)] is an important and major staple food crop in sub-Saharan Africa, especially in the dry savanna regions of West Africa. The crop provides food, cash, and fodder. As a food crop, cowpea is a primary source of protein for the ever-growing population of both rural and urban dwellers. The fodder and husks from cowpea also form an important source of protein, fiber, and energy for livestock. West Africa accounts for about 80% of the world's cowpea production. However, insect pests are major constraints to cowpea production in the West African sub-region. The crop is severely attacked at every stage of its growth by different insect pests from the pre-flowering stage right through storage. Damage by insect pests on cowpea can be as high as 80 – 100% if not effectively controlled. Current control measures against the insect pests, which mostly consist of chemical control, are not without limitations.

There is a need to develop a more comprehensive IPM strategy against cowpea insect pests by exploiting the knowledge of their biology, location of alternate host plants, and natural enemies, and combining these with the recent advances in genome sequencing technologies. This dissertation consists of five chapters and focuses on the integration of integrated pest management (IPM) and the current trends in genomic sequencing to cowpea IPM in West Africa with the aim of better understanding the insect pest populations of the cowpea crop and defining their population structure and movement patterns.

Chapter 1 which serves as the introduction to the whole thesis discusses in detail the advent of the genomics era and how IPM researchers must take advantage of the

recent development in genomic practices. It introduces the concept of IPM-omics and how this can be applied to cowpea cropping systems in West Africa. It also discusses the effective deployment of the research output to the end-users.

Chapters 2 and 3 answer questions regarding the timing and spatial scale of the migration patterns of one of the major insect pests of cowpea in West Africa, the legume pod borer (*Maruca vitrata*). I applied a set of microsatellite markers (Chapters 2 and 3) and mitochondrial *cox*1 haplotype data (Chapter 3) to characterize the *M. vitrata* populations across locations in West Africa [Burkina Faso, Niger and Nigeria (Chapter 2)] and also on four host plants of *M. vitrata* [cultivated cowpea (*Vigna unguiculata*), and three alternative host plants - *Pueraria phaseoloides, Loncocarpus sericeus,* and *Tephrosia candida*)] in southern Benin (Chapter 3). The findings from the studies in the two chapters enabled a much clearer understanding of the genetic variability, population structure, and gene flow among *M. vitrata* populations in those countries and the host plants sampled.

Chapter 4 compares the mitochondrial genome of *M. vitrata* from the New World (Puerto Rico) with the mitochondrial genome of the *M. vitrata* population from the Old World (Burkina Faso), and also with the mitogenomes of other Crambids. Species from the genus *Maruca* have a wide distribution from northern Australia and East Asia through sub-Saharan Africa to the Caribbean, Central America, and North America. The species are difficult to distinguish morphologically and have been surmised to be a species complex due to cryptic morphological differences. To be able to study evolutionary patterns among *Maruca* species, I sequenced and assembled the mitochondrial genome of the *Maruca* subspecies from Puerto Rico and compared this with the mitochondrial

genomes of *M. vitrata* from West Africa, and also with other available Crambid mitochondrial genomes. The study enabled the estimation of mutation tendencies in *M. vitrata* and also the construction of phylogenetic relationships, as well as comparative and molecular genome evolution patterns in *M. vitrata*.

Chapter 5 goes beyond my research on *M. vitrata* and concentrates on other destructive insect pests of cowpea in West Africa. In West Africa, besides *M. vitrata*, other serious insect pests also attack the cowpea crop. These insect pests include thrips (*Megalurothrips sjostedti*), aphids (*Aphis craccivora*) and pod sucking bug complex, (including *Clavigralla tomentosicollis* and *Anoplocnemis curvipes*). Collectively, these pests can wipe out a whole cowpea harvest. Part of the constraint in the application of effective control strategies against these pests is the lack of molecular markers that can enable the characterization of the pest populations. For this study, I applied Roche 454 sequencing technology to generate and subsequently assemble contigs from DNA sequencing reads for *A. curvipes*, *A. craccivora*, *C. tomentosicollis* and *M. sjostedti.* These were then used to detect polymorphisms in the different populations of these insect pests across West Africa. Findings from this study identified putative single nucleotide polymorphisms (SNPs), which can be used for characterizing the populations of the different insect pests, and also identified candidate genes putatively involved in insecticide resistance, regulation of insect growth, and response to disease transmission.

Overall, the output from the studies in this dissertation will facilitate the effective evaluation, modification, and optimization of practical cowpea IPM strategies which will in the short- and long-term help in the monitoring of the insect pest populations as well as aid in making decisions as to how, when, and where to apply appropriate control measures.

To the loving memory of my father, Oladeji Adeniran,

To my mother, Omodunbi Adeniran,

To my husband, Olusegun Agunbiade

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Dr. Barry Pittendrigh, for his tremendous academic support and invaluable guidance, which helped in the completion of this dissertation. I also appreciate the input, time, and support of the members of my doctoral committee - Dr. George Czapar, Dr. Bettina Francis, and Dr. Brian Allan.

I am extremely grateful to Dr. Brad Coates for his insightful comments on the research, and for his critical revision of the manuscripts. I am also grateful to Dr. Kyung Seok Kim, Dr. Weilin Sun, Dr. Carmen Valero, Dr. Venu Margam, and our collaborators in Africa - Dr. Manuele Tamò, Dr. Ibrahim Baoua, Dr. Clementine Dabire-Binso, Dr. Malick Ba, and Dr. Rousseau Djouaka, for their helpful discussions on the research, and feedback on manuscripts. For assistance with obtaining insect samples for the studies in this dissertation, I would like to thank our collaborators in Africa. I also appreciate the assistance of students, and technicians that helped with insect sample collection and processing. I would like to thank both past and current members of the Pittendrigh lab. For assistance with paperwork and the administrative aspects of my graduate study, I would like to thank both past (Audra Weinstein and Karen Trame) and current (Adrienne Harris and Kimberly Leigh) departmental office staff.

I am grateful to the Legumes Innovation Lab, formerly Cowpea Collaborative Research Support Program (CRSP), and the Howard Hughes Medical Institute, for providing financial support for my graduate study. Support was provided by the Legumes Innovation Lab by the Bureau for Economic Growth, Agriculture, and Trade, United States Agency for International Development (USAID), under the terms of Grant No.

EDH-A-00-07-00005. Support by the Howard Hughes Medical Institute was provided through a three-year International Student Research Fellowship.

I would like to thank all the people who have provided moral support, friendship, and prayers. Every single act of your kindness and support is appreciated, and you have all made the journey a very memorable one for me.

A very special thanks goes to my family for their love, encouragement, and support. My deepest appreciation goes to my wonderful parents, Oladeji and Omodunbi, for their love and encouragement. I am grateful to both of them for laying a solid foundation for me in education, and for all their sacrifices. I also appreciate my siblings, Adebiyi, Adeola, Adetunji, and Olufunke, for their prayers and invaluable support. My special thanks goes to my dear husband, Olusegun, for his love and patience. I am grateful to him for his sacrifices and his unfailing support throughout this journey. In all, I give thanks to God.

LIST OF PUBLICATIONS ARISING FROM THIS DISSERTATION

Agunbiade T, Steele L, Coates B, Gassmann A, Margam V, Ba M, Dabire-Binso C, Baoua I, Bello-Bravo J, Seufferheld F, Sun W, Tamò M, Pittendrigh B (2012) IPM-omics: From genomics to extension for integrated pest management of cowpea. In Boukar O, Coulibaly O, Fatokun C, Lopez K, Tamò M, (Eds.), *Improving livelihoods in the cowpea value chain through advancements in science.* Proceedings of the 5th World Cowpea Research Conference, pp. 231 – 248 (Chapter 1)

Agunbiade TA, Coates BS, Kim KS, Forgacs D, Margam VM, Murdock LL, Ba MN, Binso-Dabire CL, Baoua I, Ishiyaku MF, Tamò M, Pittendrigh BR (2012) The spatial genetic differentiation of the legume pod borer, *Maruca vitrata* F. (Lepidoptera: Crambidae) populations in West Africa. *Bulletin of Entomological Research* 102(5): 589-599 (Chapter 2)

Agunbiade TA, Coates BS, Datinon B, Djouaka R, Sun W, Tamò M, Pittendrigh BR (2014) Genetic differentiation among *Maruca vitrata* F. (Lepidoptera: Crambidae) populations on cultivated cowpea and wild host plants: Implications for insect resistance management and biological control strategies. *PLoS ONE* 9(3): e92072 (Chapter 3)

Agunbiade TA, Coates BS, Sun W, Tsai M-R, Valero C, Tamò M, Pittendrigh BR (2014) Defining the species complex of the legume pod borer, *Maruca vitrata* (Lepidoptera: Crambidae), by Comparative Mitochondrial Phylogenomics. Submitted (Chapter 4)

Agunbiade TA, Sun W, Coates BS, Djouaka R, Tamò M, Ba MN, Binso-Dabire C, Baoua I, Olds BP, Pittendrigh BR (2013) Development of reference transcriptomes for the major field insect pests of cowpea: A Toolbox for insect pest management approaches in West Africa. *PLoS ONE* 8(11) e79929 (Chapter 5)

LIST OF PUBLICATIONS RELEVANT TO THIS DISSERTATION BUT NOT FORMING PART OF IT

Margam V, Coates BS, Hellmich RL, Agunbiade T, Seufferheld JM, Sun W, Ba MN, Sanon A, Binso-Dabire C, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011) Mitochondrial genome sequence and expression profiling for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(2): e16444.

Margam VM, Coates BS, Bayles DO, Hellmich RL, Agunbiade T, Seufferheld MJ, Sun W, Kroemer JA, Ba MN, Binso-Dabire CL, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011) Transcriptome sequencing, and rapid development and application of SNP markers for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(7): e21388.

Baoua I, Ba NM, Agunbiade TA, Margam V, Binso-Dabire CL, Sanon A, Pittendrigh BR (2011) Potential use of *Sesbania pachycarpa* DC (Fabaceae: Papilionoideae) as a refugia for the legume pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae). *International Journal of Tropical Insect Science* 31: 212-218.

Bello J, Seufferheld F, Steele LD, Agunbiade T, Guillot D, Cutz G, Pittendrigh BR (2011) Scientific Animations Without Borders^{SM}: An international collaborative approach for building applicable scientific educational materials for use on cell phones, and the Internet in developing nations. *The International Journal of Science in Society* 2(4): 49-62.

Bello-Bravo J, Seufferheld F, Agunbiade T, Steele LD, Guillot D, Ba NM, Binso-Dabire CL, Baoua I, N'Diaye M, Tamò M, Pittendrigh B (2012) Scientific Animations Without Borders^{TM}: Cell-phone videos for cowpea farmers. In Boukar O, Coulibaly O, Fatokun C, Lopez K, Tamò M, (Eds.), *Improving livelihoods in the cowpea value chain through advancements in science.* Proceedings of the 5th World Cowpea Research Conference, pp. 369 – 380

Olds BP, Coates BS, Steele LD, Sun W, Agunbiade TA, Yoon KS, Strycharz JP, Lee SH, Paige KN, Clark JM, Pittendrigh BR (2012) Comparison of the transcriptional profiles of head and body lice. *Insect Molecular Biology* 21(2): 257-268.

TABLE OF CONTENTS

CHAPTER 1

IPM-OMICS: FROM GENOMICS TO EXTENSION FOR INTEGRATED PEST MANAGEMENT OF COWPEA[1]

ABSTRACT

Insect pests often develop resistance to insecticides, and such resistance represents a serious management problem. Devising methods that concurrently delay resistance and minimize injury by insects to field crops and stored grain has long been a goal of integrated pest management (IPM). A centerpiece of IPM has been the combined use of biological control agents and prudent application of chemical insecticides. Unfortunately, successful application of IPM has remained a challenge. This chapter describes the use of emerging genomic technologies that may lead to a "systems" perspective of IPM for the control of pests of cowpea and other crops. This emerging field, which we refer to as "IPM-omics", builds upon recent advances in genome sequencing technologies and detection of large-scale gene polymorphisms, which are becoming economically feasible for pest insect systems. IPM-omics will also need to involve the use of information and communications technologies both to collect critical information on pest populations and to deploy practical IPM solutions. The information obtained on the temporal fluctuations, spatial distribution, and ecological diversification within target, non-target, and natural enemy populations can be overlaid on a geographic information systems (GIS) map to

[1] This chapter has been published as Agunbiade et al. 2012. IPM-omics: From genomics to extension for integrated pest management of cowpea. In Boukar O, Coulibaly O, Fatokun C, Lopez K, Tamò M (Eds.), *Improving livelihoods in the cowpea value chain through advancements in science.* Proceedings of the 5th World Cowpea Research Conference, pp. 231 – 248. It is reprinted with the permission of the copyright owner. I acknowledge the contribution of co-authors to the publication.

predict pest outbreaks and to decide how to apply control measures. The "systems" perspective of organism communities provided through IPM-omics may also facilitate the effective evaluation, modification, and optimization of IPM strategies. However, any resultant IPM program for crop pests will also require that extension agents, government agencies, and non-governmental organizations (NGOs) have the ability to easily access and deploy the IPM research findings through information and communications technologies. Thus, we also outline the need for an online system that facilitates the sharing and peer review of practical IPM outputs. Many of these tools are currently being developed to help farmers manage insect pests of cowpea in West Africa.

INTRODUCTION

The Green Revolution, which introduced modern crop varieties and production techniques (Khush 1995; Evenson and Gollin 2003), greatly affected insect pest control. It caused a shift away from the use of crop rotation, field sanitation, flooding, and manual destruction of damaging insects and/or insect-infested plants (Smith et al. 1976) toward the use of high intensity monoculture cropping systems that depended on chemical insecticides to suppress pest insect populations. This shift began in industrialized nations in the 1940s and 1960s, and the consequent reliance on synthetic insecticides led to insecticide resistance, the suppression of beneficial insect populations, and the emergence of minor pests as major pests (Norris et al. 2003). Pesticides can provide excellent crop protection, but in addition to selecting for insecticide resistance, they have had numerous negative impacts on human health and the environment (Singh 2000; Wilson and Tisdell 2001; Georghiou 1986). In 1962, Rachel Carson's book "Silent Spring" raised concerns

and public awareness of the environmental and ecological impacts of pesticide use, and toward the end of the 1970s, this concern and awareness motivated efforts to develop lower impact crop protection methods (Hassan and Bakshi 2005).

The term "integrated pest management" (IPM) was first used by Smith and van den Bosch in 1967 to describe concurrent application of multiple control measures to reduce damage caused by insects to crop plants. The development of IPM led to dramatic changes in the technologies available for pest management. Modern IPM approaches aim to provide economically viable and sustainable control of insect damage by relying upon biological, chemical, physical, host plant resistance, and cultural control tools. Conceptually, IPM strategies recognize that many environmental factors interact and work in concert to affect the abundance of insect pests. IPM strategies also strive to exploit the knowledge of the insect pest's biology, location of alternate plant hosts, and natural enemies to achieve more informed pest management decisions. Fundamentally, the application of an IPM strategy involves a holistic or "systems" approach to pest management, where interactions between pests and the community of natural enemies are evaluated in a tiered approach once the scale of the system in question is set (for more detail, see Wise and Whalon 2009). The systems approach to IPM requires an in-depth knowledge of how management practices affect the insect's biology and other trophic interactions. It also involves the monitoring of pest populations to identify appropriate control measures as well as aids in making decisions as to how, when, and where to apply them.

Analogous to a systems approach for IPM applications, the genomics "revolution" has led to efforts to understand an organism or community on a genome-wide scale.

Recent advances in genomics are now making it possible to better understand pest populations, and are likely to provide new approaches for studying and managing insects. Despite the decreasing costs of DNA sequencing and the development of affordable high-throughput platforms, with which entire genomes can be rapidly sequenced, IPM has not exploited the use of these genomic tools.

Researchers must begin to define how such genomic tools can be used in an inclusive manner with IPM to improve pest management decisions. In addition to the development of genomics, a second parallel "revolution" that has occurred in the past two decades is the ability to deploy pest management concepts through the electronic media. Many of the control strategies that have been developed remain difficult to deploy to extension agents and farmers in Africa and other developing regions because of the remoteness of the regions and the low literacy rates among the farming communities. IPM strategies will have large-scale impacts only if they can be coupled with effective educational deployment tools that can be rapidly updated as new and practical control approaches emerge. We use the term "IPM-omics" to describe the use of genomic (and other "omics") tools to better characterize pest populations and to provide a greater depth of knowledge for the development and deployment of IPM strategies. The remainder of this chapter describes how IPM-omics can be used in the management of cowpea insect pests and how IPM-omics has the potential to further cowpea IPM in West Africa.

Advent of the genomics era

Despite the apparent prominence of molecular biology methods within modern laboratories, it is a relatively young discipline. Molecular biology techniques have made

gene cloning, sequencing, organism transformation, and genotyping of individuals almost commonplace. Many of these techniques became accessible with the development of polymerase chain reaction (PCR), which essentially is *in vitro* synthesis of millions of copies of defined genome regions. The PCR procedure has been credited to Kary Mullis, and the technique has now been adapted for many different applications such that PCR is now a fundamental and indispensable tool of molecular biology (Bartlett and Stirling 2003).

The development of genomic practices depended on the ability to perform high-throughput DNA sequencing at dramatically reduced costs along with an increased capacity to analyze large data sets. The initial sequencing of DNA by chemical modification and subsequent cleavage of specific nucleotides (i.e., Maxam-Gilbert sequencing; Maxam and Gilbert 1977), and by incorporation of di-deoxynucleotides that resulted in truncation of primer extension reactions at specific bases (i.e., Sanger sequencing; Sanger and Coulson 1975; Sanger et al. 1977) was time consuming. It also required expertise to separate DNA fragments (via denaturing polyacrylamide gel electrophoresis) and to interpret the results. Sequencing throughput was increased by the use of capillary gel electrophoresis platforms that automated the separation of up to 384 individual reaction products at a time and by the use of computers to analyze the data in the form of electropherograms. Despite improving the scale at which DNA fragments could be sequenced, these technologies or "platforms" were still not time- or cost-effective for the sequencing of entire genomes. For example, the human genome project, which was launched in 1990, required more than 10 years to complete at a cost of about three billion USD (Venter et al. 2001). These earlier sequencing platforms were also used

to acquire whole genome sequences (WGS) for insects, including the pea aphid, *Acyrthosiphon pisum* Harris (Hemiptera: Aphididae) (The International Aphid Genomics Consortium 2010); the mosquitoes, *Anopheles gambiae* Giles (Diptera: Culicidae) (Holt et al. 2002) and *Aedes aegypti* Linnaeus (Diptera: Culicidae) (Nene et al. 2004); the honey bee, *Apis mellifera* Linnaeus (Hymenoptera: Apidae) (Weinstock et al. 2006); the silkmoth, *Bombyx mori* Linnaeus (Lepidoptera: Bombycidae) (International Silkworm Genome Consortium 2008); the fruit flies, *Drosophila melanogaster* Meigen (Diptera: Drosophilidae) (Adams et al. 2000) and *D. pseudoobscura* Fabricius (Diptera: Drosophilidae) (Richards et al. 2005); the wasp, *Nasonia vitripennis* Ashmead (Hymenoptera: Pteromalidae) (Werren et al. 2010); the red flour beetle, *Tribolium castaneum* Herbst (Coleoptera: Tenebrionidae) (Tribolium Genome Sequencing Consortium 2008); and the human body louse, *Pediculus humanus humanus* Linnaeus (Phthiraptera: Pediculidae) (Pittendrigh et al. 2006; Kirkness et al. 2010; Pittendrigh et al. 2011).

More recently, sequence-by-synthesis technologies that use pyrosequencing reactions to obtain DNA sequence data have been developed (Ronaghi et al. 1998; Morozova and Marra 2008; Simon et al. 2009). Adaptations to high-throughput platforms have the capacity to generate millions of *de novo* DNA base sequences from a single run (Margulies et al. 2005). These next generation sequencing (NGS) platforms have greatly facilitated the collection of DNA sequence data and have been especially useful for genomic research on non-model organisms (McCombie et al. 1992; Ellegren 2009). NGS has contributed to the emergence of insect genomics through the accumulation of WGS data for the wasp species, *Nasonia giraulti* and *N. longicornis* (Werren et al. 2010), and

(in ongoing projects) for *Ixodes scapularis* Say (Ixodida: Ixodidae), *Rhodnius prolixus* Stål (Hemiptera: Reduviidae), *Glossina morsitans* Westwood (Diptera: Glossinidae), *Phlebotomus papatasi* Scopoli (Diptera: Psychodidae), and *Lutzomyia longipalpis* Lutz and Neiva (Diptera: Psychodidae) (Megy et al. 2009). Additionally, NGS has proven useful for the rapid resequencing of wild strains of *B. mori* (Xia et al. 2009) and multiple *Drosophila* species (Hahn et al. 2007), as well as expressed sequence tags (ESTs) (Vera et al. 2008).

EST libraries are collections of short Sanger sequencing- or NGS-derived read data obtained from complementary DNA (cDNA). These sequences are representative of genes that are expressed in a particular tissue, at a specific developmental stage, by a particular phenotype, or under particular treatment conditions under which the libraries were constructed (Adams et al. 1991; Gaines et al. 2002; Nagaraj et al. 2006), and are derived from genomic regions that are actively transcribed. ESTs obviously provide information about gene-encoding regions and are considered as low-cost alternatives to a WGS project (Rudd 2003), even though ESTs are subject to sampling bias. More specifically, cDNAs are sampled from a larger pool of potential templates during the cloning and/or DNA sequencing phases, where errors of omission can occur because of stoichiometric differences between libraries (Liu and Graber 2006). Regardless, ESTs are recognized as an effective research approach for discovering genes (Bonaldo et al. 1996; Audic and Claverie 1997), obtaining gene expression information among tissues (Dimopoulos et al. 2000; Porcel et al. 2000), discovering alternate splicing patterns (Gupta et al. 2004), and predicting novel mutations (Coates et al. 2008). Moreover, the quantitative comparison of ESTs within and between libraries is increasingly used for

gene expression analysis (RNAseq; Simon et al. 2009; Fu et al. 2009). For further information on these developing aspects of genomics research, we direct the reader to the following review articles: Ungerer et al. 2008; Rokas and Abbot 2009; Wheat 2010.

The advent of NGS has also been paralleled by the development of molecular marker screening technologies that offer precision and reliability for differentiating allelic variations at single nucleotide positions. Precision and accuracy of allele calling also has been paired with high-throughput capacity that allows for hundreds to thousands of genotyping calls to be made within 24 h (Lyamichev et al. 1999; Tang et al. 1999; Kwok 2001; Tsuchihashi and Dracopoli 2002). These molecular genotyping assays and associated platforms used for detection rely on initial prediction of segregating mutations within individuals or populations of individuals.

Core concepts underlying IPM-omics

IPM-omics, as we envision it, involves five major steps: (1) discover polymorphisms within insect populations (Figure 1.1), (2) use the polymorphisms to answer critical questions about these pest populations through detailed analysis of sets of individual insects collected from the field (Figure 1.2), (3) overlay this information with other available data sets (e.g., GIS) (Figures 1.3 and 1.4), (4) use the knowledge and outcomes to make better pest management decisions (Figure 1.5), and (5) use information and communication technologies to efficiently extend these materials to the target communities (i.e., those involved in the control of pests associated with the target crop) (Bello-Bravo et al. 2012). These steps are now possible because of the following current conditions. The advent of NGS allows one to sequence genes from a large pool of insects

from a given region in order to discover polymorphisms within these populations. The polymorphisms can then be used in high-throughput polymorphism detection systems to investigate details of pest populations. This information on its own or in combination with GIS systems should give us the ability to gain insights into pest populations that have not been previously possible, allowing for better pest management decisions. Finally, the advent of the Internet and cell phones can be used to communicate information on pest populations between farmers and researchers, and these same communication technologies can be used to provide farmers with practical pest management recommendations.

Application of genomic tools to the current pest situation in cowpea

Cowpea, *Vigna unguiculata* (L.) Walp. (Leguminosae: Papilionaceae), is an important grain legume crop in the semi-arid and dry savannah areas of the tropics (Singh and van Emden 1979). Insect pests feed on and damage cowpea at virtually every crop developmental stage and also feed on and damage the stored grain. The pest species that seriously damage cowpea are numerous (see Table 1.1) and in addition to causing direct feeding damage to vegetative and reproductive tissues, *Aphis craccivora* Koch (Hemiptera: Aphididae) also vectors plant viruses that further decrease crop yields (Singh and van Emden 1979). Although all these insects cause serious damage to cowpea, the pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae) is probably the most destructive (Jackai and Singh 1988).

The control of cowpea pests has relied largely on chemical insecticides but the efficacy of this method is variable due to the evolution of resistance to multiple classes of

insecticides (Ekesi 1999). In addition, the insecticides are prohibitively expensive or otherwise unavailable to low-income farmers in West Africa (Giga and Biscoe 1989; Alghali 1991; Bottenberg 1995; Alebeek 1996). The effective control of insect pests of cowpea has therefore relied upon conventional methods (see Introduction), which have been enhanced through the introduction of IPM strategies (Jackai and Adalla 1997). IPM-omics is currently being used to identify possible genetic variations in populations of pest species and their related natural enemies, and also to study their geographic distributions and movement patterns. These data are critical for targeting and timing appropriate control measures in the field as well as for determining the coincidence of endogenous natural enemy populations.

Although the genomes of a number of insect pests have been sequenced, there is little or no DNA sequence information for most of the cowpea pests. This scenario, however, is quickly changing. Recently, Margam et al. (2011a) used mitochondrial DNA sequence data and molecular genetic markers to determine that *M. vitrata* from West Africa and Puerto Rico represent two distinct groups and even proposed that the samples are possibly a complex of sibling species. Moreover, these data indicated that *M. vitrata* collected from West Africa (Burkina Faso, Niger, and Nigeria), Taiwan, and Australia likely constitute a single species. Previous field studies in West Africa also indicated that *M. vitrata* is more abundant throughout the year in the southern region (Sudanian zone) of the cowpea producing areas, where the rainfall is relatively high, than in the drier northern regions (the Sahelian zone) (Ba et al. 2009; Margam et al. 2010). Furthermore, light trap studies conducted in Burkina Faso (Ba et al. 2009) strongly support the hypothesis that *M. vitrata* adults migrate to the northern drier regions from the south during the rainy season.

To study these movement patterns in greater detail, researchers are developing additional molecular genetic markers from genomic sequence data. EST sequence data were collected from the larval *M. vitrata* midgut and salivary gland tissues as well as from whole adult tissues (Margam et al. 2011b; National Center for Biotechnology Information, NCBI, dbEST accessions HS097571-HS099476). The assembly of these EST data into 3729 contiguous or overlapping contig sequences has also been used to predict the location of ~ 1078 putative single nucleotide polymorphisms (SNPs). Preliminary genotyping data have been collected from Burkina Faso, Niger, and Nigeria at 70 SNP loci, and resulting analyses indicated genetic differences among populations. More specifically, individuals from three sample locations in Burkina Faso (in an East–West pattern of collection) were more genetically different than individuals collected from either Niger or Nigeria (collected in a North–South pattern) (Pittendrigh, unpublished data). These results also suggested that *M. vitrata* populations are isolated by distance in that sample sites most geographically separated show the most genetic divergence. The current and ongoing application of IPM-omics within West Africa has advanced the knowledge of the population structure for pest insects and will likely be used next to investigate the contribution of alternate host plants to the maintenance of population variability. Although cowpea pest dynamics are incompletely understood, the rapid advances in IPM-omics are increasing our understanding. This information should result in better decisions about which pest controls to deploy, and when and where to deploy them.

Several parasitoids have been identified for the biological control of most cowpea insect pests. Although some of these control agents are indigenous, whether they are or

can become sufficiently abundant to help control the pest populations is unclear. Non-native species from Taiwan also have been identified as potential biological control agents, and one example is *Apanteles taragamae* Viereck (Hymenoptera: Braconidae), which has already been experimentally released in Benin (Tamò et al. 2012). To verify the success of parasitoid releases, researchers could potentially study the parasitoid and pest populations to determine whether the pests are being controlled by the indigenous or exotic parasitoids. Molecular markers with polymorphisms unique to the introduced populations can potentially be used to determine whether parasitoids identified from the field (post-release) are in fact those that were introduced by the biological control program.

Integrating IPM with genomics: IPM-omics

In the last 60 years, the technologies available for pest management have changed dramatically. The previously described genomic tools provide IPM with new ways to understand and develop insect pest control strategies. We are now entering a time when molecular biology, or more precisely the new field of genomics, has the potential to be integrated into IPM to allow practitioners to make better decisions about pest control. The application of molecular genomics to the study of insect population ecology (i.e., ecological genomics) will become increasingly important as insect ecologists discover the power of these new tools (Sunnucks 2000). IPM can increase the durability of chemical and transgenic pest management tools by decreasing the selective pressure on pest populations to evolve resistance. Moreover, resistance management may be enhanced further if other management tools impose fitness costs on insecticide-resistant pests (Pittendrigh et al. 2008; Gassmann et al. 2009a). Fitness costs of resistance arise in the

absence of a control agent when those individuals with resistance alleles have lower fitness than homozygous susceptible individuals. Recent research on insects with resistance to insecticidal toxins produced by the bacterium *Bacillus thuringiensis* (*Bt*) indicates that ecological factors can magnify the fitness costs of *Bt* resistance (Gassmann et al. 2009a). This is an example of ecological negative cross-resistance (Pittendrigh et al. 2008). IPM agents that also increase fitness costs of *Bt* resistance include entomopathogenic viruses (Raymond et al. 2007; Sarfarz et al. 2010), entomopathogenic nematodes (Gassmann et al. 2006; Gassmann et al. 2008; Gassmann et al. 2009b; Hannon et al. 2010), entomopathogenic fungi, and host-plant resistance (Raymond et al 2005; Bird and Akhurst 2007). Computer modeling indicates that IPM agents that magnify fitness costs of resistance can delay or prevent the evolution of insecticide resistance within pest populations (Carrière and Tabashnik 2001; Pittendrigh et al. 2004; Gassmann et al. 2008; Gassmann et al. 2009b). IPM-omics could lead to more efficient and broader application of ecological negative cross-resistance (Gassmann et al. 2009c). By understanding the genetic basis of resistance, it may be possible to predict which IPM practices will elicit the largest fitness costs, thereby maintaining pest susceptibility and preserving the utility of newly developed pest management tools such as *Bt* crops and biopesticides.

The use of genomics for understanding pest populations can be broadly placed into two categories, functional genomics and population genomics (the later being a neologism of the term population genetics). Functional genomics (and other related analysis tools such as proteomics, metabolomics, and systems biology analyses) can be and have been used to understand the detailed mechanisms of how insects respond to and evolve specific responses to challenges that they experience in their environment, such as host-plant

resistance factors and pesticides (Pedra et al. 2004; Li et al. 2009). Unfortunately, functional "omics" tools are currently resource intensive and are only used in the laboratory. The use of genomics for practical applications like the development of new pesticides or the identification of new host plant-resistance factors is predicted to occur in the future (Grimmelikhuijzen et al. 2007). Additionally, genes known to be critical for insect survival may ultimately be target sites for the expression of RNA interference (RNAi) constructs by transgenic plants. However, the use of transgenic plants expressing RNAi specific to insect target genes for the control of pest insects is still in its infancy, and questions concerning resistance management for this control strategy remain to be resolved. Although little is known about potential RNAi resistance in insects, multiple resistance mechanisms to dietary RNAi have already been identified in *C. elegans* (Winston et al. 2007). It follows that practical insect control measures that may emerge out of functional genomics will probably not occur in the short term.

One of the other challenges for functional "omics" is the general need for a complete genome sequencing, annotation, assembly of the genome, and categorization of the gene classes observed. For many insect species, large genomes filled with a high number of repetitive elements may in the short term make the sequencing of these genomes prohibitively expensive, especially for crop pests, for which resources may be limited. Even when resources are sufficient, however, genomes with a large number of repetitive elements may be difficult to reassemble. An additional problem is that an increased availability of genomes will also reduce the availability of individuals able to perform manual assembly and annotation.

With the advent of high-throughput sequencing, population genomics tools have the potential to be useful in insect control in the short- and medium-term. Knowledge can be rapidly gained about polymorphisms in insect populations, and this information can be used to understand insect population structure and dynamics. Briefly, a large number of insects can be collected from a target region to optimize the amount of polymorphisms observed in a given EST sequencing run. Those polymorphisms identified in the initial sequencing runs can then be used to answer detailed questions about those insect populations. One example within the context of biological control is to use such molecular markers, in conjunction with traditional pest population sampling, to help determine the source populations of the pests during the seasons when the target crop is not grown. By determining the source of host plants or regions where the pest populations originate (e.g., testing the hypothesis that *M. vitrata* migrates from the south to the north during the rainy season), pest managers can decide where to release biological control agents.

Need for effective information sharing systems of polymorphism data and cost-effective long-term extension strategies

Finally, a mechanism will be needed to facilitate the flow of information in both directions between those scientists who develop IPM-omics strategies and those end users who apply the strategies. The scientists will need such information sharing systems to obtain the large amounts of polymorphic data required to continually and efficiently build on previous efforts. Extension agents will also require a centralized system for information exchange so that they can access information about practical pest control strategies.

Geographic information systems (GIS)

The use of geo-statistics, GIS, satellite and photo aerial images, geographical positioning systems (GPS), and cellular technologies in combination with the latest advancements in molecular genetic analysis have dramatically changed the level of resolution applied in insect ecology and IPM. The spatial sciences combined with genetic studies of molecular markers such as polymorphisms are now being used to understand the major factors affecting the distribution of insects and to understand the mechanisms of dispersion and migration. These tools are also being used to study the spatial and temporal flow of genetic material within populations and to determine how biotic and abiotic factors along with selection-adaptation strategies may contribute to the genetic structure of an insect population (Lushai and Loxdale 2004).

The use of genomics data requires the development of databases that integrate information on genetic polymorphisms in insect pest populations with geo-referenced information such as geographic coordinates of collected specimens, the time of collection, and environmental conditions of the location. The information collected (or generated in the laboratory) is analyzed using geo-statistical methods and processed into a digital platform or GIS. The information can be displayed in multiple layers as synthetic maps where correlations are visually expressed (Manel et al. 2003; Ribes-Dasi et al. 2005). The revolutionary application of these technologies allows researchers to identify the insects that are being tracked and to track insect displacements at different geographic scales while enhancing the genetic resolution such that the genetics of an individual insect within a population could be characterized (Lushai and Loxdale 2004). From a methods perspective, the use of geo-statistical analysis and GIS programs may simplify the

collection, transmission, and handling of large databases. Through the analysis and manipulation of digital data, it is possible to generate scientific predictions based on a number of possible scenarios. A series of thematic maps are usually elaborated. These maps are very powerful application tools because they show values and correlations as visual variables, and this makes them easy to understand by technical personnel, decision makers, and participating producers (e.g., farmers).

If the current trends continue, our ability to access many polymorphisms across many loci and many insect samples will require a data repository that can then be linked with geographic and ecological data sets. These combined databases will allow researchers to answer important questions about insect pest populations and their environment such as: Does the population migrate? Is the population adapted to particular hosts? Is the population resistant to particular insecticides? With the answers to these and similar questions, researchers, government agencies, and farmers can make better decisions regarding human and environmental protection through the application of more rational pest management strategies.

Extension systems

Ultimately, the success of any IPM strategy will hinge on the development, adaptation, and rapid deployment of educational materials that inform scientists and farmers about cost-effective pest control strategies. These materials will be developed and used by various stakeholders including scientists, development organizations (e.g., NGOs and Peace Corps volunteers), extension services, farmer organizations, and farmers themselves. For example, live-action videos have been created that describe the best

practices for rearing and deploying biological control agents. The sharing of these videos and similar educational materials through the Internet will help institutions in host countries develop effective biological control release programs (Bello-Bravo et al. 2012). Additionally, as innovations in these programs arise, a centralized Internet-based system will be needed so that groups can share these materials for peer review to assure the quality and easy access of the content.

A centralized Internet-based system is also needed so that practical pest control strategies can be deployed directly to NGOs, extension agents, farmer organizations, and farmers. To this end, an online peer-reviewed system (termed the Sustainable Development Virtual Knowledge Interface) is currently in development to provide for an online platform for the sharing of extension materials in an easily and widely accessible manner (Bello-Bravo et al. 2010). For example, educational videos containing useful information for cowpea farmers and designed for transmission and viewing with cell phones have already been produced. The language of these videos can be easily changed to match the language of the users. In addition, this centralized Internet-based cell phone system can be used for deploying text, Powerpoint®, audio, and PDF files. Thus, older extension materials can be added to the centralized system and deployed by cell phone to the community of cowpea farmers and extension agents.

CONCLUSIONS

Cost-effective educational strategies that produce tangible and useful educational materials will be critical for the long-term sustainability of any IPM program. Such materials will provide various stakeholders with the capacity to easily deploy such

information into their communities. It will also allow the end users to provide feedback on the educational materials. Thus, IPM-omics programs will need to integrate not only genomics (to increase the understanding of pest populations) but also GIS and other current tools that facilitate the efficient collection, analysis, and exchange of IPM information.

REFERENCES

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, WoodageT, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185-2195.

Adams MD, Kelley JM, Gocayne JD, Dubmick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* 252: 1651- 1656.

Agunbiade TA, Kim KS, Forgacs D, Margam VM, Murdock LL, Ba NM, Binso-Dabire CL, Baoua I, Ishiyaku MF, Tamò, Pittendrigh BR (2012) The spatial genetic differentiation of the legume pod borer, *Maruca vitrata* F. (Lepidoptera: Crambidae) populations in West Africa. *Bulletin of Entomological Research* 102(5): 589-599.

Agunbiade TA, Sun W, Coates BS, Djouaka R, Tamò M, Ba MN, Binso-Dabire C, Baoua I, Olds BP, Pittendrigh BR (2013) Development of reference transcriptomes for

the major field insect pests of cowpea: A Toolbox for insect pest management approaches in West Africa. *PLoS ONE* 8(11) e79929.

Agunbiade TA, Coates BS, Datinon B, Djouaka R, Sun W, Tamò M, Pittendrigh BR (2014) Genetic differentiation among *Maruca vitrata* F. (Lepidoptera: Crambidae) populations on cultivated cowpea and wild host plants: Implications for insect resistance management and biological control strategies. *PLoS ONE* 9(3): e92072.

Alebeek F (1996) Natural suppression of bruchid pests in stored cowpea, *Vigna unguiculata* (L.) Walp., in West Africa. *International Journal of Pest Management* 42: 55 – 60.

Alghali AM (1991) Studies on cowpea farming practices in Nigeria with emphasis on insect pest control. *Tropical Pest Management* 37: 71-74.

Audic S, Claverie J (1997) The significance of digital gene expression profiles. *Genome Research* 7: 986-995.

Ba NM, Margam VM, Binso-Dabire CL, Sanon A, McNeil JN, Murdock LL, Pittendrigh BR (2009) Seasonal and regional distribution of the cowpea pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae) in Burkina Faso. *International Journal of Tropical Insect Science* 29(3): 109-113.

Bartlett JMS, Stirling D (2003) PCR Protocols, 2nd Edition. Humana Press, Totowa, New Jersey.

Bello-Bravo J, Diaz R, Venugopal S, Viswanathan M, Pittendrigh BR (2010) Expanding the impact of practical scientific concepts for low-literate learners through an inclusive and participatory virtual knowledge ecosystem. *Journal of the World Universities Forum* 3:147–164.

Bird LJ, Akhurst RJ (2007) Effects of host plant species on fitness costs of *Bt* resistance in *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Biological Control* 40: 196-203.

Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: Two approaches to facilitate gene discovery. *Genome Research* 6: 791-806.

Bottenberg H (1995) Farmers' perceptions of crop pests and pest control practices in rainfed cowpea cropping systems in Kano, Nigeria. *International Journal of Pest Management* 41: 195-200.

Carrière Y, Tabashnik BE (2001) Reversing insect adaptation to transgenic insecticidal plants. *Proceedings of the Royal Society: Biological Sciences Series B* 268: 1475-1480.

Chi YH, Ahn JE, Yun DJ, Lee SY, Liu TX, Keyan ZS (2011) Changes in oxygen and carbon dioxide environment alter gene expression of cowpea bruchids. *Journal of Insect Physiology* 57(1): 220-230.

Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2008) Mining an *Ostrinia nubilalis* midgut expressed sequence tag (EST) library for candidate genes and single nucleotide polymorphisms (SNPs). *Insect Molecular Biology* 17: 607–620.

Dimopoulos G, Casavant TL, Chang S, Scheetz T, Roberts C, Donohue M, Schultz J, Benes V, Bork P, Ansorge W, Soares MB, Kafatos FC (2000) *Anopheles gambiae* pilot gene discovery project: Identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. *Proceedings of the National Academy of Sciences USA* 97: 6619-6624.

Ekesi S (1999) Variability of pathogenic activity of entomopathogenous fungi (Hyphomycetes) towards the legume flower thrips, *Megalurothrips sjostedti* (Trybom) (Thysanoptera: Thripidae) and their potentials for biological control. PhD Thesis. Ahmadu Bello University, Zaria.

Ellegren H (2009) Sequencing goes 454 and takes large–scale genomics into the wild. *Molecular Ecology* 17: 1629–1635.

Evenson RE, Gollin D (2003) Assessing the impact of the green revolution, 1960 – 2000. *Science* 300 (5620): 758-762.

Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, Menzel C, Chen W, Li Y, Zeng R, Khaitovich P (2009) Estimating accuracy of RNA-seq and microarrays with proteomics. *BMC Genomics* 10: 161.

Gaines PJ, Brandt KS, Eisele AM, Wagner WP, Bozic CM, Wisnewski N (2002) Analysis of expressed sequence tags from subtracted and unsubtracted *Ctenocephalides felis* hindgut and Malpighian tubule cDNA libraries. *Insect Molecular Biology* 11: 299–306.

Gassmann AJ, Stock SP, Carrière Y, Tabashnik BE (2006) Effect of entomopathogenic nematodes on the fitness cost of resistance to *Bt* toxin Cry1Ac in pink bollworm (Lepidoptera: Gelechiidae). *Journal of Economic Entomology* 99: 920-926.

Gassmann AJ, Stock SP, Sisterson MS, Carrière Y, Tabashnik BE (2008) Synergism between entomopathogenic nematodes and *Bacillus thuringiensis* crops: Integrating biological control and resistance management. *Journal of Applied Ecology* 45: 957-966.

Gassmann AJ, Carrière Y, Tabashnik BE (2009a) Fitness costs of insect resistance to *Bacillus thuringiensis*. *Annual Review of Entomology* 54: 147-163.

Gassmann AJ, Fabrick JA, Sisterson MS, Hannon ER, Stock SP, Carrière Y, Tabashnik BE (2009b) Effects of pink bollworm resistance to *Bacillus thuringiensis* on phenoloxidase activity and susceptibility to entomopathogenic nematodes. *Journal of Economic Entomology* 102: 1224-1232.

Gassmann AJ, Onstad DW, Pittendrigh BR (2009c) Evolutionary analysis of herbivorous insects in natural and agricultural environments. *Pest Management Science* 65: 1174-1181.

Georghiou GP (1986) The magnitude of the resistance problem. In *Pesticide resistance: Strategies and tactics for management*. National Academy Press, Washington D.C. pp14-43.

Giga DP, Biscoe J (1989) Treating maize grains for storage with registered protectants in Zimbabwe: Technical, practical and economic considerations. *Zimbabwe Science News* 23: 101-103.

Grimmelikhuijzen CJP, Cazzamali G, Williamson M, Hauser F (2007) The promise of insect genomics. *Pest Management Science* 63: 413–416.

Gupta S, Zink D, Korn B, Vingron M, Haas SA (2004) Genome wide identification and classification of alternative splicing based on EST data. *Bioinformatics* 20(16): 2579-2585.

Hahn MW, Han MV, Han S-G (2007) Gene family evolution across 12 *Drosophila* Genomes. *PLoS Genetics* 3(11): e197.

Hannon ER, Sisterson MS, Stock SP, Carrière Y, Tabashnik BE, Gassmann AJ (2010) Effects of four nematode species on fitness costs of pink bollworm resistance to *Bacillus thuringiensis* toxin Cry1Ac. *Journal of Economic Entomology* 103: 1821-1831.

Hassan ASMR, Bakshi K (2005) Pest management, productivity and environment: A comparative study of IPM and conventional farmers of northern districts of Bangladesh. *Pakistan Journal of Social Sciences* 3(8): 1007-1014.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC,

Weissenbach J, Kafatos FC, Collins FH, Hoffman SL (2002) The genome sequence of the Malaria mosquito *Anopheles gambiae*. *Science* 298: 129-149.

Jackai LEN, Singh SR (1988) Screening techniques for host plant resistance to insect pests of cowpea. *Tropical Grain Legume Bulletin* 35: 2-18.

Jackai LEN, Adalla CB (1997) Pest management practices in cowpea: A review. In Singh BB, Mohan Raj DR, Dashiell KE, Jackai LEN (Eds.), *Advances in Cowpea Research.* Co-publication of International Institute of Agriculture (IITA) and Japan International Center for Agricultural Sciences (JIRCAS). IITA, Ibadan, Nigeria, pp 240-258.

Khush GS (1995) Breaking the yield frontier of rice. *GeoJournal* 35: 329–332.

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JM, Ribeiro JM, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li HM, Meyer JM, Johnson RM, Romero-Severson J, Vanzee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao R, Barker SC, Cameron S, Bruggner RV, Regier A, Johnson J, Viswanathan L, Utterback TR, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR. (2010) Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences USA* 107: 12168-12173.

Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. *Annual Review of Genomics and Human Genetics* 2: 235–258.

Li CH, Yeakley JM, McDaniel TK, Shen R (2009) Medium- to high-throughput genotyping using VeraCode microbeads. *Methods in Molecular Biology* 496: 129-142.

Liu D, Graber JH (2006) Quantitative comparison of EST libraries requires compensation for systematic biases in cDNA generation. *BMC Bioinformatics* 7:77.

Lushai G, Loxdale HD (2004) Tracking movement in small insect pests, with special reference to aphid populations. *International Journal of Pest Management* 50(4): 307-315.

Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA, Neri BP, Brow MA (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology* 17: 292–296.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18(4): 189-197.

Margam VM, Baoua I, Ba, NM, Ishiyaku MF, Huesing JE, Pittendrigh BR, Murdock LL (2010) Wild host plants of legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae) in southern Niger and northern Nigeria. *International Journal of Tropical Insect Sceince* 30(2): 108-114.

Margam VM, Coates BS, Ba MN, Sun W, Binso-Dabire CL, Baoua I, Ishiyaku MF, Shukle JT, Hellmich RL, Covas FG, Ramasamy S, Armstrong J, Pittendrigh BR, Murdock LL (2011a) Geographic distribution of phylogenetically-distinct legume pod borer, *Maruca vitrata* (Lepidoptera: Pyraloidea: Crambidae). *Molecular Biology Reports* 38(2): 893 – 903.

Margam VM, Coates BS, Hellmich RL, Agunbiade T, Seufferheld JM, Sun W, Ba MN, Sanon A, Binso-Dabire C, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011b) Mitochondrial genome sequence and expression profiling for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(2): e16444.

Margam VM, Coates BS, Bayles DO, Hellmich RL, Agunbiade T, Seufferheld MJ, Sun W, Kroemer JA, Ba MN, Binso-Dabire CL, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011c) Transcriptome sequencing, and rapid development and application of SNP markers for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(7): e21388.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 326–327.

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences USA* 74 (2): 560–564.

McCombie WR, Adams MD, Kelley JM, FitzGerald MG, Utterback TR, Khan M, Dubnick M, Kerlavage AR, Venter JC, Fields C (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologies. *Nature Genetics* 1: 124–131.

Megy K, Hammond M, Lawson D, Bruggner RV, Birney E, Collins FH (2009) Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase. *Infection, Genetics and Evolution* 9: 308-313.

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255–264.

Nagaraj SH, Gasser RB, Ranganathan S (2006) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics* 8: 6–21.

Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, Ren Q, Zdobnov EM, Lobo NF, Campbell KS, Brown SE, Bonaldo MF, Zhu J, Sinkins SP, Hogenkamp DG, Amedeo P, Arensburger P, Atkinson PW, Bidwell S, Biedler J, Birney E, Bruggner RV, Costas J, Coy MR, Crabtree J, Crawford M, Debruyn B, Decaprio D, Eiglmeier K, Eisenstadt E, El-Dorry H, Gelbart WM, Gomes SL, Hammond M, Hannick LI, Hogan JR, Holmes MH, Jaffe D, Johnston JS, Kennedy RC, Koo H, Kravitz S, Kriventseva EV, Kulp D, Labutti K, Lee E, Li S, Lovin DD, Mao C, Mauceli E, Menck CF, Miller JR, Montgomery P, Mori A, Nascimento AL, Naveira HF, Nusbaum C, O'leary S, Orvis J, Pertea M, Quesneville H, Reidenbach KR, Rogers YH, Roth CW, Schneider JR, Schatz M, Shumway M, Stanke M, Stinson EO, Tubio JM, Vanzee JP, Verjovski-Almeida S, Werner D, White O, Wyder S, Zeng Q, Zhao Q, Zhao Y, Hill CA, Raikhel AS, Soares MB, Knudson DL, Lee NH, Galagan J, Salzberg SL, Paulsen IT, Dimopoulos G, Collins FH, Birren B, Fraser-Liggett CM, Severson DW (2004) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316: 1718-1723.

Norris RF, Caswell-Chen EP, Kogan M (2003) Concept in integrated pest management. Prentice-Hall of India Private Ltd, New Delhi. *Journal of Invertebrate Pathology* 105: 204-206.

Pedra JH, McIntyre LM, Scharf ME, Pittendrigh BR (2004) Genome-wide transcription profile of field- and laboratory-selected DDT-resistant *Drosophila*. *Proceedings of the National Academy of Sciences USA* 101(18): 7034-7039.

Pittendrigh BR, Gaffney PJ, Huesing JE, Onstad DW, Roush RT, Murdock LL (2004) "Active" refuges can inhibit the evolution of resistance in insects towards transgenic insect-resistant plants. *Journal of Theoretical Biology* 231: 461-474.

Pittendrigh BR, Clark JM, Johnston JS, Lee SH, Romero-Severson J, Dasch GA (2006) Proposed sequencing of a new target genome: The human body louse, *Pediculus humanus humanus*. *Journal of Medical Entomology* 43(6): 1103-1111.

Pittendrigh BR, Huesing JE, Sun L, Gaffeny P (2008) Negative cross-resistance: Past, present, and future potential. In Onstad DW [Ed.], *Resistance Management: Biology, Economics, and Prediction*. Academic Press, Burlington, pp. 107-124.

Pittendrigh BR, Berenbaum MR, Seufferheld MJ, Margam VM, Strycharz JP, Yoon KS, Sun W, Reenan R, Lee SH, Clark JM (2011) Simplify, simplify: Lifestyle and compact genome of the body louse provide a unique functional genomics opportunity. *Communicative and Integrative Biology* 4(2): 188-191.

Porcel BM, Tran AN, Tammi M, Nyarady Z, Rydâker M, Urmenyi TP, Rondinelli E, Peterrsson U, Andersson B, Aslund L (2000) Gene survey of the pathogenic protozoan *Trypanosoma cruzi*. *Genome Research* 10: 1103-1107.

Raymond B, Sayyed AH, Wright DJ (2005) Genes and environment interact to determine the fitness costs of resistance to *Bacillus thuringiensis*. *Proceedings of the Royal Society: Biological Sciences Series B* 272: 1519-1524.

Raymond B, Sayyed AH, Hails RS, Wright DJ (2007) Exploiting pathogens and their impact on fitness costs to manage the evolution of resistance to *Bacillus thuringiensis*. *Journal of Applied Ecology* 44: 768-780.

Ribes-Dasi M, Sió J, Planas S, Almacellas J, Torà R, Avilla J (2005) The use of Geostatistics and GIS to optimize pest control practices in precision farming systems. Information and Technology for Sustainable Fruit and Vegetable Production. FRUTIC 05, 12 -16 September 2005, Montpellier France.

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, Couronne O, Hua S, Smith MA, Zhang P, Liu J, Bussemaker HJ, van Batenburg MF, Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ, Ortiz-Barrientos D, Rives CM, Metzker ML, Muzny DM, Scott G, Steffen D, Wheeler DA, Worley KC, Havlak P, Durbin KJ, Egan A, Gill R, Hume J, Morgan MB, Miner G, Hamilton C, Huang Y, Waldron L, Verduzco D, Clerc-Blankenburg KP, Dubchak I, Noor MA, Anderson W, White KP, Clark AG, Schaeffer SW, Gelbart W, Weinstock GM, Gibbs RA (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and *cis*-element evolution. *Genome Research* 15: 1-18.

Rokas A, Abbot P (2009) Harnessing genomics for evolutionary insights. *Trends in Ecology and Evolution* 24(4):192-200.

Ronaghi M, Uhlén M, Nyrén P (1998) DNA sequencing: A sequencing method based on real-time pyrophosphate. *Science* 281 (5375): 363-365.

Rudd S (2003) Expressed sequence tags: Alternative or complement to whole genome sequences? *Trends in Plant Science* 8: 321–329.

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 94 (3): 441–448.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* 74 (12): 5463–5467.

Sarfarz RM, Cervantes V, Myers JH (2010) Resistance to *Bacillus thuringiensis* in the cabbage looper (*Tricoplusia ni*) increases susceptibility to nucleopolyhedrovirus. *Journal of Invertebrate Pathology* 105(2): 204-206.

Simon SA, Zhai J, Nandety RS, McCormick KP, Zeng J, Mejia D, Meyers BC (2009) Short-read sequencing technologies for transcriptional analyses. *Annual Review of Plant Biology* 60: 305-333.

Singh RB (2000) Environmental consequence of agricultural development: A case study from the Green Revolution state of Haryana, India. *Agriculture, Ecosystems and Environment* 82: 97-103.

Singh SR, van Emden HF (1979) Insect pests of grain legumes. *Annual Review of Entomology* 24: 255–278.

Smith RF, van den Bosch R (1967) Integrated control. In Kilgore WW, Doutt RL (Eds.), *Pest Control: Biological, Physical, and Selected Chemical Methods*. Academic Press, New York.

Smith RF, Apple JL, Bottrell DG (1976) The origins of integrated pest management concepts in agricultural crops. In Apple JL, Smith RF (Eds.), *Integrated Pest Management*. Plenum Press, New York, pp 1–16.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Evolution and Ecology* 15:199–203.

Tang K, Fu D-J, Julien D, Braun A, Cantor CR, Köster H (1999) Chip-based genotyping by mass spectrometry. *Proceedings of the National Academy of Sciences USA* 96:10016–10020.

The International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biology* 8: e1000313.

The International Silkworm Genome Consortium (2008) The genome of a Lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochemistry and Molecular Biology* 38: 1036-1045.

Tribolium Genome Sequencing Consortium (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452: 949-955.

Tsuchihashi Z, Dracopoli NC (2002) Progress in high throughput SNP genotyping methods. *Pharmacogenomics Journal* 2: 103–110.

Ungerer MC, Johnson LC, Herman MA (2008) Ecological genomics: Understanding gene and genome function in the natural environment. *Heredity*100 (2):178-83.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M,

Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. *Science* 291: 1304-1351.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH (2008) Rapid transcriptome characterization for a non–model organism using 454 pyrosequencing. *Molecular Ecology* 17: 1636–1647.

Weinstock GM, Robinson GE, Gibbs RA, Worley KC, Evans JD, Maleszka R, The Honeybee genome sequencing consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443: 931-949.

Werren JH, Richards S, Desjardins CA, Niehuis O, Jürgen G, Colbourne JK, The Nasonia Genome Working Group (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327: 343-348.

Wheat CW (2010) Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica* 138(4): 433-51.

Wilson C, Tisdell C (2001) Why farmers continue to use pesticides despite environmental, health and sustainability costs. *Ecological Economics* 39: 449-462.

Winston MW, Sutherlin M, Wright AJ, Feinberg EH, Hunter CP (2007) *Caenorhabditis elegans* SID-2 is required for environmental RNA interference. *Proceedings of the National Academy of Sciences USA*104: 10565–10570.

Wise J, Whalon M (2009) A systems approach to IPM integration, ecological assessment and resistance management in tree fruit orchards. In Ishaaya I, Horowitz AR (Eds.), *Biorational Control of Arthropod Pests: Application and Resistance Management.* Springer Publishing Ltd., Dordrecht, Heidelberg, London, New York, pp. 325-345.

Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, Cheng T, Jiang T, Becquet C, Xu X, Liu C, Zha X, Fan W, Lin Y, Shen Y, Jiang L, Jensen J, Hellmann I, Tang S, Zhao P, Xu H, Yu C, Zhang G, Li J, Cao J, Liu S, He N, Zhou Y, Liu H, Zhao J, Ye C, Du Z, Pan G, Zhao A, Shao H, Zeng W, Wu P, Li C, Pan M, Li J, Yin X, Li D, Wang J, Zheng H, Wang W, Zhang X, Li S, Yang H, Lu C, Nielsen R, Zhou Z, Wang J, Xiang Z, Wang J (2009) Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science* 326: 433-436.

FIGURES AND TABLES

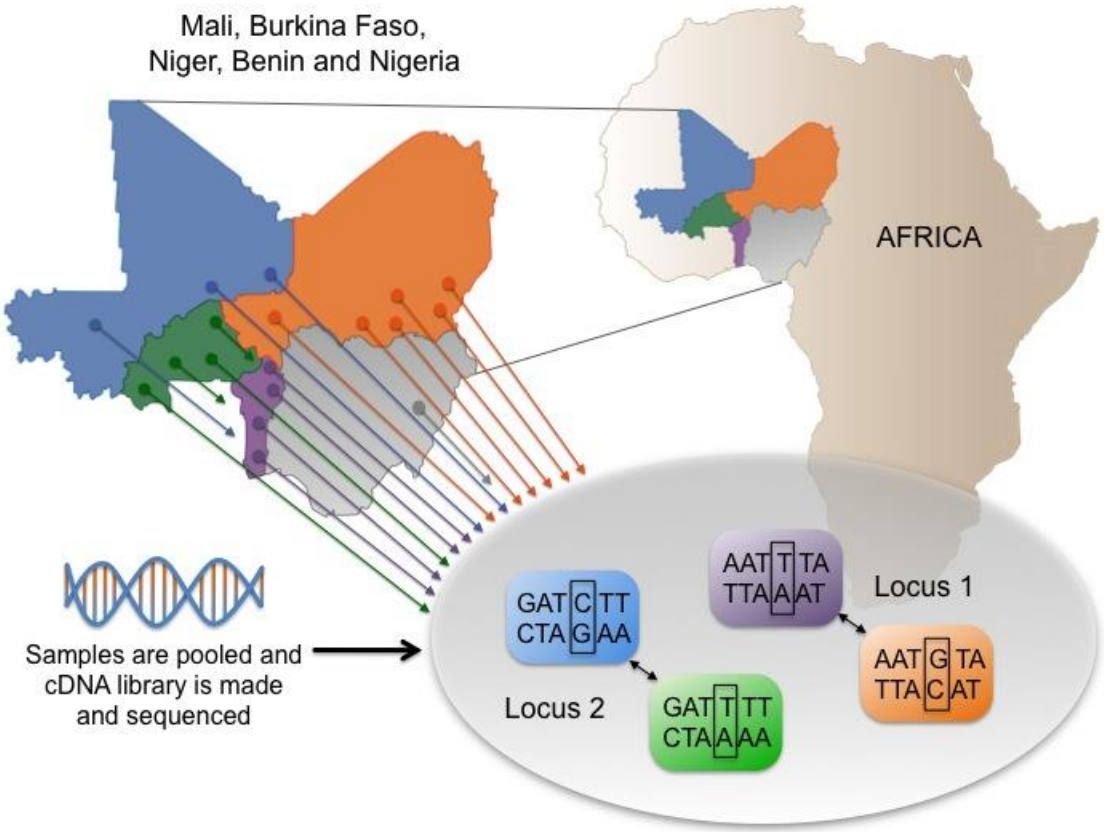


Figure 1.1. Hypothetical representation of the pooling of insect samples from different cowpea growing areas of West Africa for the construction of cDNA library (the arrows do not represent exact current sampling locations for any given insect species). The insects are taken from a diverse set of geographical locations, and the mRNA is pooled to create a cDNA library. Alignment of subsequent DNA sequence forms a set of expressed sequence tags (ESTs) potentially derived from the same locus, and thus representative of alleles at a locus. The presence of polymorphisms among alleles in the form of single nucleotide polymorphisms (SNPs) can readily be adapted for genotyping individuals with high-throughput systems.

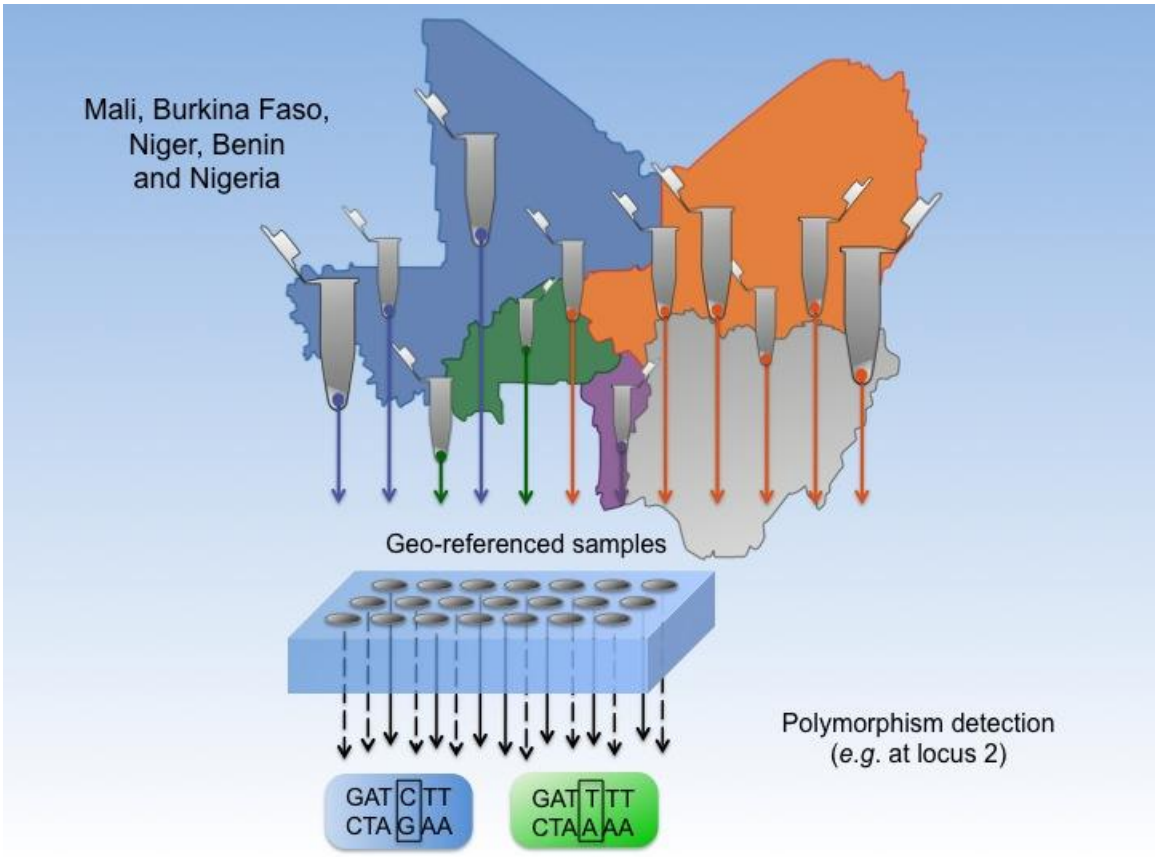


Figure 1.2. Representation of a hypothetical population genomics study within the scope of IPM-omics in which individual insects from different locations (shown by the tubes) are collected from West Africa. Polymorphisms are detected among individual insect samples with high-throughput system(s) (shown as a blue plate), and are used to differentiate genotypes that are correlated back to geographic location, host plant, ecotype, or other ecological characteristics described at the initial collection. Polymorphisms detected between individual insects are indicated as the C/G to T/A differences at the bottom of the figure.

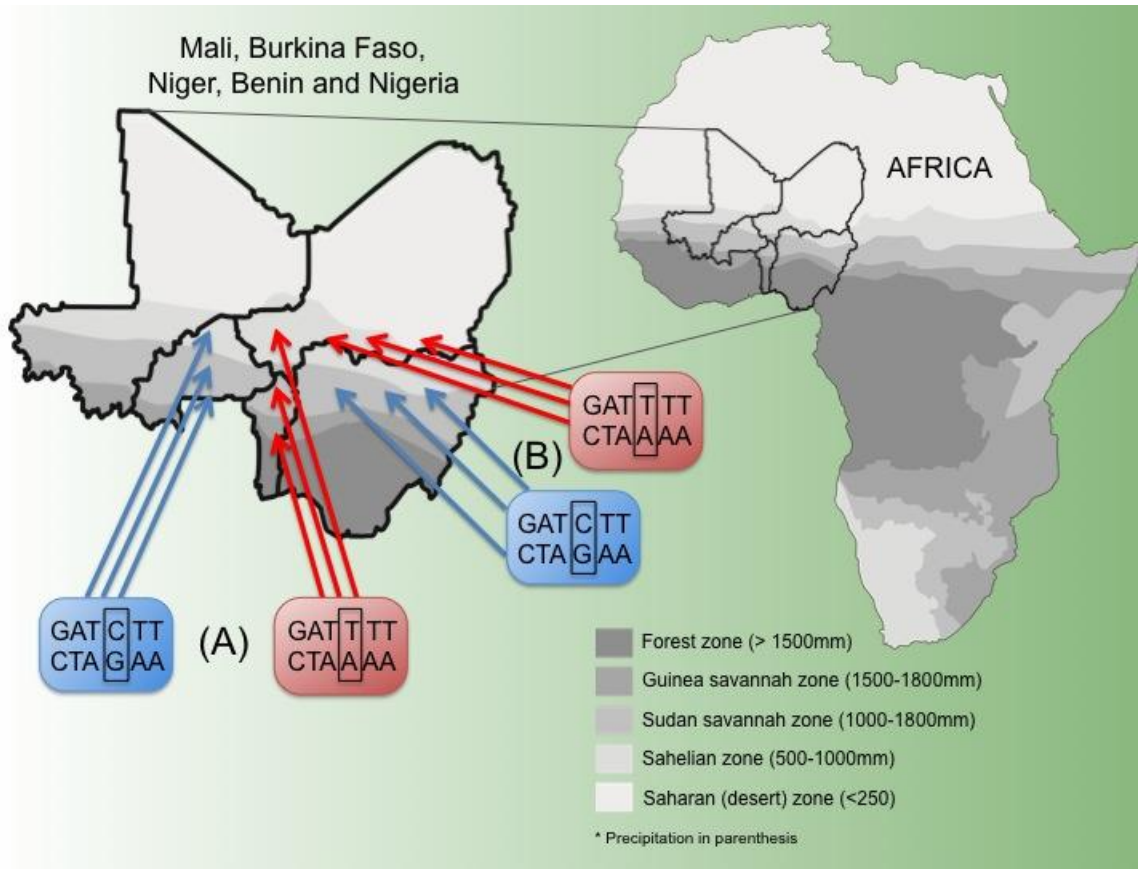


Figure 1.3. Hypothetical scenario of polymorphisms associated with the insect species in the different agro-ecological zones in West Africa. Scenario (A) would indicate a longitudinal movement pattern of an insect population, and scenario (B) would be consistent with the hypothesis that the insect populations are endemic in both areas with no major longitudinal movement patterns. These represent hypothetical situations in which polymorphisms could be used to test hypotheses associated with large-scale insect movement patterns.

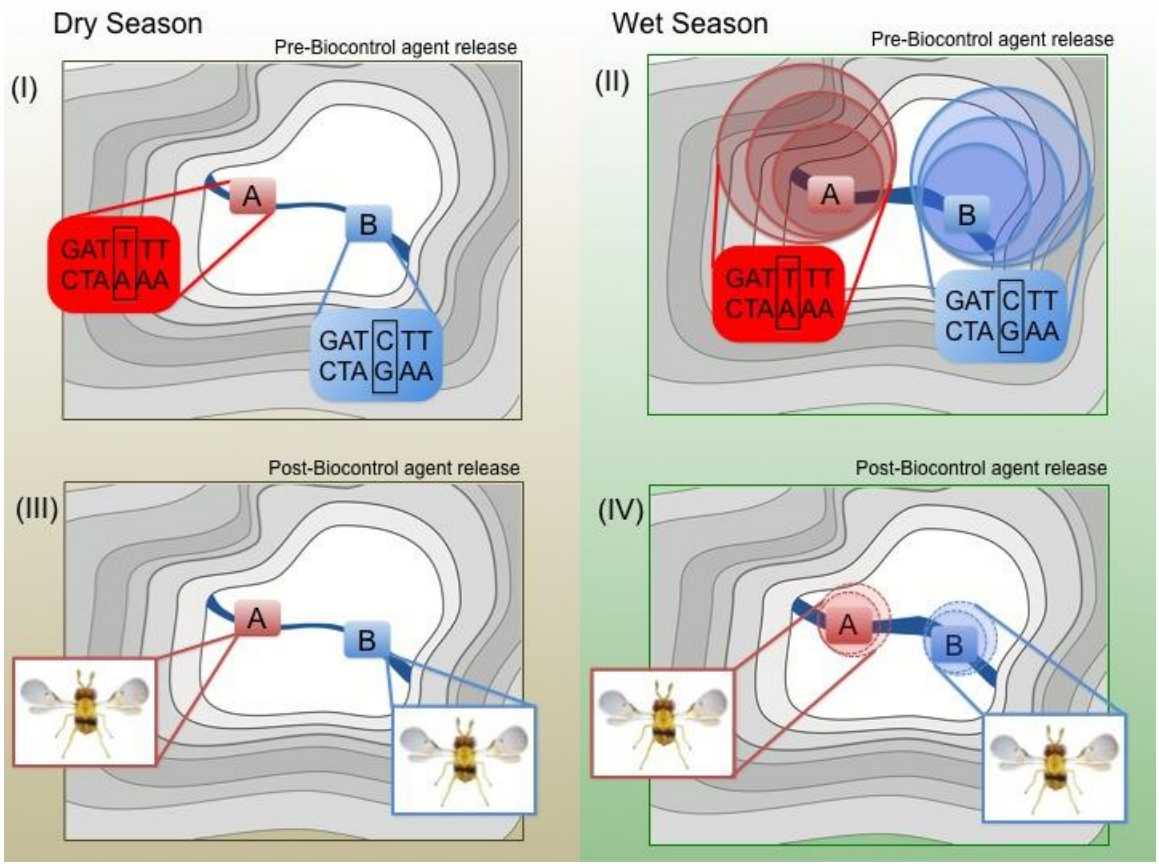


Figure 1.4. Hypothetical use of molecular markers (short DNA sequences given in the blue and red boxes) for the study of local population structure in a pest population in the dry (I) and wet (II) seasons. When two defined source pest populations remain in the dry season (A and B in I), biocontrol agents are released in the areas where the pest populations are endemic (III) to suppress the pest population in the wet season (IV). During the wet season (II), the pest population without biocontrol agents expands as indicated by the sets of red and blue rings around the A and B source populations. In (IV), the biocontrol agents have reduced the pest population levels as indicated by the smaller and more lightly colored red and blue rings around the A and B source populations.

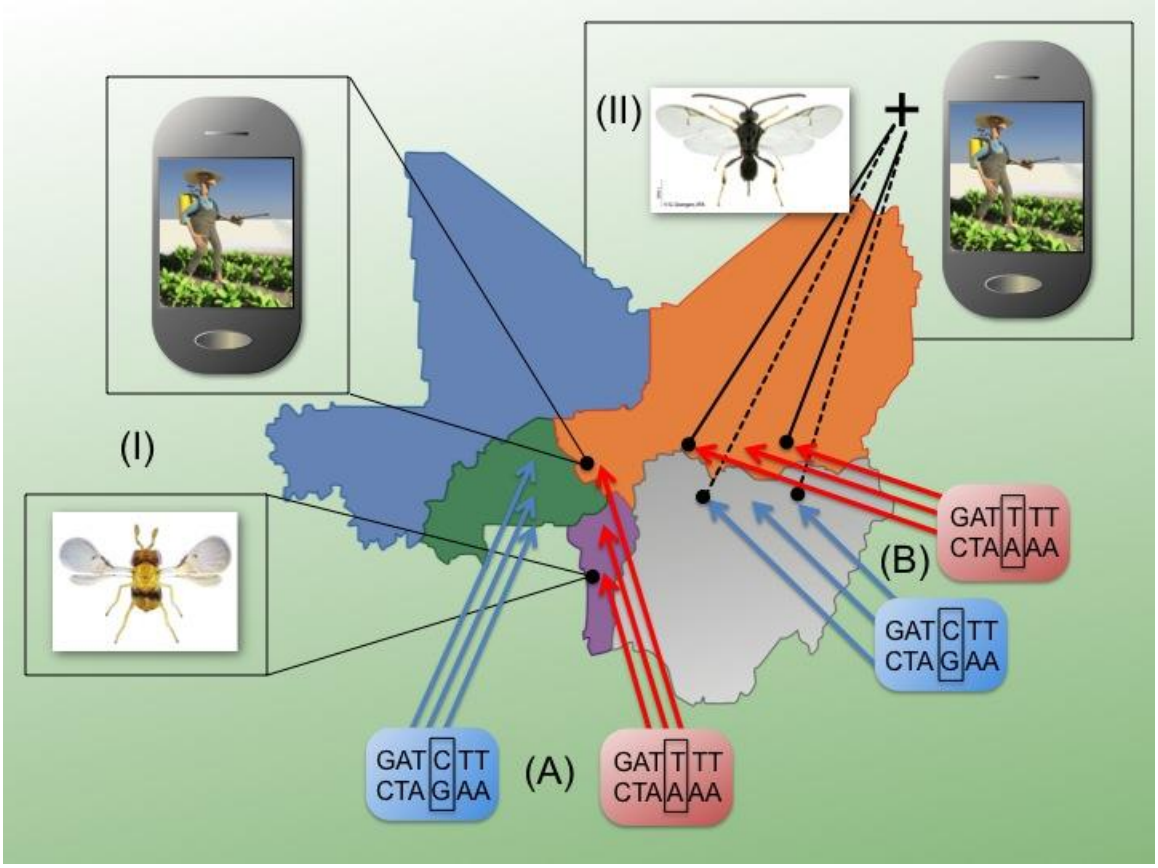


Figure 1.5. Hypothetical use of information gained from genomic studies of insect populations and how such information could be used in decisions associated with deployment of IPM strategies. Two hypothetical scenarios are illustrated: scenario A with biocontrol approach I and scenario B with biocontrol approach II. In scenario A, the insects move from the south to the north in the rainy season, and biocontrol agents are released in the south (in endemic regions) before migration for long-term suppression of the pest population (biocontrol approach I); temporary control measures (e.g., neem sprays) may be required in non-endemic regions if pest levels become too high. *Apanteles taragamae* is shown as an example of a biocontrol agent. In scenario B, the pest is endemic in both regions, with no significant longitudinal movement patterns. A possible pest control strategy in scenario B could involve the release of parasitoids (in both the south and north regions) (biocontrol approach II) as well as implementation of an education program concerning the temporary control (e.g., with neem sprays) if pest levels become damaging. *Trichogramma eldanae* is shown as an example of a parasitoid. The farmer in the figure is applying a plant-based extract spray and thus does not require personal protection equipment; if a synthetic pesticide were used in this extension material, the farmer would have to be shown with personal protection equipment.

Table 1.1. Summary of the major pests of cowpea in West Africa, important areas of research for IPM-omics strategies, genomics tools needed in the near future, and genomics tools currently available or currently being developed.

Pest Species	Important research areas of immediate concern	Current or potential control strategies	Genomic tools needed (short-term)	Genomic tools available or being developed
Maruca vitrata Fabricius (Lepidoptera: Crambidae)	**Local movement Patterns** *Implications* 1) Insect resistance management (IRM) 2) Deployment of biological control agents	1) *Bt* cowpea 2) Biological control 3) Chemical pesticides 4) Bio-pesticides 5) Viral sprays	1) Greater diversity of ESTs to perform more in-depth local studies	1) EST libraries and polymorphisms (Margam et al. 2011c) 2) Studies already performed to show two species in the old and new world 3) Microsatellites for characterization of pest populations (Agunbiade et al. 2012, 2014) 4) Mitochondrial genome characterized (Margam et al. 2011b)
Callosobruchus maculatus Fabricius (Coleptera: Bruchidae)	**Regional movement patterns** *Implications* 1) IRM plans for *Bt* cowpea 2) Deployment of biological control agents 3) Resistance to low oxygen conditions and potential molecular mechanisms	1) Triple bagging 2) Solarization 3) Fumigants 4) Bio-pesticides	1) RNAi system needs to be developed for functional –omics experiments 2) Continued molecular marker development for population studies	1) Microarrays 2) Bioassay system available 3) ESR libraries available and sequenced 4) Studies on the functional genomics of responses to low oxygen conditions (Chi et al. 2011) and genomic and proteomic responses to plant defensive proteins 5) Genome size determination

Table 1.1. (cont.)

Megalurothrips sjostedti Trybom (Thysanoptera: Thripidae)	1) Population structure and dynamics 2) Interactions: Host plant resistance/tolerance – biological control 3) Insecticide resistance	1) Host plant resistance/field tolerance 2) Chemical pesticides 3) Bio-pesticides 4) Biological control 5) Cultural control	1) EST libraries and sequence data generated 2) Molecular marker development for population studies	1) EST libraries sequenced and available (Agunbiade et al. 2013)
Aphis craccivora Koch (Hemiptera: Aphididae)	1) Population structure and dynamics 2) Interactions: Host plant resistance/tolerance – biological control 3) Insecticide resistance 4) Presence of genetically different populations	1) Host plant resistance/field tolerance 2) Chemical pesticides 3) Bio-pesticides 4) Biological control	1) EST libraries and sequence data generated 2) Molecular marker development for population studies	1) EST libraries sequenced and available (Agunbiade et al. 2013)
Clavigralla tomentosicollis Stål (Hemiptera: Coreidae)	1) Population structure and dynamics 2) Interactions: Host plant resistance/tolerance – biological control 3) Insecticide resistance 4) Deployment systems for augmentative biological control	1) Host plant resistance/field tolerance 2) Chemical pesticides 3) Biological control	1) EST libraries and sequence data generated 2) Molecular marker development for population studies	1) EST libraries sequenced and available (Agunbiade et al. 2013)

Table 1.1. (cont.)

Sericothrips adolfifriderici Karny (Thysanoptera: Thripidae)	1) Population structure and dynamics 2) Interactions: Host plant resistance/tolerance – biological control 3) Insecticide resistance	1) Host plant resistance/field tolerance 2) Chemical pesticides 3) Biological control	1) EST libraries and sequence data generated 2) Molecular marker development for population studies	1) EST libraries in progress
Anoplocnemis curvipes Fabricius (Hemipera: Coreidae)	1) Population structure and dynamics 2) Insecticide resistance 3) Deployment systems for augmentative biological control	1) Chemical pesticides 2) Bio-pesticides 3) Biological control	1) EST libraries and sequence data generated 2) Molecular marker development for population studies	1) EST libraries sequenced and available (Agunbiade et al. 2013)

CHAPTER 2

THE SPATIAL GENETIC DIFFERENTIATION OF THE LEGUME POD BORER, *MARUCA VITRATA* F. (LEPIDOPTERA: CRAMBIDAE) POPULATIONS IN WEST AFRICA[2]

ABSTRACT

The legume pod borer, *Maruca vitrata*, is an endemic insect pest that causes significant yield loss to the cowpea crop in West Africa. The application of population genetic tools is important in the management of insect pests but such data on *M. vitrata* is lacking. We applied a set of 6 microsatellite markers to assess the population structure of *M. vitrata* collected at five sites from Burkina Faso, Niger and Nigeria. Observed polymorphisms ranged from one (marker 3393) to eight (marker 32008) alleles per locus. Observed and expected heterozygosities ranged from 0.0 to 0.8 and 0.0 to 0.6, respectively. Three of the loci in samples from Nigeria and Burkina Faso deviated significantly from Hardy-Weinberg Equilibrium (HWE), whereas no loci deviated significantly in samples from Niger. Analysis of molecular variance (AMOVA) indicated that 67.3% level of the genetic variation was within individuals compared to 17.3% among populations. A global estimate of F_{ST} = 0.1 (ENA corrected F_{ST} = 0.1) was significant ($P \leq 0.05$) and corroborated by pairwise F_{ST} values that were significant

[2] This chapter has been published as Agunbiade et al. 2012. The spatial genetic differentiation of the legume pod borer, *Maruca vitrata* F. (Lepidoptera: Crambidae) populations in West Africa. *Bulletin of Entomological Research* 102(5): 589-599. It is reprinted with the permission of the copyright owner and is available at http://journals.cambridge.org and DOI: http://dx.doi.org/10.1017/S0007485312000156. I acknowledge the contribution of co-authors to the publication.

among all possible comparisons. A significant correlation was predicted between genetic divergence and geographic distance between subpopulations (R^2 = 0.6, P = 0.04), and cluster analysis by the program STRUCTURE predicted that co-ancestry of genotypes were indicative of three distinct populations. The spatial genetic variance among *M. vitrata* in West Africa may be due to limited gene flow, south-north seasonal movement pattern or other reproductive barriers. This information is important for the cultural, chemical, and biological control strategies for managing *M. vitrata*.

INTRODUCTION

The legume pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae), is one of the major pests of grain legumes in the tropics and subtropics. Its emergence as a serious pest is attributed to an extensive host plant range, distribution and persistence. The geographic range of *M. vitrata* extends from northern Australia and East Asia through sub-Saharan Africa (Taylor 1967; Raheja 1974; Katayama and Suzuki 1984; Ke et al. 1985; Sharma 1998) to the Americas (Wolcott 1933; Taylor 1967; Munroe 1995). The larval stages of *M. vitrata* are destructive within agricultural and forest eco-systems as they feed on the tender parts of the plant stems, peduncles, flower buds, flowers and pods (Singh and Jackai 1988) of more than 39 host plants belonging mostly to the family Fabaceae (Singh and van Emden 1979; Sharma et al. 1999; Arodokoun et al. 2006).

In West Africa, *M. vitrata* is one of the major pests of cowpea, especially in the countries of Nigeria, Niger and Burkina Faso, which are the major cowpea producing areas. Cowpea production in West Africa accounts for more than 80% of the world's production (Ortiz 1998; FAOSTAT 2000; Sawadogo 2009), but typical infestations by *M.*

vitrata can cause yield reductions of 20 to 80% (Taylor 1967; Raheja 1974; Katayama and Suzuki 1984; Ke et al. 1985; Singh et al. 1990; Sharma 1998). The year-to-year crop losses and challenges faced in the effective field control have led to the identification of this pest as a major threat to economic and humanitarian well-being in developing and under-developed nations. Efforts to enhance the effectiveness of biological control agents that attack *M. vitrata* were partially successful following the importation of the parasitoid wasp, *Apanteles taragamae,* into Benin from Taiwan by the International Institute of Tropical Agriculture (IITA) (Dannon et al. 2010). However, this has been hindered by the lack of population genetic data and information regarding the structure of *M. vitrata* populations in West Africa. Although significant advances have been made to understand the life-history and distribution patterns of *M. vitrata* using light trap and field studies (Ba et al. 2009; Baoua et al. 2011), extensive population-level data are still needed for deployment of biocontrol agents to be effective (Margam et al. 2011). The application of population genetic data to biological control of *M. vitrata* will provide better information on how many distinct genotypes exist and the effect this can have on the parasitoid population over time. If different distinct genotypes exist, then this could result in the formation of specialized parasitoid populations on the different host genotypes in a process termed sequential speciation. So, over time, parasitoids on a particular genotype or host population may be isolated from other parasitoids on other populations. As herbivorous insects and their parasitoids interact with their environment on a fine spatial and temporal scale, sequential radiation may be quite common (Feder and Forbes 2010). More importantly, detailed population genetics data can be used to better target biological control interventions. First, the origin of a given pest can be traced by comparing the

population genetics of the same organism across continents where it occurs, thus allowing to focus the search for efficient natural enemies in areas where both the pest and its antagonistic organisms have co-evolved (Roderick 1996). Once the most efficient natural enemies have been identified, their deployment in the field can be guided by the genetic structure of the target pest population in the likely area of introduction (Roderick and Navajas 2003).

Prior studies in sub-Saharan Africa, including Burkina Faso, have suggested that seasonal flowering patterns of the different host plants on a south-north gradient may influence the migration of *M. vitrata* (Ba et al. 2009). This seasonal movement occurs from temperate conditions along the coast into the Savannas of West Africa as the rainy season progresses (Bottenberg et al. 1997). Migrating *M. vitrata* find favorable feeding and reproductive conditions on a succession of different host plants and, thereby, increase the population size and density with each successive generation. Despite the results of previous studies, many questions remain regarding the timing and spatial scale of *M. vitrata* migration patterns. A preliminary survey that genotyped *M. vitrata* from West Africa using 11 single nucleotide polymorphism (SNP) markers indicated that the population may show population subdivision (Margam et al. 2011). Corroborating evidence is still necessary in order to validate the conclusions drawn from prior SNP analyses and it is yet to be seen whether alternate molecular genetic markers, such as highly polymorphic microsatellites, have a similar potential to differentiate populations.

We developed and applied a set of microsatellite markers to estimate the genetic variability, population structure and gene flow among *M. vitrata* in the West African countries of Niger, Nigeria and Burkina Faso. The main objective of this study was to

assess the genetic variability in the *M. vitrata* populations across West Africa, where these data will be useful for (i) determining effective areas for the release of biocontrol agents and, (ii) the recommendation of insect resistance management (IRM) protocols aimed at minimizing the threat of selection for insecticide resistance alleles in the major cowpea producing areas in West Africa.

MATERIAL AND METHODS

DNA sequence libraries

A combined assembly from Roche 454 reads of a premolt 4th to 5th instar *M. vitrata* larval EST library and Sanger-based EST reads from whole *M. vitrata* adults (referred to as the "reference assembly") was generated from *M. vitrata* collected from Maradi in Niger; Zaria in Nigeria; and Fada, Farakoba, and Kamboinse in Burkina Faso between 2005 and 2007 (see Margam et al. 2011; Figure 2.1 provides a map of the locations). Specifically, adult and larval samples from Nigeria were collected from a cowpea field around Zaria. The adults were collected using light traps. Samples from Niger were also collected from a light trap at Maradi field station while samples from Burkina Faso were collected in cowpea plots at the Institut de l'Environnement et de Recherches Agricoles (INERA) Fada, Farakoba, and Kamboinse stations. This microsatellite study was conducted using DNA samples from 72 individuals from Niger, 53 individuals from Nigeria, and 175 individuals from Burkina Faso (Fada – 40, Farakoba – 86, and Kamboinse – 49). Additionally, a microsatellite repeat enriched partial genomic library was constructed using biotinylated $(CA)_{15}$ and $(GA)_{15}$ probes. Biotin probe-selected fragments were ligated into the pBluescript SK+ vector, which was used to

transform the *Escherichia coli* strain XL1 Blue (Stratagene) by electroporation. Transformants were plated on LB agar containing 20 μg/ml ampicillin and clones were picked, cultured, and plasmid DNA purified at the Purdue University Genomics Core Facility (PUGCF), West Lafayette, Indiana. Additionally, PUGCF performed sequencing of plasmid inserts on an ABI 3730XL sequencer, as well as vector sequence trimming, and Phred quality parameter assessment and trimming at $q < 30$ (99% base call accuracy). Sanger sequence reads from the microsatellite-enriched library were assembled with CAP3 (Huang and Madan 1999) using default parameters, and contig and singleton sequences pooled into a single FASTA-formatted file.

Microsatellite prediction, repeat filtering, and marker development

The EST "reference assembly" and DNA sequence files were merged and used as input for the program SciRoKo (Kofler et al. 2007), where a search for arrays ≥ 10 units at di-, tri-, and tetranucleotide repeat loci was specified. The *M. vitrata* sequence from +250 to -250 of predicted microsatellite repeats was parsed from FASTA files using "SciRoKo's Little Helper" application (Chunk size 50000000; overhead 0), and corresponding positional information exported in tab-delimited format. Repetitive DNA has been described in proximity to microsatellite loci in Lepidopteran genomes and proven problematic for the development of locus-specific molecular genetic markers (please see Discussion). In order to identify repetitive DNA adjacent to *M. vitrata* microsatellites, we employed a bioinformatic pipeline to predict loci with homologies to known Lepidopteran transposable elements and showing sequence similarity to other unrelated *M. vitrata* microsatellites. To do this, the *Bombyx mori* transposon database was

downloaded (file silkworm_glean_transposons.fa.tar.gz at http://sgp.dna.affrc.go.jp/pubdata/index.html), and was queried with *M. vitrata* microsatellite sequences using the Blastn algorithm and results filtered for *E*-values $\leq 1 \times 10^{-15}$ (low-complexity filter was not used). NCBI nr database accessions identified as microsatellite loci from Lepidoptera were downloaded and queried with *M. vitrata* microsatellite sequences as described previously. *Maruca vitrata* sequences that produced a "hit" with an *E*-value $\leq 1 \times 10^{-15}$ to a *B. mori* transposon or Lepidopteran microsatellite sequence were removed from the merged *M. vitrata* dataset.

The remaining *M. vitrata* sequence were used as input for BatchPrimer3 (You et al. 2008; http://probes.pw.usda.gov/batchprimer3/index.html) and primer pairs were picked with the SSR screening and primers module [Pattern type = di-, tri- or tetranucleotide; Product Size = 90 to 300 (150 opt); Primer Size = 18 Min. 21 Opt. 26 Max.; Primer Tm = 58 Min. 60 Opt. 65 Max). Primer and predicted PCR amplified genome sequence information was used as input for the program MultiPLX 2.0 (Kaplinski et al. 2005; http://bioinfo.ut.ee/multiplx/), where potential primer-primer interactions were predicted using a monovalent salt concentration of 50 mM and a Mg concentration of 1.5 mM. Alignment energies for all primer pairs and products, and locus groupings used the normal stringency conditions according to author instructions. Oligonucleotides were synthesized and 5′-dye labels were added to forward primers by Integrated DNA technologies (Coralville, Iowa).

Maruca vitrata sampling and genotyping

Maruca vitrata samples were collected from West African sites and genomic DNA extracted as described by Margam et al. (2011), and DNA concentrations were adjusted to ~ 10 ng/µl with nuclease free water. PCR reactions of 25 µl were set up with 2.0 mM $MgCl_2$, 150 µM dNTPs, ~ 15 ng DNA, 1.75 pmol each primer (multiplex reactions indicated in Table 2.1), 2 µl 5X PCR buffer and 0.3125U Go*Taq* DNA polymerase (Promega, Madison, WI), then amplified using the touchdown thermocycler program TD2 (Coates et al. 2009). A total of 2µl of each of the 5 PCR product for individual samples was pooled, diluted to a total volume of 128 µl, and a 5 µl aliquot submitted to the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign for separation on an ABI Prism 3730xl Analyzer with the LIZ500 internal standard. Fragment analysis was performed using Peak Scanner software 1.0 (Applied Biosystems, Foster City, CA) for data scoring and GeneMapper software for the identification and classification of alleles present per locus and per individual.

Genetic structure of M. vitrata

The mean number of alleles per locus, observed heterozygosity and expected heterozygosity were used to estimate within population genetic variability using Arlequin 3.5.1.2. (Excoffier and Lischer 2010). Analysis of molecular variance (AMOVA) was also performed using Arlequin 3.5.1.2. Variance components were used to compute fixation indices, and their significance was tested at 1000 permutations as described by Weir and Cockerham (1984) in Arlequin. The program Micro-Checker 2.2.3 (Van Oosterhout et al. 2004) was used to estimate the frequency of null alleles and other genotyping errors such

as stuttering and allele drop out. Null alleles are suspected for a given locus when the Micro-Checker 2.2.3 program rejects Hardy-Weinberg Equilibrium (HWE) among genotypes and if the excess homozygote genotypes are evenly distributed among allele size classes. Because some of the alleles harbored potential null alleles, corrected pairwise F_{ST} estimates were calculated for all populations by applying the ENA correction in the FREENA package (Chapuis and Estoup 2007; Chapuis et al. 2008). F_{ST} values were estimated following Weir (1996) while the null allele frequencies for each locus and population was analyzed following the expectation maximization (EM) algorithm (Dempster et al. 1977). Exact tests using a Markov chain were used to test the deviation from HWE at each locus and population (Guo and Thompson 1992) as implemented in Arlequin 3.5.1.2. Locus-by-locus *F*-statistics and pairwise F_{ST} estimates were also calculated using Arlequin 3.5.1.2. Due to the fact that pairwise F_{ST} estimates among sample sites represent multiple comparisons that are considered simultaneously, applications of a significance threshold that treats these comparisons as a single comparison (that is $\alpha = 0.05$) fails to recognize that as the number of comparisons increases the probability that any comparison will differ by random chance also increases (Miller 1981). Specifically, when multiple hypothesis tests are performed, the experiment-wise (EW) Type I error is quickly increased at the rate of $1 - (1 - \alpha)^k$; where k is the number of hypothesis tests performed. To account for the presence of multiple dependent tests within our pairwise F_{ST} estimates, we implicated a correction to the significance thresholds used and determined the critical value according to B-Y method by Benjamini and Yekutieli (2001). Isolation by distance (Slatkin 1993) was tested by analyzing the independence between geographical and genetic distances (Bohonak 2002). The

relationship was assessed by the Mantel test after 1000 permutations using a program IBD 1.52 (http://www.bio.sdsu.edu/pub/andy/IBD.html).

The program STRUCTURE 2.3.3 uses a model-based clustering to predict population structure using genotypic marker data from individual samples, where the model assigns proportions of individual genotypes to one of *K* populations (Pritchard et al. 2000). To accomplish this we used an admixture model to define individual ancestry and included data of sampling location as *a priori* informative descriptors of potential shared co-ancestry among the genotypes (LOCPRIOR command) according to the STRUCTURE 2.3.3 modification described by Hubisz et al. (2009). STRUCTURE 2.3.3 was run using an initial burn-in of 100,000 iterations followed by 100,000 iterations. We ran ten replicates with each value of *K* ranging from 1 to 10. The "real" value of *K* (number of potential unique populations represented by the *M. vitrata* genotypes within our sample) was estimated from the ln Pr(X|*K*) values output for each replicate of *K*=1 to *K*=10 using the mL"(*K*)/sL(*K*) statistic described by Evanno et al. (2005). In brief, the "real" value of *K* within our dataset was determined by where the ln Pr(X|*K*) maximized the value of mL"(*K*)/sL(*K*). A graphical display of individual coancestry (*Q*-matrix) data generated within STRUCTURE 2.3.3 output was performed using the program Distruct (Rosenberg 2004).

RESULTS

DNA sequence libraries

The combined assembly of *M. vitrata* EST read data produced a total of 3499 contigs of 452.9 ± 279.9 bp (see Margam et al. 2011 for details). Sequencing of 480

clones from the $(CT)_n$ and $(GT)_n$ microsatellite repeat resulted in a total of 461 high quality sequences of 285.8 ± 200.8 bp (131.8 kb total; GenBank Accession Numbers JN685509 – JN685580). Assembly of the $(CT)_n$ and $(GT)_n$ microsatellite repeat sequences using CAP3 resulted in 46 contigs (5.9 ± 7.2 sequences per contig) and 69 singletons that were merged into a single dataset with a mean length of 442.1 ± 300.3 bp.

Microsatellite prediction, repeat filtering, and marker development

A search for microsatellite-like repeats within 115 sequences in file "MvMsatCAGA.fasta" predicted a total of 118, 26, and 11 di-, tri-, and tetranucleotide repeat motifs, respectively. All of the 118 putative dinucleotides were predicted from the (AC) and (AG) repeat enriched libraries, and respectively showed a mean length of 57.1 nt (28.5 repeats) and 56.6 nt (28.3 repeats; Table 2.2). In contrast, all putative tri- and tetranucleotide repeats were characterized from EST library sequences. The mean mismatch of nucleotides within the predicted array was ≥ 3.7-fold higher among dinucleotides compared to either the tri- or tetranucleotide repeat groups. Filtering of sequences ± 250 bp of putative *M. vitrata* microsatellite repeats against *B. mori* transposon-like sequences resulted in identification of 12 putative repetitive elements (*E*-values $\leq 2.0 \times 10^{-16}$; similarity ≥ 81.5%; L = 164.7 ± 111.0 bp; remaining data not shown). An analogous Blastn search of the NCBI nr nucleotide database accessions from Lepidoptera that contained microsatellite sequences indicated that sequence similarities existed with 28 *M. vitrata* microsatellite flanking sequences (*E*-values $\leq 2.0 \times 10^{-16}$; similarity ≥ 86.7%; L = 245.5 ± 77.8 bp; remaining data not shown). In total, 40 sequences were identified within *M. vitrata* microsatellite flanking sequences (34.8%),

and were removed from the file prior to PCR primer design. All of the filtered sequences were derived from the anonymous (AC) and (AG) microsatellite-enriched library.

The design of oligonucleotide primers from the remaining 75 sequences using BatchPrimer3 resulted in 24 pairs (11 from anonymous microsatellite enriched library and 13 from EST sequence). Preliminary analysis by PCR resulted in the successful amplification for one of 11 anonymous microsatellite markers (9.1%) and five of 13 EST-derived microsatellite markers (38.5%; results not shown). PCR primer multiplex pair design using MultiPLX resulted in the prediction of two loci being suitable for co-amplification (markers C3393 and C0444), where primer alignment energies were at a maximum of −5.1 kcal mol^{-1} compared to −6.0 ±0.9 kcal mol^{-1} for all remaining possible primer pairs (remaining data not shown).

Genetic structure of M. vitrata

The six microsatellite loci used to screen *M. vitrata* samples collected at five African sites showed significant deviation from HWE at 13 instances following 30 locus-by-site calculations (Appendix A), thus subsequent estimations of population subdivision used ENA corrected values to account for potential influence of null alleles. The mean number of alleles per locus was similar across all sample sites, and the overall observed heterozygosity was less than expected ($F_{IS} \geq 0.1$) for all populations except at Fada, Burkina Faso (F_{IS} = -0.0; Table 2.3). Also the F_{IS} estimates were negative for only two loci (CO241 and CO325) across all populations (Table 2.4). The locus-by-locus F_{ST} estimates derived from the five populations ranged from −0.0 (ENA corrected F_{ST} = 0.0; marker 32008) to 0.3 (ENA corrected F_{ST} = 0.0; marker 7_02K06; remaining data not

shown). The subsequent global estimates of F_{ST} across all loci and all populations were moderate (uncorrected F_{ST} = 0.1, ENA corrected F_{ST} = 0.1) and significant (95% CI = 0.0 – 0.3, 95% CI with ENA correction = 0.0 – 0.2, with the lower bounds rounded to zero in one decimal place). The partitioning of population genetic variance from AMOVA results indicated that ≥ 67.3% resides within individuals, and correspondingly 17.3% and 15.4% of the total genetic variation, was among populations and among individuals (Table 2.5). Pairwise comparison of F_{ST} estimations showed that a significant level of differentiation exist for all the possible comparisons after the sequential B-Y adjustment for the ENA corrected F_{ST} estimates across all loci (Table 2.6). Specifically, a critical significance level was achieved at 0.05/2.929 = 0.017. Results indicated that the *P*-values obtained from pairwise F_{ST} estimates (≤ 0.005) were all statistically significant at the B-Y adjusted thresholds (Table 2.6). Regression of uncorrected F_{ST} estimates and geographic distance (km) among West African sample sites showed a significant dependence of genetic variation on geographic distance (R^2 = 0.6, Mantel *P* = 0.04), and showed the relative genetic similarity (F_{ST} estimates) of genotypes at Niger and Nigeria, and among Burkina Faso samples (Figure 2.2). The "real" number of populations (*K*) estimated from the microsatellite-defined *M. vitrata* genotypes from the mL"(*K*)/sL(*K*) statistic calculated from STRUCTURE 2.3.3 output achieved a maximum value of 14.5 at *K* = 3 and suggested that three genetically-distinct *M. vitrata* ancestries exist in West Africa. The three genetically distinct ancestries across collection sites were represented in Figure 2.3 as vertical bars with Niger, Nigeria and Fada, Burkina Faso primarily red; Farakoba, Burkina Faso primarily green; Kamboinse, Burkina Faso primarily green and red and a minor cluster represented by yellow across all sites.

DISCUSSION

Microsatellite markers in Lepidoptera

The isolation and subsequent application of microsatellite loci as molecular genetic markers in Lepidopteran insects has often been difficult (Nève and Meglécz 2000; Ji and Zhang 2004; Zhang 2004), where studies have developed five or fewer loci for most species studied so far (reviewed in Ji et al. 2003). These difficulties have been attributed to the high degrees of nucleotide sequence similarity between regions that flank different microsatellite loci (Meglécz et al. 2004) or the low frequency of microsatellites in Lepidopteran genomes (Ji et al. 2003; Prasad et al. 2005). Additionally, evidence suggests that Lepidopteran microsatellites may be derived from and mobilized by transposable elements (TEs) (Coates et al. 2009, 2010, 2011, 2012; Tay et al. 2010). Despite these associations with repetitive DNA, microsatellite markers have been developed for more than 40 Lepidopteran species in 35 genera (GenBank, 15 August 2006) and have proven to be useful for population genetic analyses. Within this study, we initially identified 155 unique *M. vitrata* microsatellite loci from which six were eventually considered informative genetic markers. Forty of 155 loci (25.8%) contained known repetitive element-like sequences previously identified within the *B. mori* genome assembly, which suggests that *M. vitrata* microsatellites may be associated with repetitive DNA as observed in other Lepidopteran species (Meglécz et al. 2004). High failure rates observed during microsatellite marker development are rarely reported, and the extent to which repetitive DNA can cause these failures has not been thoroughly investigated (Tay et al. 2010). In the current study, we aimed to identify *M. vitrata* microsatellite loci, which may contain repetitive sequence and remove them from consideration for molecular

marker development. Despite these measures, an approximate 25% success rate for *M. vitrata* microsatellite markers suggest that screening for repetitive DNA from distantly related species may not be sufficient to identify all affected loci.

Microsatellite loci that deviate significantly from HWE show evidence of null alleles according to the distribution of homozygote-size classes. Microsatellite null alleles are commonly found among a wide range of taxa, but have a particularly high incidence among species of Lepidoptera (Meglécz et al. 2004), Diptera (Lehmann et al. 1997), and Orthoptera (Chapius et al. 2005). The extent to which null alleles tend to overestimate the true population differentiation has not been investigated (Chapuis and Estoup 2007), but can lead to overestimates of population differentiation due to effects on subsequent calculations of F_{ST} and genetic distances (Slatkin 1995; Paetkau et al. 1997). Although null alleles were present at all microsatellite loci, correction with the ENA algorithm nonetheless allowed effective population genetic analysis. Specifically, analysis of these microsatellite markers resulted in conclusions analogous to those obtained from *M. vitrata* single nucleotide polymorphism (SNP) markers applied to the same populations (see next section).

Genetic structure of M. vitrata in West Africa

At the population level, our ENA corrected microsatellite genotype data suggested that a moderate level of genetic differentiation may be present among *M. vitrata* sample sites in West Africa. Consistent with previous studies in other insects, a majority of the total genetic variation is within individuals compared to between sample sites (Coates and Hellmich 2003; Juan et al. 2004; Timmermans et al. 2005). Furthermore, our

microsatellite results agree with an analogous study using *M. vitrata* SNP markers in West Africa (Margam et al. 2011), where genetic structure was detected between Niger and Nigeria, and Burkina Faso sample sites and an overall $K = 3$ obtained from STRUCTURE results. The current study also indicated three distinct populations of *M. vitrata* in West Africa and suggests that analysis of microsatellite and SNP markers can provide equivalent results and conclusions. The findings from that study, however, showed significant pairwise F_{ST} estimates between eastern (Niger and Nigeria) and western sample sites from Burkina Faso, and all other comparisons indicated a lack of genetic divergence while findings from our study indicated a significant divergence in all pairwise populations using a critical *P*-value determined using B-Y adjustment method (Benjamini and Yekutieli 2001). Our population differentiation results derived from microsatellite data may be expected since SNPs typically have two alleles per locus due to their low mutational rate (Hancock 1999; Zhang and Hewitt 2003) compared to multi-allelic microsatellites that have higher power per locus for estimating genetic divergence or gene flow when using *F*-statistics or assignment tests (Vignal et al. 2002; Brumfield et al. 2003; Morin et al. 2004). Studies further suggest that measures of pairwise genetic relationships using SNPs would require analysis of more than five times more loci as compared to microsatellites (Blouin et al. 1996; Glaubitz et al. 2003), and suggest that the current set of microsatellite markers will be more efficient at detecting population subdivision compared to SNP markers within future population genetic analyses.

Effect of migration on the population structure in M. vitrata

Migration is a fundamental population process and is crucial to understanding the dynamics and persistence of populations of insects (Dingle 1996). Despite the field observations that *M. vitrata* has high rates of migration, significant difference in F_{ST} estimates and partitioning of 17.3% of population variance between sample sites within AMOVA analysis suggested that *M. vitrata* show reduced gene flow and genetic structuring in West Africa. This is evidenced in the F_{IS} estimates obtained for almost all the study sites except Fada, Burkina Faso where there was some evidence of outbreeding suggesting high gene flow in this population. Specifically, an isolation by distance (IBD) model suggested that subpopulation differentiation (F_{ST}) or structured gene flow is a function of the geographical distance between them (Slatkin 1993). Our analyses indicate that although *M. vitrata* has a seasonal south-north migration, there appears to be evidence of reduced gene flow. In agreement with this, Peterson and Denno (1998) reported that insect species with high- and low-dispersal rates tend to show less IBD compared to species with moderate dispersal capabilities. High levels of gene flow tend to homogenize the observed genetic variation across a geographic range, whereas low gene flow can result in the effective genetic isolation of subpopulations and genetic drift. We observed a positive correlation between geographic and genetic distances among the *M. vitrata* sample sites within this study ($R^2 = 0.56$), suggesting that *M. vitrata* may have a structured migratory pattern. Although isolation by distance is not typical for species with high mobility (Arguedas and Parker 2000), it has been reported for other Lepidoptera such as *Chazara briseis* Linnaeus (Lepidoptera: Satyridae) (Johannesen et al. 1997), and *Hesperia dacotae* Skinner (Lepidoptera: Hesperiidae) (Britten and Glasford 2002). Light trap data

and sampling of *M. vitrata* in West Africa showed that *M. vitrata* does not infest cowpea during the dry season in the northern extremes of its geographic range, even if cowpea is present (Bottenberg et al. 1997; Ba et al. 2009). Climatic factors such as lack of rain, and temperature or relative humidity may influence this observed spatial restriction. Research has also shown that *M. vitrata* survive on alternate host plants within the more humid southern regions, and migrate to the northern regions over a period of several months (Bottenberg et al. 1997). During the seasonal migration, the population of *M. vitrata* finds favorable conditions for multiplying on a succession of different host plants, which may result in temporal mating barriers due to differential larval maturation rates on the alternate host plants. This serves to emphasize the importance of ecological migration (movement of individuals) on population genetic migration (transfer of alleles). Gene flow cannot occur without ecological migration, but the effects of ecological migration on the population genetic structure can be tempered by factors such as temporal mating barriers between dispersing and resident individuals.

CONCLUSIONS

The implementation of effective cultural, chemical, and biological control strategies to limit *M. vitrata* feeding damage to cowpea crops in West Africa is dependent upon a basic understanding of population structure and migration. Our results and previous ecological studies have implications for IRM strategies involving *Bt*-cowpea in West Africa. The implication is that in the north, resistance may spread only slowly among *M. vitrata* populations because the populations eventually die out during the dry season; in other words, the southern populations in endemic zones act as a source

population. Also, the long-distance migration from the south to the north might be a source of susceptible populations into the northern part, which can slow down the evolution of resistance (if there are pockets of *M. vitrata* populations that survive in the north throughout the year). In the more humid south, *M. vitrata* can be found on different host plants throughout the year. Further studies should be conducted on population characterization on the different host plants, especially in the more humid south, to determine if there are differences in populations on the different host plants or if geographic barriers restrict gene flow between them. Migration results in changes in allele frequencies that are greater in the short term and smaller in the long term, leading to under- and overestimation of effective population size, respectively, if it is ignored (Wang and Whitlock 2003). Using these polymorphic markers, studies should also be conducted to test how migration rate affects the effective population size of the *M. vitrata* populations. The results of this study also have implications for the implementation of control strategies involving the release of biocontrol agents against *M. vitrata*. In keeping with an endemic zone to migratory zone hypothesis, the conclusions of this study agree with the conclusions by Margam et al. (2011), which suggested that the deployment of biocontrol agents (for classical biological control) would be most logical in the endemic zone directly south of migratory regions where *M. vitrata* is a significant pest during the cowpea growing season.

REFERENCES

Arguedas N, Parker PG (2000) Seasonal migration and genetic population structure in house wrens. *Condor* 102: 517 – 528.

Arodokoun DY, Tamò M, Cloutier C, Brodeur J (2006) Larval parasitoids occurring on *Maruca vitrata* Fabricius (Lepidoptera: Pyralidae) in Benin, West Africa. *Agriculture, Ecosystems and Environment* 113: 320-325.

Ba NM, Margam VM, Binso-Dabire CL, Sanon A, McNeil JN, Murdock LL, Pittendrigh BR (2009) Seasonal and regional distribution of the cowpea pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae) in Burkina Faso. *International Journal of Tropical Insect Science* 29(3): 109-113.

Baoua I, Ba NM, Agunbiade TA, Margam V, Binso-Dabire CL, Sanon A, Pittendrigh BR (2011) Potential use of *Sesbania pachycarpa* DC (Fabaceae: Papilionoideae) as a refugia for the legume pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae). *International Journal of Tropical Insect Science* 31: 212-218.

Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Annals of Statistics* 29: 1165-1188.

Blouin MS, Parson M, Lacaille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* 5 (3): 393-401.

Bohonak AJ (2002) IBD (Isolation By Distance): A program for analyses of isolation by distance. *Journal of Heredity* 93: 153-154.

Bottenberg H, Tamò M, Arodokoun D, Jackai LEN, Singh BB, Youm O (1997) Population dynamics and migration of cowpea pests in northern Nigeria: Implications for integrated pest management. In Singh BB, Mohan Raj DR, Dashiell KE, Jackai LEN (Eds.), *Advances in cowpea research*. Co-publication of International Institute of Tropical Agriculture (IITA) and Japan International Center for Agricultural Sciences (JIRCAS). IITA, Ibadan, Nigeria, pp 271–284.

Britten HB, Glasford JW (2002) Genetic population structure of the Dakota skipper (Lepidoptera: Hesperia dacotae): A North American native prairie obligate. *Conservation Genetics* 3: 363–374.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18 (5): 249-256.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24(3): 621-631.

Chapuis MP, Loiseau A, Michalakis Y, Lecoq M, Estoup A (2005) Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria*. *Molecular Ecology Notes* 5*:* 554-557.

Chapuis MP, Lecoq M, Michalakis Y, Loiseau A, Sword GA, Piry S, Estoup A (2008) Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Molecular Ecology* 17(16): 3640-3653.

Coates BS, Hellmich RL (2003) Two sex-chromosome-linked microsatellite loci show geographic variance among North American *Ostrinia nubilalis*. *Journal of Insect Science* 3:29.

Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD, Lewis LC (2009) Comparative performance of single nucleotide polymorphism (SNP) and microsatellite markers for population genetic analysis. *Journal of Heredity* 100: 556-564.

Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2010) A Helitron-like transposon superfamily from Lepidoptera disrupts (GAAA)n microsatellites and is responsible for flanking sequence similarity within a microsatellite family. *Journal of Molecular Evolution* 70: 278-288.

Coates BS, Kroemer JA, Sumerford DV, Hellmich RL (2011) A novel class of miniature inverted repeat transposable elements (MITEs) that contain hitchhiking (GTCY)n microsatellites. *Insect Molecular Biology* 20: 15–27.

Coates BS, Hellmich RL, Grant DM, Abel CA **(**2012) Mobilizing the genome of Lepidoptera through novel sequence gains and end creation by non-autonomous *Lep*1 *Helitrons*. *DNA Research* 19(1): 11- 21.

Dannon EA, Tamò M, Van Huis A, Dicke M (2010) Effects of volatiles from *Maruca vitrata* larvae and caterpillar-infested flowers of their host plant, *Vigna unguiculata* on the foraging behavior of the parasitoid *Apanteles taragamae. Journal of Chemical Ecology* 36: 1083–1091.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)* 39(1): 1-38.

Dingle H (1996) Migration: The biology of life on the move. New York: Oxford University Press.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611-2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

FAOSTAT (2000) Site internet : http//www. FAO. org/statistics.

Feder JL, Forbes AA (2010) Sequential speciation and the diversity of parasitic insects. *Ecological Entomology* 35: 67-76.

Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Molecular Ecology* 12(4): 1039-1047.

Guo S, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.

Hancock JM (1999) Microsatellites and other simple sequences: Genomic context and mutational mechanisms. In Goldstein DB, Schlötterer C. (Eds.), *Microsatellites: Evolution and applications*. Oxford University Press, Oxford, pp 1-9.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research* 9: 868-877.

Hubisz M, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322-1332.

Ji YJ, Zhang DX (2004) Characteristics of microsatellite DNA in Lepidopteran genomes and implications for their isolation. *Acta Zoologica Sinica* 50: 608-614.

Ji YJ, Zhang DX, Hewitt GM, Kang L, Li DM (2003) Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. *Molecular Ecology Notes* 3: 102 – 104.

Johannesen J, Schwing U, Seufert W, Seitz A, Veith M (1997) Analysis of gene-flow and habitat patch network for *Chazara briseis* (Lepidoptera: Satyridae) in an agricultural landscape. *Biochemical Systematics and Ecology* 25: 419–427.

Juan A, Crespo MB, Cowan RS, Lexer C, Fay MF (2004) Patterns of variability and gene flow in *Medicago citrina*, an endangered endemic of islands in the western Mediterranean, as revealed by amplified fragment length polymorphism (AFLP). *Molecular Ecology* 13: 2679–2690.

Kaplinski L, Andreson R, Puurand T, Remm M (2005) MultiPLX: Automatic grouping and evaluation of PCR primers. *Bioinformatics* 21: 1701-1702.

Katayama J, Suzuki I (1984) Seasonal prevalence of pod borers (*Ostrinia scapulalis*, *Maruca testulalis* and *Matsumuraeses* sp.) in azuki-beans and injury caused by larval infestation. *Bulletin of Kyoto Prefectural Institute of Agriculture* 12: 27–34.

Ke LD, Fang JL, Li ZJ (1985) Bionomics and control of the legume pod-borer, *Maruca testulalis* Geyer. *Acta Entomologica Sinica* 28(1): 51–59.

Kofler R, Scholtterer C, Lelley T (2007) SciRoKo**:** A new tool for whole genome microsatellite search and investigation. *Bioinformatics* 23: 1683-1685.

Lehmann T, Besansky NJ, Hawley WA, Fahey TG, Kamau L, Collins FH (1997) Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology* 6: 243-253.

Margam VM, Coates BS, Bayles DO, Hellmich RL, Agunbiade T, Seufferheld MJ, Sun W, Kroemer JA, Ba MN, Binso-Dabire CL, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011) Transcriptome sequencing, and rapid development and application of SNP markers for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(7): e21388.

Meglécz E, Petenian F, Danchin E, D'Acier AC, Rasplus JY, Faure E (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology* 13: 1693–1700.

Miller RG (1981) Simultaneous Statistical Inference 2nd Ed. Springer Verlag New York.

Morin PA, Luikart G, Wayne RK, SNP-Workshop_Grp (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19: 208-216.

Munroe EG (1995) Pyraustinae. In Heppner JB (Ed.), *Atlas of neotropical Lepidoptera. Checklist: Part 2 Hyblaeoidea, Pyraloidea, Tortricoidea.* Association for Tropical Lepidoptera, Gainesville, FL, pp 58–79.

Nève G, Meglécz E (2000) Microsatellite frequencies in different taxa. *Trends in Ecology and Evolution* 15: 376–377.

Ortiz R (1998) Cowpeas from Nigeria: A silent food revolution. *Outlook on Agriculture* 27(2): 125 – 128.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* 147: 1943 – 1957.

Peterson MA, Denno RF (1998) Life history strategies and the genetic structure of phytophagous insect populations. In Mopper S, Strauss SY (Eds.), *Genetic structure and local adaptation in natural insect populations: Effects of ecology, life history, and behaviour.* New York, Chapman and Hall, pp. 253–322.

Prasad MD, Muthulakshmi M, Madhu M, Archak S, Mita K, Nagaraju J (2005) Survey and analysis of microsatellites in the Silkworm, *Bombyx mori*: Frequency, distribution, mutations, marker potential and their conservation in heterologous species. *Genetics* 169: 197–214.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Raheja AI (1974) Report on the insect pests of grain legumes in northern Nigeria. In 1st IITA grain legume improvement workshop, 1973. International Institute of Tropical Agriculture, Ibadan, Nigeria, pp 295–299.

Roderick GK (1996) Geographic structure of insect populations: Gene flow, phylogeography, and their uses. *Annual Review of Entomology* 41: 263-290.

Roderick GK, Navajas M (2003) Genes in new environment: Genetics and evolution in biological control. *Nature Reviews Genetics* 4: 889-899.

Rosenberg NA (2004) Distruct: A program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.

Sawadogo A, Thio B, Kiemde S, Drabo I, Dabire C, Ouedraogo J, Mullens TR, Ehlers JD, Roberts PA (2009) Distribution and prevalence of parasitic nematodes of cowpea (*Vigna unguiculata*) in Burkina Faso. *Journal of Nematology* 41:120-127.

Sharma HC (1998) Bionomics, host plant resistance, and management of the legume pod borer, *Maruca vitrata*: A review. *Crop Protection* 17: 373-386.

Sharma HC, Saxena KB, Bhagwat VR (1999) The legume pod borer, *Maruca vitrata*: Bionomics and management. Information bulletin 55. International Crops Research Institute for the Semi-arid Tropics (ICRISAT), Patancheru, India.

Singh SR, Jackai LEN (1988) The legume pod-borer, *Maruca testulalis* (Geyer): Past, present and future research. *Insect Science and its Applications* 9: l – 5.

Singh SR, van Emden HF (1979) Insect pests of grain legumes. *Annual Review of Entomology* 24: 255–278.

Singh SR, Jackai LEN, Dos Santos JHR, Adalla CB (1990) Insect pests of cowpea. In Singh SR (Ed.), *Insect pests of tropical food legumes*, Chichester, John Wiley & Sons, Chichester and UK, pp 43-90.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47: 264-279.

Slatkin M (1995) Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution* 12(3): 473-480.

Tay WT, Behere GT, Batterham P, Heckel DG (2010) Generation of microsatellite repeat families by RTE retrotransposons in Lepidopteran genomes. *BMC Evolutionary Biology* 10: 144.

Taylor MFJ, McKechnie SW, Pierce N, Kreitman M (1993) The Lepidopteran mitochondrial control region: Structure and evolution. *Molecular Biology and Evolution* 10:1259-1272.

Timmermans MJ, Ellers J, Mariën J, Verhoef SC, Ferwerda EB, Van Straalen NM (2005) Genetic structure in *Orchesella cincta* (Collembola): Strong subdivision of European populations inferred from mtDNA and AFLP markers. *Molecular Ecology* 14: 2017–2024.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4(3): 535-538.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 34: 275-305.

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 163: 429-446.

Weir BS (1996) Genetic Data Analysis II. Methods for discreet population genetic data Sinauer Associates, Inc, Sunderland, MA.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Wolcott GN (1933) The lima bean pod borer caterpillars of Puerto Rico. *Journal of the Department of Agriculture of Puerto Rico* 17: 241–255.

You FM, Huo N, Gu YQ, Luo M-C, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9: 253.

Zhang DX (2004) Lepidopteran microsatellite DNA: Redundant but promising. *Trends in Ecology and Evolution* 19: 507-509.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Molecular Ecology* 12 (3): 563-584.

FIGURES AND TABLES

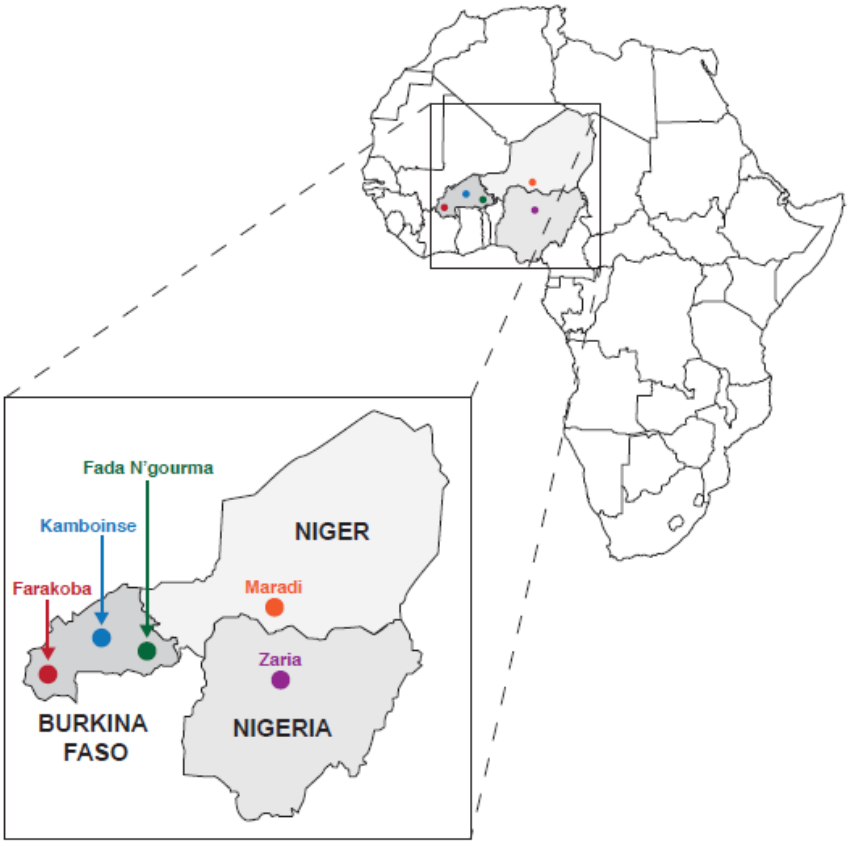


Figure 2.1. Map showing locations in West Africa (Niger, Nigeria and Burkina Faso) where the *M. vitrata* samples were collected.

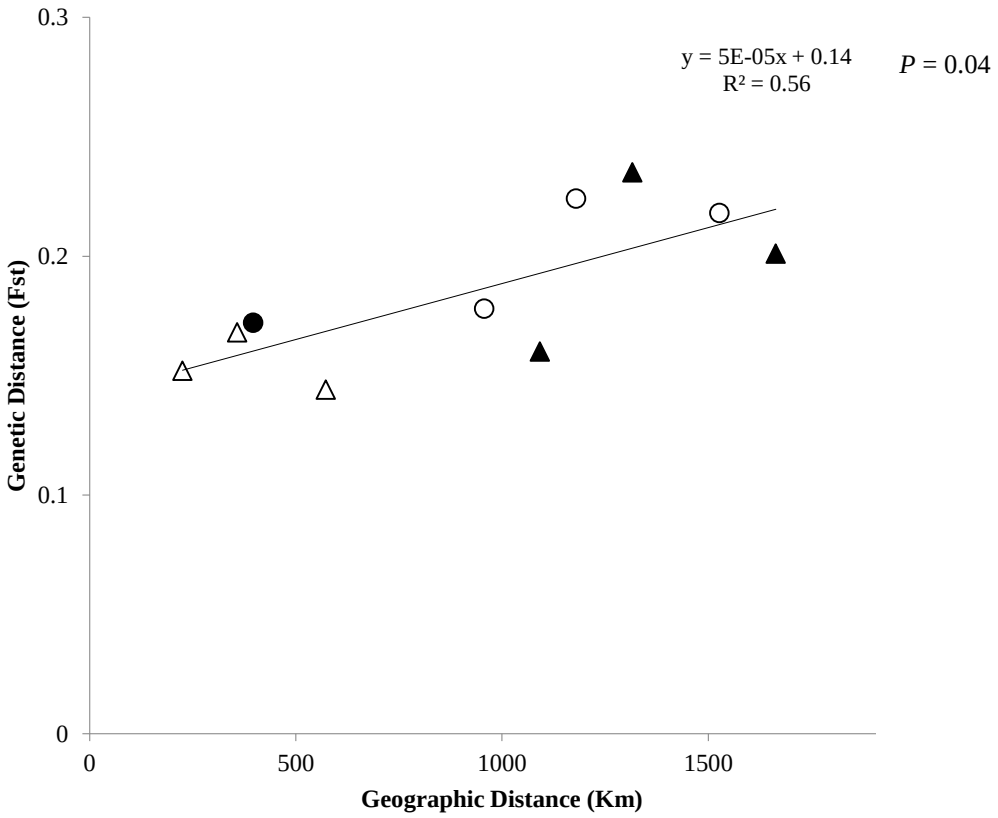


Figure 2.2. Geographic distance versus genetic distance among populations of *M. vitrata* using F_{ST}. Correlations and probabilities were estimated from a Mantel test with 10,000 repeats of bootstrapping. ∆, Comparisons among Burkina Faso (Fada, Farkoba and Kamboinse; ●, Niger and Nigeria; ▲, Niger and locations in Burkina Faso; ○, Nigeria and locations in Burkina Faso are indicated.

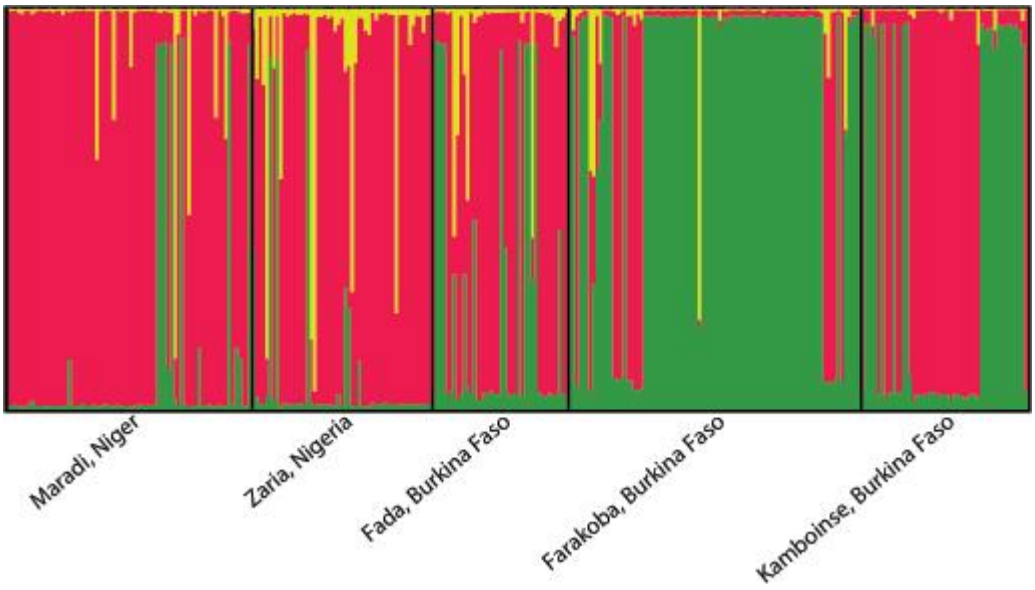


Figure 2.3. Partitioning of the co-ancestry among microsatellite-defined *M. vitrata* genotypes generated from STRUCTURE using the LOCPRIOR command for Maradi, Niger; Zaria, Nigeria; and Fada, Farakoba and Kamboinse locations in Burkina Faso. For each, the estimated co-ancestry was derived from the *Q*-matrix for each individual and represented as vertical lines showing the proportion of the K^{th} segments that made up the individual genotype.

Table 2.1. *M. vitrata* primer sequences used for microsatellite amplification reactions.

Locus	Primer (dye label) and sequence (5' -3')	Repeat	Size (bp)
C32008 [E]	F-(MAX)AAAAAGCGCTTATATGTTTGTTATAGT R-GAAATTTTTAACGGAGATACAATCA	$(CATA)_3$	163
7_02K06 [A]	F-(FAM)ATTTGTCAGAATGGTATCTTACGT R-CCTCTGGGTCATAATTATATTGTTCA	$(GAT)_6$	151
C3393 [E, 1]	F-(ROX)AGACCCCCAAAGTGGAGAA R-ACGTTCACGAACCTCCTGTT	$(GAA)_5$	91
C0444 [E, 1]	F-(FAM)AAAGGAACTACGCCGTCAGG R-GTTGAGCGATCTTGGCACAG	$(CAA)_8$	102
C0241 [E]	F-(TAM)GACGAAACAAGGCCTACCAG R-GGTACTTCYGACGTTGTTCG	$(GAT)_9$	165
C0325 [E]	F-(ROX)CGAAAAGAAACACCGCTCTG R-CAGTCTGTTCAGWCTCTTCAGTGG	$(GAA)_7$	173

E, EST-derived primer pair; A, anonymous genomic sequence-derived primer pair; 1, PCR multiplexed primers

Table 2.2. Microsatellite motifs predicted from *M. vitrata* DNA sequence sources (combined library), that was partitioned into (CA) and (GA) repeat enriched library (microsatellite library), and expressed sequence tag (EST) library (EST library) sources. The mean length of repeat arrays (L) and frequency of nucleotide mismatch within repeat arrays (MisMch).

	Combined library			Microsatellite library			EST library		
Motif	**Count**	**L**	**MisMch**	**Count**	**L**	**MisMch**	**Count**	**L**	**MisMch**
AC	107	57.1	1.6	107	57.1	1.6	0		
AG	11	56.6	1.3	11	56.6	1.3	0		
ATC	14	25.8	0.4	0			14	25.8	0.4
AAT	12	18.4	0.3	0			12	18.4	0.3
AAAT	11	17.8	0.2	0			11	17.8	0.2
	155								

Table 2.3. Characteristics of the *M. vitrata* individuals across West Africa showing sample size (N), number of alleles (Na), observed heterozygosity (H_O), expected heterozygosity (H_E), and fixation index (F_{IS}) per sample site (values rounded up to 1 decimal place).

Location	Na (Mean per locus)	*Ho*	H_E	F_{IS}	Loci not in HWE
Niger	25 (4.2)	0.2	0.2	0.1	0
Nigeria	27 (4.5)	0.2	0.3	0.2	4
Fada	18 (3.0)	0.3	0.3	−0.0	3
Farakoba	23 (3.8)	0.2	0.3	0.2	3
Kamboinse	18 (3.0)	0.2	0.3	0.4	3

Table 2.4. Estimates of total number of alleles (Na), mean estimated frequency of null alleles, fixation index (F_{IS}) and F_{ST} across microsatellite loci (values rounded up to 1 decimal place).

Locus	Na	Mean null alleles frequency	F_{IS}	F_{ST} uncorrected (ENA corrected)
C0444	7	0	0.1	0.0 (0.0)
C32008	10	0.1	0.1	−0.0 (0.0)
C3393	3	0.1	0.9	0.0 (0.0)
7_02K06	6	0.2	0.7	0.4 (0.3)
CO241	7	0	−0.1	0.0 (0.0)
CO325	5	0	−0.1	0.1 (0.1)

Table 2.5. Analysis of Molecular Variance (AMOVA) for Maradi, Niger; Zaria, Nigeria; and Fada, Farakoba, and Kamboinse locations in Burkina Faso.

Source of Variation	Df	SS	% of Variation
Among Population	4	56.1	17.3
Among Individuals	295	190.5	15.4
Within Individuals	300	133	67.3
Total	599	379.6	100

Table 2.6. Pairwise comparisons of *M. vitrata* samples showing the corrected F_{ST} estimates (below diagonal) and significance of corresponding comparisons (*P*-values) as indicated above the diagonal. Significance thresholds were evaluated using a Benjamini and Yekutieli (B-Y) adjusted $\alpha = 0.017$.

	Niger	Nigeria	Fada	Farakoba	Kamboinse
Niger	-	0.005*	< 0.001*	<0.001*	0.001*
Nigeria	0.2	-	< 0.001*	< 0.001*	0.004*
Fada	0.2	0.2	-	< 0.001*	0.002*
Farakoba	0.2	0.2	0.1	-	0.001*
Kamboinse	0.2	0.2	0.2	0.2	-

CHAPTER 3

GENETIC DIFFERENTIATION AMONG *MARUCA VITRATA* F. (LEPIDOPTERA: CRAMBIDAE) POPULATIONS ON CULTIVATED COWPEA AND WILD HOST PLANTS: IMPLICATIONS FOR INSECT RESISTANCE MANAGEMENT AND BIOLOGICAL CONTROL STRATEGIES[3]

ABSTRACT

Maruca vitrata Fabricius (Lepidoptera: Crambidae) is a polyphagous insect pest that feeds on a variety of leguminous plants in the tropics and subtropics. The contribution of host-associated genetic variation on population structure was investigated using analysis of mitochondrial cytochrome oxidase 1 (*cox*1) sequence and microsatellite marker data from *M. vitrata* collected from cultivated cowpea (*Vigna unguiculata* L. Walp.), and alternative host plants *Pueraria phaseoloides* (Roxb.) Benth. var. *javanica* (Benth.) Baker, *Loncocarpus sericeus* (Poir), and *Tephrosia candida* (Roxb.). Analyses of microsatellite data revealed a significant global F_{ST} estimate of 0.05 ($P \leq 0.001$). The program STRUCTURE estimated 2 genotypic clusters (co-ancestries) on the four host plants across 3 geographic locations, but little geographic variation was predicted among genotypes from different geographic locations using analysis of molecular variance (AMOVA; among group variation -0.68%) or *F*-statistics (F_{ST}^{Loc} = -0.01; P = 0.62). These

[3] This chapter has been published as Agunbiade et al. 2014. Genetic differentiation among *Maruca vitrata* F. (Lepidoptera: Crambidae) populations on cultivated cowpea and wild host plants: Implications for insect resistance management and biological control strategies. *PLoS ONE* 9(3): e92072. It is an open access article and is available at www.plosone.org and DOI:10.1371/journal.pone.0092072. I acknowledge the contribution of co-authors to the publication.

results were corroborated by mitochondrial haplotype data (φ_{ST}^{Loc} = 0.05; P = 0.92). In contrast, genotypes obtained from different host plants showed low but significant levels of genetic variation (F_{ST}^{Host} = 0.04; P =0.01), which accounted for 4.08% of the total genetic variation, but was not congruent with mitochondrial haplotype analyses (φ_{ST}^{Host} = 0.06; P = 0.27). Variation among host plants at a location and host plants among locations showed no consistent evidence for *M. vitrata* population subdivision. These results suggest that host plants do not significantly influence the genetic structure of *M. vitrata*, and this has implications for biocontrol agent releases as well as insecticide resistance management (IRM) for *M. vitrata* in West Africa.

INTRODUCTION

Host plant adaptation by herbivorous insects has resulted in monophagous species that are highly specialized on a single host, whereas polyphagous insect species have evolved to feed upon a wide array of different host plants (e.g., Futuyma 1983; Jaenike 1990; Mitter and Farrell 1991). Host plants may have a major role in the differentiation and diversification of herbivorous insects, and are important in our current understanding of global biodiversity and niche exploitation by insect populations (Ehrlich and Raven 1964; Strong et al. 1984; Farrell 1998). The diversity of ecosystems, which polyphagous species encounter, makes the study of genetic variation based on host plants important for the understanding of adaptation and niche formation. Within a single species, genetic variation can arise among subpopulations that utilize different host plants through variation in oviposition or feeding preferences, rates of development on different host plants, as well as subsequent survivorship, fecundity and mating preferences of adults

(Funk et al. 2002). Mating barriers and reduced gene flow have been predicted among individuals from insect species that show adaptation to different host plants (Nason et al. 2002; Sword et al. 2005), and resulted in assortative mating within populations (Feder et al. 1998). In addition to broader implications in species formation, assortative mating based on host plant preference can impact the practical application of insect pest management strategies, such as the release of biocontrol agents and the implementation of insect resistance management (IRM) strategies based on genetically modified crops.

The legume pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae) is a polyphagous insect pest of grain legumes that has a wide distribution throughout tropical and subtropical regions worldwide. Feeding damage caused by larval *M. vitrata* to cowpea crops occurs on flower buds, flowers and seed pods. This insect species develops without diapause and uses multiple alternative host plants during the dry season in West Africa when cowpea crops are not in cultivation (Taylor 1978; Bottenberg et al. 1997; Arodokoun et al. 2003). Larval *M. vitrata* feeding has been documented on over 50 alternative host plants (Taylor 1978; Arodokoun et al. 2003; Sharma 1998), and most often found on cultivated and wild host plants from the family, Fabaceae (Leumann 1994; Arodokoun 1996). *Pterocarpus santalinoides* L'Hér. ex DC., *P. phaseoloides* and *Centrosema pubescens* (except cv. Belalto) are used for oviposition and subsequent larval development during the long dry season, whereas *Lonchocarpus sericeus* and *L. cyanescens* (Schumach and Thonn.) Benth. are similarly used during the main rainy season, and *Tephrosia platycarpa* Guill. and Perr. during the short rainy season (Arodokoun et al. 2003). The reservoirs of *M. vitrata* maintained on alternative host plants results in difficulties for cultural and chemical insecticide control. As proposed by Tamò

et al. (2002), the possible Asian origin of *M. vitrata* may contribute to the lack of corresponding native natural enemies capable of regulating its populations in those alternative host plant habitats in West Africa, and thus might also lead to heavy infestations observed on cowpea crops. Efforts to introduce biological control candidate species have had limited success, and yet unrecognized biotic factors such as *M. vitrata* alternative host plant differentiation, could hinder the effective spread of introduced control agents (Tamò et al. 2012).

Protein crystalline (Cry) toxins produced by the gram-positive soil bacterium *Bacillus thuringiensis* (*Bt*) show insecticidal activities against many Lepidopteran insects. Transgenic cowpea that express the *Bt* toxin Cry1Ab are being developed for the protection of this crop for use in West African cropping systems (Huesing et al. 2011). Although transgenic *Bt*-cowpea offers a promising approach to crop improvement, sustainability of the technology will likely depend on the mitigation of resistance development in *M. vitrata* populations and availability of suitable alternative host plants to act as refuges. Specifically, the high-dose refuge model is the most widely accepted IRM strategy (Alstad and Andow 1995), and has been implemented as an effective resistance management plan to delay the development of resistance to *Bt* in target pest insect populations (Gould 1998). The high-dose component of this IRM strategy requires that crops express levels of *Bt* toxin sufficient to kill 100% of homozygous susceptible and heterozygous larvae. Refuges are non-*Bt* plants in proximity to *Bt* crops on which the targeted pests can also complete development (Gould 1998). In theory, refuge plants are able to produce a large population of adults that will mate randomly with any potential homozygous resistant individual that might complete development on a *Bt* crop plant. By

shear stochastic sampling, rare homozygous resistant individuals that emerge from *Bt* fields are most likely to mate with a refuge plant-derived homozygous susceptible individual. This increases the probability that any resistant insects emerging from the *Bt* crops are more likely to mate with a susceptible adult emerging from the refuges, thereby generating heterozygous progeny that are not capable of surviving exposure to the high dose of *Bt* toxin expressed by transgenic crop in order to delay or prevent an increase in resistance allele frequency within target insect populations (Bourguet et al. 2000). Wild-growing alternative host plants can also serve as natural refuges for target pests, and have been reported as effective refuges for IRM of transgenic crops (Zhang and Tang 2000; Tan et al. 2001; Wu et al. 2004; Abney et al. 2007; Jackson et al. 2008). In the case of *M. vitrata*, there are several alternative host plants which are available throughout the cowpea growing season and which might act as natural refuges. Assessing the suitability of alternative hosts as effective refuge plants for *Bt*-cowpea will be important for developing IRM programs for *M. vitrata* in West Africa. However, it is not clear when *Bt*-cowpea will be used broadly in West Africa, which highlights the need to enhance the efficacy of current pest control solutions.

The control of *M. vitrata* in West Africa currently relies on the use of cultural and chemical control methods and increasingly on the use of biological control agents. Alternative host plant use and any potential genetic differentiation among populations based on this biological phenomenon may also impact how biocontrol agents are deployed (Olivieri et al. 2004). The lack of alternative hosts may be a contributing factor in the observation that, although many biological control introductions result in establishment, most are unsuccessful in reducing pest densities (Gurr and Wratten 1999). Therefore, most

managers of agricultural systems seek to manipulate habitat complexity to encourage the conservation and enhancement of natural enemies in the hopes of improving pest suppression (see reviews by Wratten and van Emden 1995; Gurr et al. 2005; Landis et al. 2000). A key factor that enhances predator and parasitoid populations in complex landscapes is the availability of nectar and pollen subsidies. Many natural enemies, particularly Hymenopteran parasitoids, lacewings, syrphid flies, and tachinid flies are herbivorous as adults and require carbohydrates for successful reproduction. A literature review by Altieri and Letourneau (1982) showed that the successful establishment of certain parasitoids in cropping systems depends on the presence of weeds that provide nectar for the adult female wasps. Laboratory and field studies have also demonstrated positive impacts on parasitoid fecundity, lifespan, or searching efficiency as a result of floral resources in bordering non-crop areas (Lavandero et al. 2006; Gourdine et al. 2005; Lee and Heimpel 2008; Bianchi and Wäckers 2008). However, although alternative host plants have been reported to enhance parasitoid and predator efficiency in conservation biological control strategies, extensive population-level data are still needed for deployment of biocontrol agents to be effective. The application of population genetic data to biological control of *M. vitrata* will provide better information on how many distinct genotypes exist on the different host plants and the effect this can have on the parasitoid population over time. The use of population structure data will therefore enable the identification of the genetic differentiation of *M. vitrata* on cultivated cowpea and available alternative host plants and the effective host plants that can be planted alongside the cultivated cowpea in order to maximize parasitoid efficiency.

Genetic variation among *M. vitrata* larvae on four host plants including cowpea in West Africa was assessed using haplotype sequencing of the mitochondrial cytochrome c oxidase-1 gene (*cox*1) fragment, as well as genotyping using a set of microsatellite markers previously developed by Agunbiade et al. (2012). Levels of genetic and haplotype variation, population structure, and gene flow were estimated among *M. vitrata* collected from different host plants in southern regions of Benin. The results of this research are important for assessing the effectiveness of alternative host plants for use as a refuge for *Bt*-cowpea crops, and to potentially identify the most appropriate host plant to apply biocontrol agents. These data will be used to enhance ongoing efforts to reduce the impact of *M. vitrata* feeding damage and to improve yields in cowpea cropping systems of West Africa.

MATERIALS AND METHODS

Insect sampling and DNA extraction

Larval *M. vitrata* were collected from cultivated cowpea (*Vigna unguiculata*), and three alternative host plants – *P. phaseoloides* (dry season host), *T. candida* (short rainy season host), and *L. sericeus* (main rainy season host), in three divisions representing 6 departments in Southern Benin in 2012 (Figure 3.1). The divisions were Mono-Couffo, Zou-Collines and Ouémé-Plateau. Within each division, we collected from different locations to lessen the possibility that the same female individual laid larvae collected. Forty-nine, 50 and 49 individual *M. vitrata* samples were collected from *V. unguiculata* in Ouémé-Plateau, Zou-Collines, and Mono-Couffo, respectively. Forty-seven and 45 individual *M. vitrata* samples were collected from *L. sericeus* in Ouémé-Plateau, and Zou-

Collines, respectively. Fifty-two, 52 and 58 individual *M. vitrata* samples were collected from *T. candida* in Ouémé-Plateau, Zou-Collines, and Mono-Couffo, respectively, and 49, 49 and 48 individual *M. vitrata* samples were collected from *P. phaseoloides* in Ouémé-Plateau, Zou-Collines, and Mono-Couffo, respectively. Genomic DNA was extracted from the insect samples using DNeasy animal tissue kit and following manufacturer instructions (Qiagen, Valencia, CA). The DNA concentrations were adjusted to 10ng/µl and used for genotyping.

Microsatellite genotypes

Microsatellite markers C0241, 7_02K06, C0444, C32008 and 01_B12 were used for genotyping *M. vitrata* samples (Table 3.1), amplified in multiplex PCR reactions and detected as previously described by Agunbiade et al. (2012). The microsatellite markers were obtained as previously described in Agunbiade et al. (2012) and the DNA sequence libraries submitted to GenBank under the accession numbers from JN685509 to JN685580. The mean number of alleles per locus, observed heterozygosity and expected heterozygosity were calculated for genotypes by location and by host plant within each location using Arlequin 3.5.1.3 (Excoffier et al. 2005). The potential occurrence of null alleles and other genotyping errors (stuttering and allele drop out) were tested using the program Micro-Checker 2.2.3 (Van Oosterhout et al. 2004), and null alleles were suspected at a given locus when Micro-Checker rejects Hardy-Weinberg Equilibrium (HWE) and excess homozygosity was evenly distributed among allelic size classes. Null allele-corrected pairwise F_{ST} estimates were calculated for all populations by applying the ENA correction in the FREENA package (Chapuis and Estoup 2007; Chapuis et al. 2008;

available at http://www1.montpellier.inra.fr/URLB/). Uncorrected F_{ST} values were estimated following Weir (1996), whereas corrected F_{ST} estimates were made when null allele were predicted following the expectation maximization (EM) algorithm (Dempster et al. 1977).

Analysis of molecular variance (AMOVA), global *F*-statistics (Weir and Cockerham 1984) and pairwise F_{ST} estimates were calculated also using Arlequin 3.5.1.3 (Excoffier et al. 2005). Four different analyses were performed based on assumed partitioning of the population based on host plant and/or geographic location; *analysis 1*: variation among host plants (pooled across all locations), *analysis 2*: variation among geographic locations (pooled for all host plants), *analysis 3*: differentiation between host plant within each geographic location, and *analysis 4*: differentiation between geographic location for each host plant group. Significance for each comparison was corrected for Type I error by application of the B-Y method (Benjamini and Yekutieli 2001).

The program STRUCTURE 2.3.4 uses a model-based clustering to predict population structure using genotypic marker data from individual samples, where the model assigns proportions of individual genotypes to one of *K* populations (Pritchard et al. 2000). STRUCTURE analysis of microsatellite genotype data was run using an initial burn-in of 100,000 iterations followed by 100,000 iterations, and ten replicates with each potential value of *K (*range 1 to 10) were run with an assumed population admixture model. STRUCTURE runs were performed using LOCPRIOR command, where genotypes were defined based on host plant at each geographic location. The "real" value of *K* (number of potential unique populations represented by the *M. vitrata* genotypes) was estimated as described by Evanno et al. (2005) using the program Structure harvester

(Earl and vonHoldt 2012; available at http://taylor0.biology.ucla.edu/structureHarvester/). A graphical display of individual co-ancestry (*Q*-matrix) data was generated from STRUCTURE output using the program Distruct (Rosenberg 2004).

Isolation by distance (IBD) model of genetic differentiation was tested by comparing F_{ST} (1 - F_{ST}) with the logarithm of geographic distances, and significance evaluated using Mantel tests with 10,000 randomizations of the data. All IBD analyses were conducted using the IBDWS (Jensen et al. 2005; available at http://ibdws.sdsu.edu/~ibdws/).

Mitochondrial haplotypes

Oligonucleotide primers HC02198 5′-TAA ACT TCA GGG TGA CCA AAA AAT CA-3′ and LCO1490 5′-GGT CAA CAA ATC ATA AAG ATA TTG G-3′ (Folmer et al. 1994) were used for PCR amplification of ~ 650 bp mitochondrial cytochrome *c* oxidase I (*cox*1) DNA barcode region. All PCR, *Sac*I PCR-RFLP and DNA sequencing reactions were performed according to Margam et al. (2011a), except cycle sequencing using BigDye™ reactions (Applied Biosystems, Foster City, CA), which were performed at the Iowa State University DNA Sequencing and Synthesis Facility, Ames, IA where data was trimmed for Phred scores < 20. The haplotype data were submitted to GenBank under the accession numbers from KJ175700 to KJ176247.

DNA sequence data were aligned for each individual using CLUSTALX 1.8 (Thompson et al. 1997). Haplotype differentiation of sequence data was estimated among 1) host plant or 2) geographic location from which samples were collected using φ-statistics, which is an approximation of *F*-statistics, based on haplotype frequencies

(Benjamini and Yekutieli 2001; Excoffier et al. 1992). The φ-statistics and AMOVA estimates were obtained using Arlequin as described previously, except the Kimura 2-parameter model was used for φ-statistic calculation with an empirical estimated gamma parameter = 0.05. AMOVA was used to partition haplotype variance between 1) host plants across geographic locations (sample sites) or 2) geographic location across different host plants. Pairwise φ_{ST} estimates were made between host plant groups using Arlequin and significance for multiple tests within each comparison determined following application of the B-Y method (Benjamini and Yekutieli 2001) as described above.

RESULTS

Microsatellite genotypes

The observed heterozygosity (H_O) across all loci ranged from 0.02 to 0.89 while the expected heterozygosity ranged from 0.02 to 0.67. Nineteen of the 55 exact tests across host plants and geographic locations showed significant deviation from HWE. Two of the markers were monomorphic (CO241 on *L. sericeus* at Zou-Collines and 7_02K026 on *P. phaseoloides* at Ouémé-Plateau) (Appendix B). Micro-Checker analysis indicated that markers 7_02K06 and C0241 showed evidence of null allele presence in all populations that were tested. There was no evidence of stuttering or allele drop out in any of the microsatellite markers. Results of genetic differentiation estimates among *M. vitrata* were based on four sets of analyses; *analysis 1*: variation among host plants (pooled across all locations), *analysis 2*: variation among geographic locations (pooled for all host plants), *analysis 3*: differentiation between host plants within each geographic location, and *analysis 4*: differentiation between geographic locations for each host plant group.

Analysis 1: When microsatellite genotypes were divided into four groups based on the host plants from which *M. vitrata* larvae were collected, the global estimates of subpopulation differentiation across all loci were low but significant based on uncorrected (F_{ST} = 0.06) and ENA-corrected microsatellite genotype data (F_{ST}^{ENA} = 0.05; Table 3.2). AMOVA results indicated that 93.03% of the genetic variation for *M. vitrata* was within host plant group while 5.71% was estimated among host plants (remaining data not shown). Pairwise F_{ST} estimates of host plant differentiation based on uncorrected and ENA-corrected microsatellite data across all loci ranged from 0.01 to 0.09 (Table 3.3), and indicated that all comparisons were significant. *Analysis 2*: In comparison, microsatellite genotypes based on geographic location resulted in uncorrected F_{ST} estimates of 0.02 (F_{ST}^{ENA} = 0.02; Table 3.2). Additionally, Mantel tests showed an absence of IBD through no detectable correlation between genetic and geographic distances (R^2 = -0.12 P = 0.49; remaining results not shown). *Analysis 3*: Analysis of host plant variation within a single location effectively removed a potential confounding influence of geographic variation on host plant differentiation. Subsequent pairwise F_{ST} estimates ranged from -0.01 to 0.28, and significant differentiation was predicted for 11 of 15 comparisons at B-Y Method adjusted significant thresholds (Table 3.4). Analogously, *Analysis 4* evaluated variation between geographic locations for *M. vitrata* collected from the same host plant, which predicted significant differentiation in 5 of 10 comparisons at B-Y adjusted significant thresholds (Table 3.5).

STRUCTURE analysis indicated that there were 2 populations among all the samples on the different host plants and across locations (Figure 3.2). A maximum value of 7.41 was generated for mL"(K)/sL(K) at K = 2, which represented the “real” population

number (*K*) that STRUCTURE predicted from microsatellite dataset. The estimated co-ancestries were partitioned into these two distinct clusters among the *M. vitrata* microsatellite genotypes, and were partitioned among host plant groups from 3 geographic locations (Figure 3.2). Cluster 1 (orange) was proportionately most common among *M. vitrata* samples from *V. unguiculata* at Ouémé-Plateau and Mono-Couffo, Benin as well as from *T. candida* at Mono-Couffo, Benin.

Mitochondrial haplotypes

The mitochondrial *cox*1 gene fragment that was PCR amplified in this study was also previously used to investigate haplotype variation among *M. vitrata* in West Africa by Margam et al. (2011a). Alignment of novel *cox*1 sequence data from 548 individuals collected from 4 different host plants at 3 different geographic locations resulted in a 619 bp consensus sequence which showed a mean nucleotide diversity of 0.0019±0.0014 (mean number of pairwise sequence differences 1.17±0.76). Results of AMOVA showed that 94.15% of the haplotype variation was within populations based on host plant from which larvae were collected, whereas 4.80% of the variation was among populations (remaining data not shown). A global estimate of haplotype differentiation among host plant groups was also low (φ_{ST} = 0.05) but significant ($P < 0.001$). Pairwise φ_{ST} estimates which was analogous to analyses 1 to 4 used for microsatellite data (see previous section), ranged from -0.01 to 0.20 (Table 3.6), and showed significant differentiation for 11 of 55 comparisons at the B-Y adjusted significance threshold of 0.01 (Appendix B). For example, these results showed significant variation between φ_{ST} estimates between *T. candida* and both *V. unguiculata* and *L. sericeus* at Ouémé-Plateau, Benin. Also, *M.*

vitrata collected from *V. unguiculata* at Ouémé-Plateau, Zou-Collines and Mono-Couffo, Benin showed no significant mitochondrial haplotype variation ($P \geq 0.148$), but *M. vitrata* collected from *T. candida* showed significant variation between all 3 geographic locations ($P \leq 0.002$).

DISCUSSION

Microsatellite markers developed from species of Lepidoptera can have high frequencies of non-PCR amplifying "null" alleles that potentially result in the overestimation of homozygosity, and have been reported in population genetic studies from a range of taxa (Dakin and Avise 2004; DeWoody et al. 2006). Microsatellite markers from Lepidopteran insects and molluscs have been reported to have particularly high frequencies of null alleles (review in Chapuis and Estoup 2007). Associations between null alleles and highly variable flanking regions have been repeatedly demonstrated (see Chapuis and Estoup 2007). Recent evidence suggests that null alleles at some microsatellite loci may be affected by movement of transposable elements (Coates et al. 2010). Indeed, two of the microsatellite loci (7_02K06 and C0241) showed the presence of null alleles, but the molecular basis for the non-PCR amplification of alleles was not investigated. Regardless of the cause, resulting F_{ST} estimates from this study were corrected using the ENA algorithm, which has previously been shown to allow for accurate analysis of population genetic microsatellite data. Both ENA-corrected as well as uncorrected F_{ST} estimates from microsatellite data analyses provided congruent results that suggested significant levels of genetic variation exist between *M. vitrata* collected

from the different host plants, but this variation is not consistently present among comparisons at different geographic locations.

Larval *M. vitrata* are a major pest of cultivated cowpea, *V. unguiculata*, in the tropics and subtropics, and are difficult to control through applications of chemical insecticides because sprays cannot contact larvae that have burrowed into the flowers and pods. The development and implementation of cowpea that expresses the *Bt* Cry1Ab toxin holds the promise to effectively control *M. vitrata* feeding damage, but the evolution of resistance in several species of Lepidoptera to *Bt* toxin has also raised concerns regarding the longevity of this technology (Murdock 2002). Prior to release of cowpea varieties to farmers in West Africa, an understanding of the biology, ecology, and population structure is fundamental in making sound and effective IRM decisions, which may prolong the field efficacy of this *Bt* technology. Significant levels of genetic differentiation were previously estimated among *M. vitrata* collected from *V. unguiculata* in the West African countries of Niger, Nigeria and Burkina Faso using data from SNPs (Margam et al. 2011b) and microsatellite markers (Agunbiade et al. 2012). Genetic differentiation among *M. vitrata* populations was positively correlated with geographic distance (Margam et al. 2011b). Additionally, mitochondrial haplotypes were previously shown to be differentiated among *M. vitrata* collected from cowpea in the West African nations of Nigeria, Niger and Burkina Faso, with 2 distinct haplotype groups being predicted (Margam et al. 2011a). Winged insects that are capable of long distance flight (reviewed by Showers 1997) are typically genetically homogenous (den Boer 1978; Llewellyn et al. 2003; Wei et al. 2013), where admixture effectively results in a single random mating population that lacks any significant gene flow barriers (Lyons et al. 2012). *Maruca vitrata* persist in southern

coastal repositories during the dry season and undergo a seasonal range expansion as the population migrates to northern regions when climatic conditions become more favorable at the onset of the rainy season (Bottenberg et al. 1997; Ba et al. 2009). This pattern of seasonal migration may cause genetic structuring due to the Wahlund effect or other unknown population genetic factors (Margam et al. 2011b; Agunbiade et al. 2012), but the influence of a number of other potential confounding factors was not previously investigated.

IRM programs for *Bt*-cowpea in West Africa will likely use a high-dose/refuge strategy, where refugia of non-transgenic plants will be essential for maintaining a reservoir of susceptible alleles. The high-dose/refuge strategy is considered central to managing resistance to *Bt* toxins, but the level of gene flow and random mating within and between populations of target insects is also important for the spread of susceptible genotypes in the population (Bourguet et al. 2000; Onstad et al. 2012). Refugia can be comprised of cultivated non-transgenic crop plants or perhaps any other host plants that can support significant population sizes for the targeted insect pest species. Weedy species that are alternative hosts to arthropod pests may also serve as an effective form of refugia. Models based on studies of maize cropping systems suggest that increased habitat diversity, including weedy vegetation, could reduce the rate of spread of rotation-resistant western corn rootworm (Onstad et al. 2003). Studies have also reported that the utilization of wild host plants can be effective refuges within IRM strategies for transgenic crops (Zhang and Tang 2000; Tan et al. 2001; Wu et al. 2004; Abney et al. 2007; Jackson et al. 2008). Although *M. vitrata* are known to feed on multiple non-cowpea plants, the level of gene flow between individuals feeding on cowpea and these other plants remains

unknown, and may affect the efficacy of IRM strategies. Many species of Lepidoptera are polyphagous and are opportunistic insects that feed on multiple alternative host plants, but instances of differential rates of development are proposed to result in reduced gene flow due to temporal variation in adult mating periods, such that assortative or structured mating systems have evolved (Malausa et al. 2005; Calcagno et al. 2007). Breakdown of gene flow between sympatric populations of a species has been hypothesized to cause host race formation (Nason et al. 2002).

Low but significant levels of genetic differentiation was estimated from microsatellite marker and mitochondrial haplotype data between *M. vitrata* collected from cultivated cowpea (*V. unguiculata*) and alternative native host plants *P. phaseoloides*, *L. sericeus* and *T. candida*. Analogous sampling of *M. vitrata* from alternative hosts was not conducted in previous studies by Margam et al. (2011b), and Agunbiade et al. (2012), and provided new insights into possible genetic structure in West Africa. Results of the current study might suggest little host plant-related *M. vitrata* population structure from initial analyses of microsatellite (F_{ST} = 0.05) and haplotype data (φ_{ST} = 0.04). Also in contrast to previous results by Margam et al. (2011b), and Agunbiade et al. (2012), genetic variation in this current study was shown to be low between the 3 collection sites and not correlated with geographic distance. This might be due to our sampling that was restricted to just the southern region of Benin. Additional analyses which potentially removed the confounding influence of geographic variance showed significant pairwise genetic differentiation between *M. vitrata* collected from all of the different host plants at Ouémé-Plateau, but this pattern was not consistent at the Zou-Collines or Mono-Couffo locations. Similar inconsistent results were observed among pairwise comparisons of *M. vitrata* from

different geographic locations but collected from the same host plant. These findings were supported by analysis with the program STRUCTURE, where co-ancestry represented by Cluster 2 (blue) was prevalent among *M. vitrata* collected from all different host plants, with the exception of individuals collected from *V. unguiculata* at Ouémé-Plateau and *T. candida* at Mono-Couffo.

With respect to the high dose-refuge strategy, the apparently weak and inconsistent genetic differentiation of *M. vitrata* on different host plants might suggest that high levels of gene flow would occur between susceptible individuals on wild alternative hosts and rare resistant individuals that survive on *Bt*-cowpea. Although not conclusive, our findings might also suggest that the wild hosts surveyed in this study may serve as effective refuge plants in any eventual implementation of *Bt*-cowpea in West Africa. Lack of consistent host plant differentiation among *M. vitrata* across multiple geographic locations might also suggest that the females have not become "tuned" for oviposition on specific host plants, such that host-races are not likely to have formed. More likely, complex temporal interactions between plant phenologies and attraction of female *M. vitrata* for oviposition may play a role in determining host plant usage and subsequent levels of gene flow at a specific locality in a specific year. Thus variation in local environments could influence oviposition and/or subsequent larval development on host plants, such that random and significant perturbations on genetic distribution might be detected. Alternatively, climatic conditions have been shown to support basal insect population sizes during conditions previously thought to be restrictive (Merret 1986), such that some alternative non-cultivated hosts might harbor reservoirs of *M. vitrata* during the dry season. Sampling of these presumable small reservoir populations in this study might have inadvertently

skewed our estimates of within population differentiation, and could complicate any future population genetic studies where these confounding factors are not taken into account. Regardless, our data might not suggest that random mating will occur between rare resistant moths emerging from *Bt*-cowpea and susceptible moths derived from non *Bt*-cowpea or native host plant refuges. The rate of development among *Bt* resistant individuals has been documented, such that assortative mating might be possible due to temporal delay in emergence of subsequent adults. In such a scenario, the mating period of reproductive adults may show limited overlap and could result in reduced gene flow. Under the assumptions of the high-dose/refuge strategy, temporal delays between adult emergence from *Bt*-cowpea, non *Bt*-cowpea and alternative host plants will affect the probabilities at which the rare resistant individuals mate with susceptible adults, and could lead to the rapid increase in homozygous resistant genotypes within the pest insect population if significant temporal delays are encountered.

The interactions between insect pests, their natural enemies, and the natural vegetation often leads to more efficient biological control, not only because of the increased availability of refugia and alternative prey for natural enemies during off-seasons, but also because of the higher diversity in the natural vegetation (e.g. Altieri et al. 1993; Waage and Hawksworth 1991). Tamò et al. (1997) reported that the availability of alternative host plants positively affects parasitism rates, and should consequently reduce overall pest densities. Because of the semi-migratory habit of *M. vitrata,* Tamò et al. (2003) suggested two different levels from which to consider possible biological control interventions. The first option during the cropping season in cowpea fields, would be the inundative release of locally available, mass-reared trichogrammatids, preferably in

conjunction with the use of pheromone trap-derived thresholds (Downham et al. 2002), particularly in areas where *M. vitrata* does not have suitable alternative host plants during the dry season, but rather invades the cowpea fields like a migrant pest (e.g., coming from the south, as it is the case for the Kano region, see Bottenberg et al. 1997). The second option would be more appropriate in areas where alternative host plants are abundant and constitute a major factor influencing the dynamics of *M. vitrata* populations. In this case, inoculative releases of larval parasitoids such as *Therophilus javanus* or *T. marucae* (Hymenoptera: Braconidae) will be targeting *M. vitrata* populations on those host plants, with the objective of reducing overall pod borer populations at the landscape level. Based on the results obtained in the present study, the second option would seem more appropriate in the introduction and release of biocontrol agents against *M. vitrata*.

REFERENCES

Abney MR, Sorenson CE, Bradley JR Jr (2007) Alternate crop hosts as resistance management refuges for tobacco budworm, *Heliothis virescens* (Lepidoptera: Noctuidae) in North Carolina. *The Journal of Cotton Science* 11: 35-39.

Agunbiade TA, Kim KS, Forgacs D, Margam VM, Murdock LL, Ba NM, Binso-Dabire CL, Baoua I, Ishiyaku MF, Tamò, Pittendrigh BR (2012) The spatial genetic differentiation of the legume pod borer, *Maruca vitrata* F. (Lepidoptera: Crambidae) populations in West Africa. *Bulletin of Entomological Research* 102(5): 589-599.

Alstad DN, Andow DA (1995) Managing the evolution of insect resistance to transgenic plants. *Science* 268: 1894-1896.

Altieri MA, Cure JR, Garcia MA (1993) The role and enhancement of parasitic Hymenoptera biodiversity in agroecosystems. In LaSalle J. and Gauld ID (Eds.), *Hymenoptera and biodiversity*. CAB International, Wallingford, UK, pp 257–275.

Altieri MA, Letourneau DK (1982) Vegetation management and biological control in agroecosystems. *Crop Protection* 1: 405-430.

Arodokoun D (1996) Importance des plantes-hôtes alternatives et des ennemis naturels indigènes dans le contrôle biologique de *Maruca testulalis* Geyer (Lepidoptera: Pyralidae), ravageur de *Vigna unguiculata* Walp. PhD thesis, University of Laval, Québec, Canada, 181 pp.

Arodokoun DY, Tamò M, Cloutier C, Adeoti R (2003) The importance of alternative host plants for the annual cycle of the legume pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Pyralidae). *Insect Science and its Application* 23: 103-113.

Ba NM, Margam VM, Binso-Dabire CL, Sanon A, McNeil JN, Murdock LL, Pittendrigh BR (2009) Seasonal and regional distribution of the cowpea pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae) in Burkina Faso. *International Journal of Tropical Insect Science* 29(3): 109-113.

Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Annals of Statistics* 29: 1165-1188.

Bianchi FJJA, Wäckers FL (2008) Effects of flower attractiveness and nectar availability in field margins on biological control by parasitoids. *Biological Control* 46: 400-408.

Bottenberg H, Tamò M, Arodokoun D, Jackai LEN, Singh BB, Youm O (1997) Population dynamics and migration of cowpea pests in northern Nigeria: Implications for integrated pest management. In Singh BB, Mohan Raj DR, Dashiell KE, Jackai LEN (Eds.), *Advances in cowpea research*. Co-publication of International Institute of Tropical Agriculture (IITA) and Japan International Center for Agricultural Sciences (JIRCAS). IITA, Ibadan, Nigeria, pp 271–284.

Bourguet D, Bethenod MT, Pasteur N, Viard F (2000) Gene flow in European corn borer *Ostrinia nubilalis*: Implications for the sustainability of transgenic insecticidal maize. *Proceedings of the Royal Society of London Series B* 267: 117-122.

Calcagno V, Thomas Y, Bourguet D (2007) Sympatric host races of the European corn borer: Adaptation to host plants and hybrid performance. *Journal of Evolutionary Biology* 20: 1720-1729.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24(3): 621-631.

Chapuis MP, Lecoq M, Michalakis Y, Loiseau A, Sword GA, Piry S, Estoup A (2008) Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Molecular Ecology* 17(16): 3640-3653.

Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2010) A Helitron-like transposon superfamily from Lepidoptera disrupts (GAAA)n microsatellites and is responsible for flanking sequence similarity within a microsatellite family. *Journal of Molecular Evolution* 70: 278-288.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* 93: 504-509.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)* 39(1): 1-38.

den Boer MH (1978). Isozyme and migration in the African armyworm *Spodoptera exempta* (Lepidoptera: Noctuidae). *Journal of Zoology London* 185: 539-553.

DeWoody J, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* 6: 951-957.

Downham MCA, Tamò M, Hall DR, Datinon B, Dahounto D, Farman DI (2002) Development of sex pheromone traps for monitoring the legume pod-borer, *Maruca vitrata* Fabricius (Lepidoptera: Pyralidae). In Fatokun CA, Tarawali SA, Singh BB, Kormawa PM and Tamò M (Eds.), *Challenges and opportunities for enhancing sustainable cowpea production*. Proceedings of the World Cowpea Conference III held at the International Institute of Tropical Agriculture (IITA) Ibadan, Nigeria, pp 124-135.

Earl DA, vonHoldt BM (2012) Structure Harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4 (2): 359-361.

Ehrlich PR, Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution* 18: 586-608.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611-2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1:47-50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

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

Feder JL, Berlocher SH, Opp SB (1998) Sympatric host-race formation and speciation in *Rhagoletis* (Diptera: Tephritidae) a tale of two species for Charles D. In Mopper S, Strauss SY (Eds.), *Genetic structure and local adaptation in natural insect populations.* Chapman & Hall, New York, pp 408 – 441.

Folmer MB, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3(5): 294-299.

Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: Model systems for the comparative study of speciation ecology. *Genetica* 116: 251-267.

Futuyma DJ (1983) Selective factors in the evolution of host choice by phytophagous insects. In S. Ahmad (Ed.), *Herbivorous Insects: Host seeking behaviour and mechanisms*. Academic Press, New York, pp 227-244.

Gould F (1998) Sustainability of transgenic insecticidal cultivars: Integrating pest genetics and ecology. *Annual Review of Entomology* 43: 701-726.

Gourdine JS, Simmons AM, McCutcheon GS, Leibee GL (2005) Floral nectars and honey enhance survival of *Diadegma insulare* (Hymenoptera: Ichneumonidae), a parasitoid of the diamondback moth (Lepidoptera: Plutellidae). *Journal of Entomological Science* 40: 96-99.

Gurr GM, Wratten SD (1999) "Integrated biological control": A proposal for enhancing success in biological control. *International Journal of Pest Management* 45: 81-84.

Gurr GM, Wratten SD, Tylianakis J, Kean J, Keller M (2005) Providing plant foods for natural enemies in farming systems: Balancing practicalities and theory. In Wäckers FL, van Rijn PCJ, Bruin J (Eds.), *Plant-provided food for carnivorous insects: A protective mutualism and its applications*. Cambridge University Press, Cambridge, UK, pp 267-304.

Huesing J, Romeis J, Ellstrand N, Raybould A, Hellmich R, Wolt J, Ehlers J, Dabiré C, Fatokun C, Hokanson K, Ishiyaku MF, Margam V, Obokoh N, Mignouna J, Nangayo F, Ouedraogo J, Pasquet R, Pittendrigh B, Schaal B, Stein J, Tamò M, Murdock L (2011) Regulatory considerations surrounding the deployment of *Bt*-expressing cowpea in Africa: Report of the deliberations of an expert panel. *GM Crops* 2: 211-224.

Jackson RE, Bradley JR, van Duyn J, Leonard BR, Allen KC, Luttrell R, Ruberson J, Adamczyk J, Gore J, Hardee DD, Voth R, Sivasupramaniam S, Mullins JW, Head G (2008) Regional assessment of *Helicoverpa zea* populations on cotton and non-cotton crop hosts. *Entomologia Experimentalis et Applicata* 126: 89-106.

Jaenike J (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics* 21: 243-273.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* 6: 13. v.3.23 http://ibdws.sdsu.edu/ Assessed 2014 Feb 20.

Landis DA, Wratten SD, Gurr GM (2000) Habitat management to conserve natural enemies of arthropod pests in agriculture. *Annual Review of Entomology* 45: 175-201.

Lavandero BI, Wratten SD, Didham RK, Gurr GM (2006) Increasing floral diversity for selective enhancement of biological control agents: A double-edged sword? *Basic and Applied Ecology* 7: 236-243.

Lee JC, Heimpel GE (2008) Effect of floral nectar, water, and feeding frequency on *Cotesia glomerata* longevity. *BioControl* 53: 289-294.

Leumann C (1994) Populations dynamik des Hülsenbohrers *Maruca testulalis* auf der Kuherbse *Vigna unguiculata* und auf *Pterocarpus santalinoides*, einer alternativen Wirtspfl anze, während der Trockenzeit im Süden von Bénin. Diplomarbeit, ETH Zürich, Switzerland, 83 pp.

Llewellyn KS, Loxdale HD, Harrington R, Brookes CP, Clark SJ, Sunnucks P (2003) Migration and genetic structure of the grain aphid (*Sitobion avenae*) in Britain related to climate and clonal fluctuation as revealed using microsatellites. *Molecular Ecology* 12: 21-34.

Lyons JI, Pierce AA, Barribeau SM, Sternberg ED, Mongue AJ, De Roode JC (2012) Lack of genetic differentiation between monarch butterflies with divergent migration destinations. *Molecular Ecology* 21(14): 3433-3444.

Malausa T, Bethenod MT, Bontemps A, Bourguet D, Cornuet JM, Ponsrard S (2005) Assortative mating in sympatric host races of the European corn borer. *Science* 308: 258-260.

Margam VM, Coates BS, Ba MN, Sun W, Binso-Dabire CL, Baoua I, Ishiyaku MF, Shukle JT, Hellmich RL, Covas FG, Ramasamy S, Armstrong J, Pittendrigh BR,

Murdock LL (2011a) Geographic distribution of phylogenetically-distinct legume pod borer, *Maruca vitrata* (Lepidoptera: Pyraloidea: Crambidae). *Molecular Biology Reports* 38(2): 893 – 903.

Margam VM, Coates BS, Bayles DO, Hellmich RL, Agunbiade T, Seufferheld MJ, Sun W, Kroemer JA, Ba MN, Binso-Dabire CL, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011b) Transcriptome sequencing, and rapid development and application of SNP markers for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(7): e21388.

Merret PJ (1986) Evidence of persistent populations of the African armyworm, *Spodoptera exempta* (Walker) (Lepidoptera: Noctuidae), in east-central Tanzania. *Bulletin of Entomological Research* 76: 247-263.

Mitter C, Farrell B (1991) Macroevolutionary aspects of insect- plant relationships. In E. Bemays (Ed.), *Insect-Plant Interactions*, CRC Press, Boca Raton 3: 35-78.

Murdock LL (2002) Bioassay data, *Bt* and *Maruca*, alpha-amylase inhibitor and cowpea weevil; a status report. In *Workshop on the genetic transformation of cowpea*, NGICA, Capri, Italy, 20 pp.

Nason JD, Heard SB, Williams FR (2002) Host-associated genetic differentiation in the goldenrod elliptical-gall moth, *Gnorimoschema gallaesolidaginis* (Lepidoptera: Gelechiidae). *Evolution* 56(7):1475-1488.

Olivieri I, Singer MC, Magalhâes S, Courtiol A, Dubois Y, Carbonell D, Justy F, Beldade P, Parmesan C, Michalakis Y (2008) Genetic, ecological, behavioral and geographic differentiation of populations in a thistle weevil: Implications for speciation and biocontrol. *Evolutionary Applications* 1:112-128.

Onstad DW, Crowder DW, Mitchell PD, Guse CA, Spencer JL, Levine E, Gray ME (2003) Economics versus alleles: Balancing IPM and IRM for rotation resistant western corn rootworm (Coleoptera: Chrysomelidae). *Journal of Economic Entomology* 96: 1872-1885.

Onstad DW, Kang J, Ba NM, Tamò M, Jackai L, Dabire C, Pittendrigh BR (2012) Modeling evolution of resistance by *Maruca vitrata* (Lepidoptera: Crambidae) to transgenic insecticidal cowpea in Africa. *Environmental Entomology* 41(5): 1045-1276.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Rosenberg NA (2004) Distruct: A program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.

Sharma HC (1998) Bionomics, host plant resistance, and management of the legume pod borer, *Maruca vitrata*: A review. *Crop Protection* 17: 373-386.

Showers WB (1997) Migratory ecology of the black cutworm. *Annual Review of Entomology* 42: 393-425.

Strong DR, Lawton JH, Southwood R (1984). Insects on plants: Community patterns and mechanisms. Blackwell Scientific. Oxford. England.

Sword GA, Joern A, Senior LB (2005) Host plant-associated genetic differentiation in the snakeweed grasshopper, *Hesperotettix viridis* (Orthoptera: Acrididae). *Molecular Ecology* 14(7): 2197-2205.

Tamò M, Arodokoun DY, Zenz N, Agboton C, Adeoti R (2002) The importance of alternative host plants for the biological control of two main cowpea insect pests, the pod borer, *Maruca vitrata* (Fabricius) and the flower thrips, *Megalurothrips sjostedti* (Trybom). In Fatokun CA, Tarawali SA, Singh BB, Kormawa PM, Tamò M (Eds.), *Challenges and Opportunities for Enhancing Sustainable Cowpea Production*. Proceedings of the 3rd World Cowpea Conference, 5–10 September 2000, Ibadan, Nigeria, pp 81–93.

Tamò M, Bottenberg H, Arodokoun DY, Adeoti R (1997) The feasibility of classical biological control of two major cowpea insect pests. In Singh BB, Mohan Raj DR, Dashiell KE, and Jackai LEN (Eds.), *Advances in cowpea research*. Co-publication of International Institute of Tropical Agriculture (IITA) and Japan International Center for Agricultural Sciences (JIRCAS). IITA Ibadan, Nigeria, pp 259-270.

Tamò M, Ekesi S, Maniania NK, Cherry A (2003) Biological control, a non-obvious component of IPM for cowpea. In Neuenschwander P, Borgemeister C, Langewald J (Eds.), *Biological control in IPM systems in Africa*. CAB International, Wallingford, Oxon, pp 256–309.

Tamò M, Srinivasan R, Dannon E, Agboton C, Datinon B, Dabire C, Baoua I, Ba M, Haruna B, Pittendrigh BR (2012) Biological control: A major component for the long-term cowpea pest management strategy. In Boukar O, Coulibaly O, Fatokun C, Lopez K, Tamò M (Eds.), *Improving livelihoods in the cowpea value chain through advancements in science*. Proceedings of the 5th World Cowpea Research Conference, pp 249-259.

Tan S, Chen X, Li D, Zhang H (2001) Can other host species of cotton bollworm be non-*Bt* refuges to prolong the effectiveness of *Bt*-cotton? *Chinese Science Bulletin* 46: 1804-1807.

Taylor TA (1978) *Maruca testulalis*, an important pest of tropical grain legumes. In Singh SR, Emden HF, Taylor TA (Eds.), *Pests of grain legumes: Ecology and control.* London, Academic pp 193–200.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Research* 25: 4876-4882.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4(3): 535-538.

Waage JK, Hawksworth DL (1991) Biodiversity of microorganisms and invertebrates: Its role in sustainable agriculture. In *Proceedings of the first workshop on the ecological foundations of sustainable agriculture (WEFSA 1),* CAB International, Wallingford, UK, pp 149–162.

Wei S-J, Shi B-C, Gong Y-J, Jin G-H, Chen X-X, Meng X-F (2013) Genetic structure and demographic history reveal migration of the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae) from the southern to northern regions of China. *PLoS ONE* 8(4): e59654.

Weir BS (1996) Genetic Data Analysis II. Methods for discreet population genetic data Sinauer Associates, Inc, Sunderland, MA.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Wratten SD, van Emden HF (1995) Habitat management for enhanced activity of natural enemies of insect pests. In Glen DM, Greaves MP, Anderson HM (Eds.), *Ecology and integrated farming systems*. Wiley, Chichester, UK, pp 117-145.

Wu K, Feng H, Guo Y (2004) Evaluation of maize as a refuge for management of resistance to *Bt* cotton by *Helicoverpa armigera* (Hubner) in the Yellow River cotton-farming region of China. *Crop Protection* 23: 523-530.

Zhang TZ, Tang CM (2000) Utilization and management of *Bt* transgenic cotton plants and their insecticidal activity. *Chinese Science Bulletin* 45(2): 119-127.

FIGURES AND TABLES

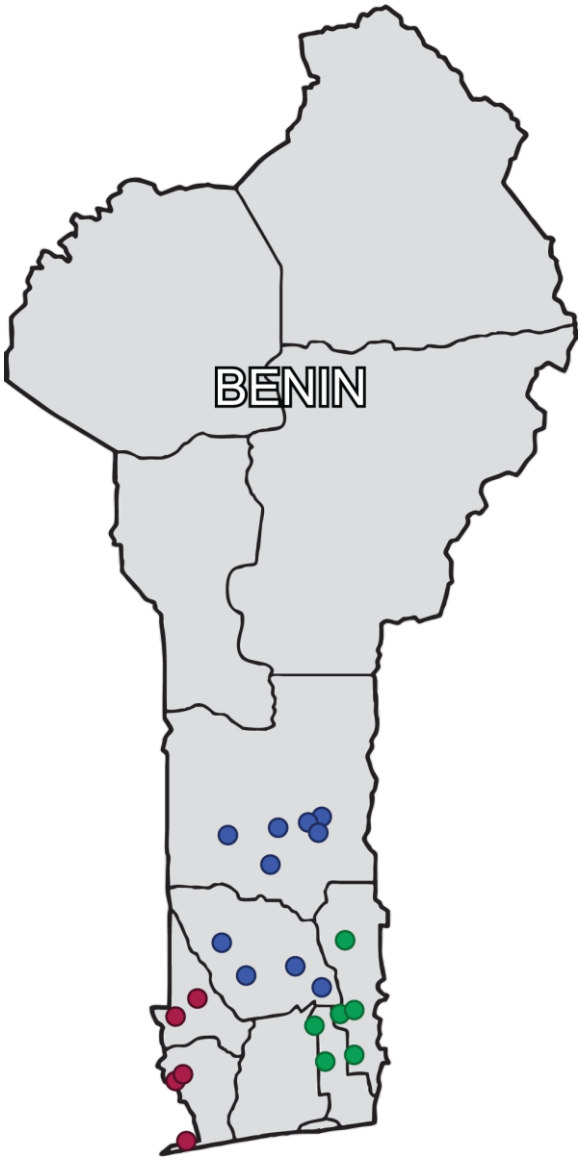


Figure 3.1. Map showing collection sites in southern Benin (red circles – Mono-Couffo, blue circles – Zou-Collines, and green circles – Ouémé-Plateau).

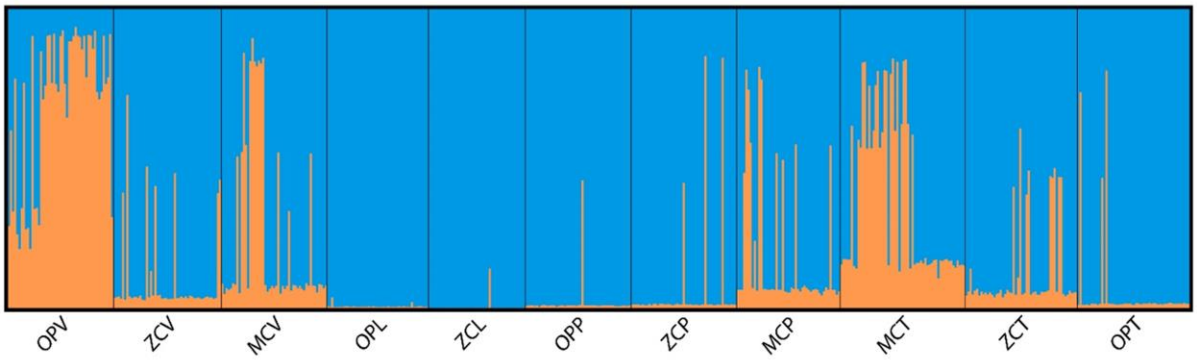


Figure 3.2. Partitioned co-ancestries among microsatellite-defined *M. vitrata* genotypes generated using the program STURUTURE with the LOCPRIOR command. For each, the estimated co-ancestry was derived from the Q-matrix for each individual and represented as vertical lines showing the proportion of the *K* = 2 segments that made up the individual genotype. Genotypes identified from the host plants *V. unguiculata*, *L. sericeus*, *P. phaseoloides* and *T. candida* across the locations are defined [OPV – Ouémé-Plateau (*V. unguiculata*), ZCV – Zou-Collines (*V. unguiculata*), MCV – Mono-Couffo (*V. unguiculata*), OPL – Ouémé-Plateau (*L. sericeus*), ZCL – Zou- Collines (*L. sericeus*), OPP – Ouémé-Plateau (*P. phaseoloides*), ZCP – Zou-Collines (*P. phaseoloides*), MCP – Mono-Couffo (*P. phaseoloides*), MCT – Mono-Couffo (*T. candida*), ZCT – Zou-Collines (*T. candida*) and OPT – Ouémé-Plateau (*T. candida*)].

Table 3.1. *M. vitrata* primer sequences used for microsatellite amplification reactions.

Locus	Primer (dye label) and sequence (5'-3')	Repeat	Size (bp)
C32008 [E]	F-(MAX)AAAAAGCGCTTATATGTTTGTTATAGT	$(CATA)_3$	163
	R-GAAATTTTTAACGGAGATACAATCA		
7_02K06 [A]	F-(FAM)ATTTGTCAGAATGGTATCTTACGT	$(GAT)_6$	151
	R-CCTCTGGGTCATAATTATATTGTTCA		
C0444 [E, 1]	F-(FAM)AAAGGAACTACGCCGTCAGG	$(CAA)_8$	102
	R-GTTGAGCGATCTTGGCACAG		
C0241 [E]	F-(TAM)GACGAAACAAGGCCTACCAG	$(GAT)_9$	165
	R-GGTACTTCYGACGTTGTTCG		
01_B12	F--(TAM)CGGGATGTTACATATACCCAGCA	$(CA)_{12}$	119
	R-CGTACCAATTCATTGAGACTCTCTT		

E, EST-derived primer pair; A, anonymous genomic sequence-derived primer pair; 1, PCR multiplexed primers.

Table 3.2. Global and locus-by-locus estimates of subpopulation differentiation using uncorrected (F_{ST}) and ENA-corrected microsatellite genotype data (F_{ST}^{ENA}) between four host plant groups (*V. unguiculata*, *P. phaseoloides*, *L. sericeus*, and *T. candida*) or geographic location in Benin (Ouémé-Plateau, Zou-Collines and Mono-Couffo).

Locus	Host plant groups		Geographic location	
	F_{ST}	F_{ST}^{ENA}	F_{ST}	F_{ST}^{ENA}
Global	0.056	0.054	0.016	0.024
C0241	0.003	0.013	0.002	0.018
7_02K06	0.109	0.111	0.055	0.077
01_B12	0.123	0.111	0.012	0.019
C32008	0.011	0.011	0.006	0.005
C0444	-0.001	0.002	-0.002	0.002

Table 3.3. Pairwise estimates of subpopulation differentiation across all microsatellite loci with and without ENA-correction (F_{ST}) (below diagonal) and significance of corresponding comparisons (*P*-values) as indicated above the diagonal.

	V. unguiculata	***L. sericeus***	***T. candida***	***P. phaseoloides***	***V. unguiculata***	***L. sericeus***	***T. candida***	***P. phaseoloides***
	Uncorrected				**Corrected**			
V. unguiculata	-	0.001*	0.010*	<0.001*	-	0.001*	<0.001*	<0.001*
L. sericeus	0.09	-	<0.001*	<0.001*	0.13	-	<0.001*	<0.001*
T. candida	0.02	0.03	-	0.010*	0.01	0.09	-	0.010*
P. phaseoloides	0.04	0.03	0.01	-	0.04	0.13	0.01	-

Table 3.4. Estimates of *M. vitrata* subpopulation differentiation from pairwise F_{ST} between host plant groups at each geographic location (below diagonal) and significance of corresponding comparisons (*P*-values) as indicated above the diagonal.

a) Differentiation between host plant group within Oueme-Plateau, Benin (B-Y corrected $\alpha = 0.020$)

	V. unguiculata	***L sericeus***	***T. candida***	***P. phaseoloides***
V. unguiculata	–	<0.001*	<0.001*	<0.001*
L. sericeus	0.277	–	<0.001*	<0.001*
T. candida	0.196	0.063	–	<0.003*
P. phaseoloides	0.22	0.172	0.028	–

b) Differentiation between host plant group within Zou-Collines, Benin (B-Y corrected $\alpha = 0.020$)

	V. unguiculata	***L sericeus***	***T. candida***	***P. phaseoloides***
V. unguiculata	–	0.85	0.285	0.082*
L. sericeus	0.074	–	<0.001*	<0.001*
T. candida	-0.005	0.056	–	0.301
P. phaseoloides	0.001	0.1	0.001	–

c) Differentiation between host plant group within Mono-Couffo, Benin (B-Y corrected $\alpha = 0.027$)

	V. unguiculata	***L sericeus***	***T. candida***	***P. phaseoloides***
V. unguiculata	–	NA	0.010*	0.017*
L. sericeus	NA	–	NA	NA
T. candida	0.011	NA	–	0.055
P. phaseoloides	0.005	NA	0.013	–

Table 3.5. Estimates of *M. vitrata* subpopulation differentiation from pairwise F_{ST} between locations from the same host plant (below diagonal) and significance of corresponding comparisons (*P*-values) as indicated above the diagonal.

a) Differentiation between geographic location for *V. unguiculata* (B-Y corrected α = 0.027)

	Oueme-Plateau	**Zou-Collines**	**Mono-Couffo**
Oueme-Plateau	–	<0.001*	<0.001*
Zou-Collines	0.157	–	0.082
Mono-Couffo	0.084	0.01	–

b) Differentiation between geographic location for *L. sericeus* (B-Y corrected α = 0.05)

	Oueme-Plateau	**Zou-Collines**	**Mono-Couffo**
Oueme-Plateau	–	0.608	NA
Zou-Collines	-0.006	–	NA
Mono-Couffo	NA	NA	–

c) Differentiation between geographic location for *T. candida* (B-Y corrected α = 0.027)

	Oueme-Plateau	**Zou-Collines**	**Mono-Couffo**
Oueme-Plateau	–	0.604	<0.001*
Zou-Collines	-0.003	–	<0.001*
Mono-Couffo	0.072	0.046	–

d) Differentiation between geographic location for *P. phaseoloides* (B-Y corrected α = 0.027)

	Oueme-Plateau	**Zou-Collines**	**Mono-Couffo**
Oueme-Plateau	–	0.141	0.004*
Zou-Collines	0.005	–	0.219
Mono-Couffo	0.023	0.004	–

Table 3.6. Pairwise estimates of mitochondrial *cox*1 haplotype differentiation among *M. vitrata* collected from different host plants (φ_{ST}) (below diagonal) and significance of corresponding comparisons (*P*-values) as indicated above the diagonal. Significance determined at a B-Y adjusted significance threshold of $\alpha \leq 0.020$.

	V. unguiculata	***L. sericeus***	***T. candida***	***P. phaseoloides***
V. unguiculata	–	0.010*	<0.001*	0.020*
L. sericeus	0.02	–	<0.001*	<0.001*
T. candida	0.02	0.01	–	<0.001*
P. phaseoloides	0.01	0.04	0.06	–

CHAPTER 4

DEFINING THE SPECIES COMPLEX OF THE LEGUME POD BORER, *MARUCA VITRATA* (LEPIDOPTERA: CRAMBIDAE), BY COMPARATIVE MITOCHONDRIAL PHYLOGENOMICS

ABSTRACT

Mitochondrial DNA are useful tools for defining phylogenetic and evolutionary relationships of animal species. The legume pod borer, *Maruca vitrata* (Lepidoptera: Crambidae) is a pan-tropical species of Lepidoptera that is comprised of two unique strains that respectively inhabit the American continents (New World strain), and regions spanning from Africa through Southeast Asia, and Northern Australia (Old World strain). In this study, we present the complete mitochondrial genome sequence of *M. vitrata* from the New World, assembled *de novo* from whole genome shotgun sequence data generated on an Illumina HiSeq 2000. Phylogenomic comparisons were made to other previously published mitochondrial genome sequences from Crambid moths including Old World strain of *M. vitrata*. The 15, 385 bp *M. vitrata* (New World) sequence has a 80.7% A+T content, and encodes the 13 protein-coding, 2 ribosomal RNA, and 22 transfer RNA genes that have the typical orientation and arrangement of Lepidopteran mitochondrial DNAs. The derived arrangement of $tRNA^{Met}$-$tRNA^{Ilu}$-$tRNA^{Gln}$ following the A+T-rich control region, which has also been found in all Lepidoptera, is observed in *M. vitrata* (New World). The intergenic spacer region between $tRNA^{ser}$ and *nad*1 genes contains the ATACTAA motif, which is also present in all the six available Crambid mitogenomes, included in this study. The A+T-rich control region contains an 18 bp poly-T repeat

preceded by ATAG motif, and three tandem repeats $(AT)_3$, $(AT)_4$ and $(AT)_9$. Sequence variation between *M. vitrata* New World and Old World strains show an excess of synonymous substitution as a result of purifying selection, and suggest that divergence occurred ~1.87 mya. The phylogenetic relationships constructed based on amino acid sequences of the 13 protein coding genes also support the previously defined relationships among species of Lepidoptera. The mitochondrial genome of *M. vitrata* (New World strain) shares features common to those of other Crambid species, and prove quantitative estimation of divergence with the Old World strain. Furthermore, the *de novo* assembly of this mitochondrial genome from next generation sequencing (NGS) reads represents a readily available tool for the generation of data for similar phylogenomic studies.

INTRODUCTION

The mitochondrial genome (mitogenome) encodes proteins that are involved in electron transport and oxidative phosphorylation that is essential for energy production of the cell (Wolstenholme 1992; Boore 1999; Cameron 2014). Insect mitogenomes are circular double stranded DNA (dsDNA) molecules that typically range in size from 14 to 20 kb (Boyce et al. 1989; Wolstenholme 1992), and contain a conserved set of 13 protein coding genes (PCGs), 22 transfer RNAs (tRNAs) and 2 ribosomal RNA genes (rRNAs) that are required for translation of mitogenome-encoded proteins (Boore 1999). The A+T-rich non-coding control region that functions in the initiation of transcription and replication tends to be large in insect mitogenomes (Wolstenholme 1992), where the length is highly variable among different insects due to high rates of nucleotide substitution, insertions/deletions, and variable numbers of tandem repeats (Fauron and

Wolstenholme 1980; Inohira et al. 1997). Animal mitogenomes have a highly conserved gene content, maternal inheritance, and lack recombination (Avise 2000), which has facilitated their use in studies of phylogenetic relationships, population genetic structuring, as well as comparative evolution (Zhang et al. 1995; Boore 1999; Nardi et al. 2003; Arunkumar et al. 2006). Comparisons have allowed for estimation of short-term evolutionary patterns within and between closely related species (Ballard 2000; Coates et al. 2005). Although the gene content of the metazoan mitogenomes is highly conserved, variable lengths of A+T-rich coding regions, and gene order and orientation of tRNAs and PCGs are observed among insects (Hong et al. 2009).

To date, complete or near complete mitogenomes has been sequenced from more than 600 arthropod species (Cameron 2014). Although the insect order Lepidoptera contains approximately 200,000 species, 72 complete or near complete mitogenomes in 17 families and belonging to the Lepidopteran lineage Ditrysia are available (Cameron 2014). Among the family Crambidae, mitogenomes have been obtained for six species - *Ostrinia nubilalis* and *O. furnacalis* (Coates et al. 2005), *Diatraea saccharalis* (Li et al. 2011), *Maruca vitrata* (Old World) (Margam et al. 2011), *Cnaphalocrocis medinalis* (Chai et al. 2012), and *Chilo suppressalis* (Chai et al. 2012). Complete mitogenome sequences for Lepidoptera have been used to study the divergences between sibling and congeneric species (Yukuhiro et al. 2002; Coates et al. 2005), and as a stepping-stone to facilitate population level studies (Kim et al. 2006). Furthermore, full mitogenomes for Lepidoptera are important for the determination of phylogenetic relationships and identification of cryptic species (Lee et al. 2006; Chai et al. 2012).

Obtaining full mitogenome sequences has been a challenging and resource demanding task, where long-range PCR and subsequent primer walking (Huyse et al. 2007), or overlapping PCR products methods have been used to generate data by Sanger sequencing (Coates et al. 2005). Recently, next-generation sequencing (NGS) technologies have led to more straightforward pipelines for the assembly of complete mitogenome sequences (Jex et al. 2010a; Knaus et al. 2011; Ma et al. 2012). Mitogenomes have been sequenced from NGS libraries prepared from long-PCR amplified sequences (Maricic et al. 2010; Morin et al. 2010; Horn et al. 2011) or from sequence captured DNA molecules (Vasta et al. 2009). The long-range PCR strategies may remain useful for NGS-based sequencing, especially when limited genetic material can be recovered from small invertebrates (Jex et al. 2010b). Ultra-deep sequencing on NGS platforms can acquire data that is sufficient for the subsequent assembly of full mitogenomes at high read depth (Knaus et al. 2011; Ma et al. 2012). Also, due to the high copy number of mitochondria in most eukaryotic cells, NGS can reduce the impact of rare nuclear-integrated copies of mitochondrial DNAs or other non-target products (Ho and Gilbert 2010; Gilbert et al. 2007).

Species from the genus *Maruca* (Lepidoptera: Crambidae) are serious insect pests that feed on cowpea crops throughout the tropical and subtropical regions from northern Australia, and East Asia through sub-Saharan Africa to the Caribbean, Central America, and Hawaii (see Agunbiade et al. 2012 for references). Feeding by *M. vitrata* cause considerable crop loss and reduced yields to cultivated cowpea in many developing African countries, and is a target for improved and sustainable control methods (Agunbiade et al. 2012). Comparison of *M. vitrata* mitochondrial cytochrome c oxidase

subunit I (*cox*1) sequence data originating from samples collected in Africa and Puerto Rico revealed a high and unexpected levels of sequence variation, which suggested the existence of related subspecies or strains - *M. vitrata* in Africa (Old World) and the subspecies of *M. vitrata* from Puerto Rico (New World) (Margam et al. 2011). An analogous relationship among New and Old World species of Lepidoptera has also been predicted between *Helicoverpa zea* and *H. armigera* (Behere et al. 2007), but accurate estimates of time since divergence between these species is often lacking. In this study, Illumina HiSeq2000 sequencing reads were obtained from *M. vitrata* (New World), and used to successfully assemble the full mitogenome sequence. These data were applied to detect mitogenome-wide variance with its sister species, *M. vitrata* (Old World) and estimate time divergence between these New and Old World species. Furthermore, this study provides a pipeline for the rapid NGS acquisition and assembly of mitogenomes.

MATERIALS AND METHODS

Sample collection and DNA extraction

Insect samples of *M. vitrata* adults were collected from white bean (*Phaseolus vulgaris*) near Lares, Puerto Rico. All specimens were preserved in 100% ethanol and stored at 4°C. Total genomic DNA was extracted from a single individual adult using the DNeasy animal tissue kit following the manufacturer's instructions (QIAGEN, Valencia, CA) according to Margam et al. (2011). The quality of the DNA was assessed through electrophoresis on a 1% agarose gel before submission for library construction and sequencing.

Illumina HiSeq2000 ™ library construction and sequencing

A shotgun genomic DNA library was constructed using the TruSeq DNA Sample prep kit (Illumina, San Diego, CA). Briefly, 1 μg of genomic DNA was sonicated on a Covaris M220 (Covaris, Woburn, MA) using the default methods for sonication to a size of 500 bp library as described by the manufacturer. Sonicated DNA was blunt-ended, 3'-end A-tailed and ligated to an indexed adaptor. The adaptor-ligated gDNA was amplified by PCR to selectively enrich for those fragments that have adapters on both ends. Amplification was carried out for 8 cycles with the Kapa HiFi polymerase (Kapa Biosystems, Woburn, MA) to reduce the likeliness of multiple identical reads due to preferential clonal amplification. The library was size selected on a 2% agarose gel for fragments 600 bp to 800 bp in length. The size selected library was run on Agilent bioanalyzer DNA 7500 LabChips (Agilent, Santa Clara, CA) to determine the average fragment size and to confirm the presence of DNA of the expected size range and quantitated by qPCR on an ABI 7900. The indexed library was loaded onto a single lane of an 8-lane flowcell for cluster formation and paired-end sequenced at 100-bp lengths on an Illumina HiSeq 2000™ using the TruSeq SBS sequencing kits version 3. The raw .bcl files were converted into fastq files using the software Casava 1.8.2 (Illumina). All shotgun genomic DNA library construction and Illumina HiSeq 2000™ sequencing was carried out at the W. M. Keck Center for Comparative and Functional Genomics, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign.

Mitogenome assembly

Fastq output of read data was trimmed with Phred quality scores < 20 using the script TrimmingReads.pl (NGSToolKit; Patel and Jain 2012). A total of 30 million reads from the trimmed fastq sequence were loaded into the Velvet Assembler program 1.2.10 (Zerbino and Birney 2008) and assembled using the de Bruijn graph method with a hash size (*k*-mer) = 21, no coverage cutoff, and a minimum contig length (-min_contig_lgth) = 10000. Querying the National Center for Biotechnology Information (NCBI) non-redundant (nr) nucleotide database with the Blastn algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al. 1997) identified resulting contigs containing mitogenome sequence. The contigs that produced database "hits" were filtered for those with *E*-values $\geq 10^{-40}$ and percent similarities ≥ 75 to identify contigs with encoded mitochondrial gene sequences.

Sequence annotation

The FASTA-formatted consensus sequence was imported into the MITOS web server (Bernt et al. 2013; http://mitos.bioinf.uni-leipzig.de/index.py), a program used for *de novo* metazoan mitogenome annotation. To better define the limits of PCGs, coding sequences for each mitochondrial gene were aligned to that from published Crambid mitogenomes (See Appendix C) using the alignment algorithm on the CLC Genomics Workbench 6.5.2. The 5' end of the mitochondrial genes were inferred to be the first legitimate start codon in the open reading frame, and manual adjustments were made when needed. The 3' ends were inferred to be at the first in-frame stop codon encountered. When the stop codon was located within the sequence of a downstream gene encoded on

the same strand, a truncated stop codon (T) adjacent to the beginning of the downstream gene was designated as the termination codon. The positions and secondary structures for the tRNAs genes were confirmed using ARWEN 1.2 (Laslett and Canback 2008; http://mbio-serv2.mbioekol.lu.se/ARWEN/index.html). The graphical circular structure of the mitogenome was drawn using GenomeVx, a suite of tools for generating physical maps of plastids and mitogenomes (Conant and Wolfe 2008; http://wolfe.ucd.ie/GenomeVx/). All annotations were transferred to Sequin 13.05 (http://www.ncbi.nlm.nih.gov/Sequin/), and submitted to the NCBI nr database under accession number KJ466365.

Genome sequence analysis

The nucleotide sequences of the 13 PCGs of previously published 32 Lepidoptera species (Appendix C) were downloaded from the METAMIGA database (Feijao et al. 2006; http://amiga.cbmeg.unicamp.br/). The nucleotide compositions, codon usage and relative synonymous codon usage (RSCU) in the *M. vitrata* (New World) mitogenome were calculated with MEGA 5.2.2 (Tamura et al. 2011). The composition skew analysis was carried out to describe the base composition of nucleotide sequences, by measuring the relative number of As to Ts using the formula: AT skew = [A-T]/[A+T] and GC skew = [G-C]/[G+C], respectively (Perna and Kocher 1995).

The DNA and protein sequence similarities were calculated between PCGs of *M. vitrata* (New World) and orthologous genes from other species in the Family Crambidae using FASTA program, a tool for biological sequence comparison (Pearson and Lipman 1988; http://fasta.bioch.virginia.edu/fasta_www2/). The online program, microsatellite

repeats finder (http://insilico.ehu.es/mini_tools/microsatellites/) was used to analyze the A+T-rich control region for tandem repeats.

Phylogenetics and comparative genomic analysis

The 13 PCGs from 32 complete or near complete mitogenome sequences for the Lepidopterans, and also the complete mitogenome of *Drosophila melanogaster* within GenBank (Appendix C) were downloaded using METAMIGA (Feijao et al. 2006; http://amiga.cbmeg.unicamp.br/). The PCGs of these mitogenomes and that of *M. vitrata* (New World) were then concatenated for each insect species and imported into MEGA 5.2.2 (Tamura et al. 2011) and a multiple sequence alignment was performed with the ClustalW algorithm using default parameters (gap opening penalty 15, gap extension penalty 6.66, weight matrix IUB, and transition weight of 0.5). The mitogenome of *D. melanogaster* was included as an out-group. Maximum likelihood (ML) analysis of derived amino acid sequences, to infer the phylogenetic relationships among mitogenome sequences, was performed on MEGA 5.2.2. Support at each bifurcating node of the consensus tree was provided by 1000 bootstrap pseudo-replicates of the data (Felsenstein 1985). All gaps were deleted. Initial tree for the heuristic search was obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Jones, Taylor, and Thorton (JTT) probability matrix (Jones et al. 1992), and then selecting the topology with best log likelihood value.

The Bayesian inference (BI) analysis was performed based on the concatenated 13 PCG amino acid sequences of the mitogenomes of the 34 insects, including *M. vitrata* (New World) using MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and

Huelsenbeck 2003) and the Markov's chains were run simultaneously for 1,000,000 generations sampled every 1000 generations. The tree generated was viewed using FigTree 1.4.0 (Rambaut 2012).

Estimation of divergence times

The nucleotide sequence of *M. vitrata* (Old World; GenBank accession HM751150.1) was downloaded in FASTA format, and aligned with the nucleotide sequence obtained in this study for the *M. vitrata* New World sample (GenBank accession KJ466365) using MEGA 5.2.2 (Tamura et al. 2011) with the parameters described above. Gaps in the resulting alignment due to ambiguous bases (Ns) in HM7511150.1 as well as overhangs were deleted. The number of nucleotide differences between the two aligned sequences were calculated using DiffSeq from the EMBOSS Package (Rice et al. 2000). A clock-like rate of 2% mitochondrial DNA sequence change per million years has been estimated among *Drosophila* species (Powell 1986). Since no analogous estimates are available among Lepidoptera, we applied the *Drosophila* rate estimation as described by Coates et al. 2005.

RESULTS AND DISCUSSION

Mitogenome assembly and annotation

Paired-end Illumina sequencing generated a total of 336, 472, 660 million reads, of which 30 million were subsampled, and used to successfully generate a 15, 385 bp *de novo* assembly of the previously uncharacterized *M. vitrata* (New World) mitogenome (Table 4.1 and Figure 4.1). The size of the complete mitogenome from *M. vitrata* (New

World) is similar to the size previously reported from other species of Lepidoptera; ranging from 15, 314 bp in *Coreana raphaelis* (Kim et al. 2006) to 15, 928 bp in *Bombyx mandarina* (Yukuhiro et al. 2002). The annotated *M. vitrata* (New World) mitogenome contains the 13 PCGs, 22 tRNAs, and 2 rRNA genes which are typical for metazoans (Wolstenholme 1992; Table 4.1 and Figure 4.1). Analogous to all other Lepidopteran mitogenomes, of the total 37 genes, 23 were encoded on the major (forward) strand, whereas the 14 others were encoded on the minor (reverse) strand (Table 4.1 and Figure 4.1). The 13 PCGs spanned 11, 369 bp, and accounted for 73.9% of the entire *M. vitrata* (New World) mitogenome sequence. The *M. vitrata* (New World) mitogenome has the typical Lepidopteran gene order *tRNA*$^{\text{Met}}$-*tRNA*$^{\text{Ile}}$-*tRNA*$^{\text{Glu}}$-*nad*2 (Yukuhiro et al. 2002; Kim et al. 2006; Lee et al. 2006; Cameron and Whiting 2008a; Hong et al. 2008; Coates et al. 2005; Taylor et al. 1993), which differs from the ancestral gene order *tRNA*$^{\text{Ile}}$-*tRNA*$^{\text{Glu}}$-*tRNA*$^{\text{Met}}$ *nad*2 found in other insects (Boore 1999). This translocation of *tRNA*^{Met} was also present in the *M. vitrata* (New World) mitogenome. This was first reported by Taylor et al. (1993) in Lycaenids and Noctuids, but has now been found in five Lepidopteran superfamilies (Papilionoidea, Noctuoidea, Pyraloidea, Tortricoidea, and Bombycoidea) and is likely a synapomorphic character of all Lepidopteran mtDNAs (Salvato et al. 2008).

The whole *M. vitrata* (New World) mitogenome showed a high A+T content of 80.70% (A = 40.30%, T = 40.4%, G = 7.99% and C = 11.31%) (Table 4.2), which is consistent with the high A+T content of insect (Clary and Wolstenholme 1985; Crozier and Crozier 1993) and Lepidopteran mitogenomes (Liu et al. 2008). The nucleotide composition of the PCGs of the *M. vitrata* (New World) mitogenome, the A+T%, and G+C% values for the PCGs as well as the AT- and GC-skews, were calculated for all

available complete mtDNA genomes of the Crambids and are presented in the scatter plots of Figure 4.2. The average AT-skew of the Crambid mitogenomes is 0.01, ranging from -0.002 to 0.03, whereas the average GC-skew is -0.20, ranging from -0.17 to -0.26. The AT-skew of *M. vitrata* (New World) is -0.002 and the GC-skew is -0.17. A similar negative AT-skew (-0.02%) was also previously reported for *C. medinalis* (Chai et al. 2012) (Table 4.2).

Protein-coding genes

The arrangement of the PCGs is consistent with those of other animal mitogenomes (Table 4.1 and Figure 4.1). Nine of the PCGs were on the forward strand (*nad*2, *cox*1, *cox*2, *atp*8, *atp*6, *cox*3, *nad*3, *nad*6 and *cytb*) while the remaining 4 were on the reverse strand (*nad*1, *nad*4, *nad*4l and *nad*5). The putative start codons of PCGs used by *M. vitrata* (New World) (ATN; Table 4.1) are identical to those previously known in animal mitogenomes (Wolstenholme 1992). CGA was used as the start codon for *cox*1 as previously reported for *D. yakuba* (Clary and Wolstenholme 1985) and also for Lepidopteran species such as *B. mori* (Yukuhiro et al. 2002), *O. nubilalis* and *O. furnacalis* (Coates et al. 2005), *Adoxophyes honmai* (Lee et al. 2006), *C. raphaelis* (Kim et al. 2006), *Antheraea pernyi* (Liu et al. 2008), *B. mandarina* (Pan et al. 2008), *Ochrogaster lunifer* (Salvato et al. 2008), *Artogeia melete* (Hong et al. 2009), *Eriogyna pyretorum* (Jiang et al. 2009), *Hyphantria cunea* (Liao et al. 2010) and *M. vitrata* (Old World) (Margam et al. 2011). The termination codon TAA was used in all PCGs with the exception of *nad*3, which had TAG. Also, consistent with some other insect mitogenomes, *cox*1 and *cox*2 both had incomplete stop codon, T (Coates et al. 2005; Clary and

Wolstenholme 1985) (Table 4.1). The presence of incomplete stop codons is a feature shared with all Lepidopteran mitogenomes sequenced to date (Yukuhiro et al. 2002; Kim et al. 2006; Lee et al. 2006; Cameron and Whiting 2008; Hong et al. 2008; Coates et al. 2005) and more in general with many arthropod mitogenomes (Boore 1999). Although the motif TTAG has been observed to be located immediately upstream of the putative CGA start codon of some Lepidopteran mitogenomes and proposed to serve in a non-standard initiation process (Yukuhiro et al. 2002), we observed the motif ATAG immediately upstream of the CGA start codon of the *M. vitrata* (New World) *cox*1 gene. This motif was and has been conserved in some Lepidopteran mitogenomes but lacking in some (Margam et al. 2011). Among the available Crambid mitogenomes, *O. furnacalis* and *O. nubilalis* both have the motif TTAG upstream of the start codon of the *cox*1 gene while *M. vitrata* (Old World) has ATAG, which is similar to what obtains in *M. vitrata* (New World) mitogenome. Reduction in mitogenome size has led to minimal intergenic space as well as overlap between adjacent genes, where *M. vitrata* (New World) *atp8* and *atp6* PCGs have a predicted seven nucleotide overlap (Table 4.1). This feature is common to all Lepidopteran mitogenomes known (Yukuhiro et al. 2002; Kim et al. 2006; Lee et al. 2006; Cameron and Whiting 2008a; Hong et al. 2008; Coates et al. 2005) and is found in many animal mitogenomes (Boore 1999).

The A+T content of the PCGs is 79.4%, which is lower than the A+T content of the mitogenome as a whole (80.7%). The total number of codons in the forward strand excluding stop codons was 3777 and showed a strong bias toward AT-rich codons with the four most prevalent codons being Leu, UAA (481); Ile, AUU (443); Phe, UUU (344); Met, AUA (267) (Table 4.3). The codon usage prevalence is consistent with what obtains

in *M. vitrata* (Old World). Comparison of the 13 PCGs with the other six available Crambid mitogenomes showed between 75.3% and 97.9% nucleotide similarity, and 88.9% and 100% amino acid similarity with *M. vitrata* (New World) (Table 4.4). At the protein level, the average percentage similarity was highest in comparisons with *M. vitrata* (Old World) than with other Crambids. Derived *cox*1, *cox*3, *atp*6 and *nad*3 peptide sequences showed 100% similarity between *M. vitrata* (New World) and *M. vitrata* (Old World) (Table 4.4).

Transfer and ribosomal RNA genes

The *M. vitrata* (New World) mitogenome encoded 22 tRNAs typically found in animal mitogenomes (Boore 1999), and all formed cloverleaf-like secondary structures with the exception of $tRNA^{Ser1}$, in which its dihydrouridine (DHU) arm formed a simple loop (Figure 4.3). The lack of DHU arm in $tRNA^{Ser1}$ is a common condition in metazoan mitogenomes (Lavrov et al. 2000) but is not reported for all Lepidopterans. For example, *A. honmai* has all the tRNAs with a complete clover leaf structure (Lee et al. 2006). Of the 22 tRNA genes, 14 genes are encoded on the forward strand while 8 are encoded on the reverse strand. The length of the tRNAs ranged from 63 bp to 71 bp.

Consistent with what is observed in all other insect mitogenomes, the mitogenome of *M. vitrata* (New World) contained two rRNAs with a total length of 2369 bp. The large rRNA (*rrnL*) had a length of 1304 bp while the small rRNA (*rrnS*) had a length of 765 bp (Table 4.1).

Non-coding regions

The *M. vitrata* (New World) mitogenome contains characteristics typical of other Lepidopteran mitogenomes. The 7 bp intergenic spacer located between $tRNA^{Ser2}$ and *nad1* contains the ATACTAA motif (Figure 4.4), which is conserved across the order Lepidoptera (Cameron and Whiting 2008). This motif is possibly fundamental to site recognition by the transcription termination peptide (mtTERM protein) (Taanman 1999) and is present in most insect mitogenomes (Cameron and Whiting 2008). Insect mitogenomes often contain a non-coding A+T-rich region, which varies considerably in length among insect species, or even within the same species (Zhang and Hewitt 1997). Another intergenic space and which is the largest consisted of the A+T-rich control region, which ranges in size in Crambid moth and is consistent with previous findings from other insects (Hua et al. 2008; Oliveira et al. 2008; Ma et al. 2009). Tandem repetitive sequences are common in the control region for most insects that can vary in length and copy number (Dotson and Beard 2001), but repeats in Lepidoptera are considered to share greater degrees of conservation compared to other insect groups (Cameron and Whiting 2008). The A+T rich control region in *M. vitrata* (New World) was estimated at 341 bp and contained (AT) tandem repeats (Figure 4.5 and Appendix D). The A+T-rich region contained a conserved ATAG motif followed by an 18 bp poly-T stretch (Figure 4.5). This is similar to the pattern found in *C. medinalis*, C. *suppressalis* and *D. saccharalis* but the length of the poly-T stretch varies between species. This structure is suggested to function as a signal for mitochondrial DNA replication initiation (Hu et al. 2010; Yin et al. 2010; Kim et al. 2010). A poly-A commonly observed in other Lepidopteran mitogenomes was also found immediately upstream of $tRNA^{Met}$ but in *M.*

vitrata (New World), is 9bp long. The control region also has tandem repeats with the most common repeat being AT - $(AT)_3$, $(AT)_4$ and $(AT)_9$ (Appendix D). Similar patterns were observed in *Hyphantria cunea* $(AT)_8$ (Liao et al. 2010) and *Ochrogaster lunifer* $(AT)_7$ (Salvato et al. 2008).

Phylogenetic and comparative genomic analysis

Phylogenetic analyses were carried out using nucleotide sequences of mitogenomes from 32 Lepidopteran insects, and rooted with *D. melanogaster* as an out-group (Figures 4.6a and 4.6b). BI and ML analyses generated identical tree topologies (Figures 4.6a and 4.6b). The major division in Ditrysia, between moth and butterfly species was weakly supported, but the relationship among moth superfamilies Bombycoidea, Noctuoidea, Geometroidea, Pyraloidea and Tortricoidea mostly showed high levels of bootstrap support. The inter-relationships among superfamilies conformed to those proposed by Kristensen and Skalski (1999). These results suggest that molecular phylogenies based on derived protein sequence from complete mitogenomes may be valuable for deciphering Lepidopteran evolutionary lineages, although higher rates of mutation observed in mitochondrial sequences compared to nuclear loci may cause long branch attraction or other phenomenon which reduce the ability to resolve deeper branches within the tree. As expected, phylogenetic analysis of the mitogenome indicated that *M. vitrata* (Old World) and *M. vitrata* (New World) are highly related, and may comprise sister species similar to that shown between *O. nubilalis* and *O. furnacalis* (Coates et al. 2005).

Protein-coding sequence was shown to be highly similar between *M. vitrata* (New World) and *M. vitrata* (Old World) mitogenomes (see above; Table 4.4). Comparison between *M. vitrata* (New World) and *M. vitrata* (Old World) mitochondrial gene coding sequences predicted a total of 476 nt substitutions from a 11, 250 nt consensus (4.23% divergence) of which 74 were non-synonymous mutations that resulted in an amino acid change. This suggests that purifying selection may act to curb the rate of amino acid change in the genus *Maruca*, but has also been shown for most, if not, all insect mitogenomes. Alignment of entire mitogenome sequences from *M. vitrata* (Old World) and *M. vitrata* (New World) resulted in a 14,039 consensus (gaps excluded) and showed a total of 526 nucleotide differences (3.75%) (Appendix E). Making a simple assumption of a clock-like rate of 2% change per million years estimated from the *Drosophila* species group (Powell 1986) and acknowledging the potential that variance in this rate may have affected the degree of haplotype divergence among *Maruca* subspecies, these species may have diverged ~1.87 mya. This simple method for estimation was used due to lack of fossil evidence to assist in calibrating a molecular clock.

CONCLUSIONS

The mitogenome from *M. vitrata* (New World) is typical of mitogenomes from the insect Order Lepidoptera. Short-term evolutionary change in mitogenomes with the sister species *M. vitrata* (Old World) showed a distinct bias for substitution at non-synonymous sites. These results are important in ongoing efforts to define genetic and genomic variation between these two recently described sister species, and along with phylogenetic analyses presented herein suggest that whole mitogenome sequences may be tools for

species delineation. Furthermore, this study indicates that mitogenomes can be *de novo* assembled from NGS sequence data, and procedures outlined here may offer a rapid and computationally accessible method for future mitochondrial genomic studies.

REFERENCES

Agunbiade T, Steele L, Coates B, Gassmann A, Margam V, Ba M, Dabire-Binso C, Baoua I, Bello-Bravo J, Seufferheld F, Sun W, Tamò, Pittendrigh BR (2012) IPM-omics: From genomics to extension for integrated pest management of cowpea. In Boukar O, Coulibaly O, Fatokun C, Lopez K, Tamò M (Eds.), *Improving livelihoods in the cowpea value chain through advancements in science*. Proceedings of the 5th World Cowpea Research Conference, pp. 231 – 248.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.

Arunkumar KP, Metta M, Nagaraju J (2006) Molecular phylogeny of silk moths reveals the origin of domesticated silk moth, *Bombyx mori* from Chinese *Bombyx mandarina* and paternal inheritance of *Antheraea proylei* mitochondrial DNA. *Molecular Phylogenetics and Evolution* 40: 419-427.

Avise JC (2000) Phylogeography: The history and formation of species. Harvard University Press, Cambridge, MA 447pp.

Ballard JWO (2000) Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *Journal of Molecular Biology* 51(1): 64–75.

Behere GT, Tay TW, Russell AD, Heckel GD, Appleton RD, Kranthi KR, Batterham P (2007) Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. *BMC Evolutionary Biology* 7: 117.

Bernt AD, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69(2): 313-319.

Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research* 27: 1767–1780.

Boyce TM, Zwick ME, Aquadro CF (1989) Mitochondrial DNA in the bark weevils: Size, structure and heteroplasmy. *Genetics* 123: 825-836.

Cameron SL (2014) Insect mitochondrial genomics: Implications for evolution and phylogeny. *Annual Review of Entomology* 59: 95-117.

Cameron SL, Whiting MF (2008) The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene* 408: 112-123.

Chai HN, Du YZ, Zhai BP (2012) Characterization of the complete mitochondrial genomes of *Cnaphalocrocis medinalis* and *Chilo suppressalis* (Lepidoptera: Pyralidae). *International Journal of Biological Sciences* 8: 561-579.

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution* 22: 252–271.

Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2005) Partial mitochondrial genome sequences of *Ostrinia nubilali*s and *Ostrinia furnicalis*. *International Journal of Biological Sciences* 1:13-18.

Conant GC, Wolfe KH (2008) GenomeVx: Simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24: 861-862.

Crozier RH and Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: Complete sequence and genome organization. *Genetics* 133(1): 97-117.

Dotson EM, Beard CB (2001) Sequence and organization of the mitogenome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Molecular Biology* 10: 205–215.

Fauron CMR, Wolstenholme DR (1980) Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine+thymine-rich region of mitochondrial DNA molecules. *Nucleic Acids Research* 8: 2439-2452.

Feijao PC, Neiva LS, Azeredo-Espin AML, Lessinger AC (2006) AMiGA: The arthropodan mitochondrial genomes accessible database. *Bioinformatics* 22: 902–903.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39(4): 783-791.

Gilbert MT, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalén L, Kuznetsova T, Kosintsev P, Campos PF, Higham T, Collins MJ, Wilson AS, Shidlovskiy F, Buigues B, Ericson PG, Germonpré M, Götherström A, Iacumin P, Nikolaev V, Nowak-Kemp M, Willerslev E, Knight JR, Irzyk GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W, Schuster SC (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317: 1927–1930.

Ho SYW, Gilbert MTP (2010) Ancient mitogenomics. *Mitochondrion* 10: 1–11

Hong G, Jiang S, Yu M, Yang Y, Li F, Xue F, Wei Z (2009) The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). *Acta Biochimica et Biophysica Sinica (Shanghai)* 41: 446–455.

Hong MY, Lee EM, Jo YH, Park HC, Kim SR, Hwang JS, Jin BR, Kang PD, Kim KG, Han YS, Kim I (2008) Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other Lepidopteran insects. *Gene* 413: 49-57.

Horn S, Durka W, Wolf R, Ermala A, Stubbe A, Stubbe M, Hofreiter M (2011) Mitochondrial genomes reveal slow rates of molecular evolution and the timing of speciation in beavers

(Castor), one of the largest rodent species. *PLoS ONE* 6:e14622.

Hu J, Zhang D, Hao J, Huang D, Cameron S, Zhu C (2010) The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae: Heliconiinae: Acraeini): Sequence, gene organization and a unique tRNA translocation event. *Molecular Biology Reports* 37: 3431–3438.

Hua JM, Li M, Dong PZ, Cui Y, Xie Q (2008) Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). *BMC Genomics* 9: 610.

Huelsenbeck JP and F Ronquist (2001) "MrBayes: Bayesian inference of phylogenetic trees". *Bioinformatics* 17(8): 754–755.

Huyse T, Plaisance L, Webster BL, Mo TA, Bakke TA, Bachmann L, Littlewood DTJ (2007) The mitochondrial genome of *Gyrodactylus salaris* (Platyhelminthes: Monogenea), a pathogen of Atlantic salmon (*Salmo salar*). *Parasitology* 134: 739–747.

Inohira K, Hara T, Matsuura ET (1997) Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. *Molecular Biology and Evolution* 14: 814-822.

Jex AR, Hall RS, Littlewood DTJ, Gasser RB (2010) An integrated pipeline for next-generation sequencing and annotation of mitogenomes. *Nucleic Acids Research* 38: 522-533.

Jex AR, Littlewood DTJ, Gasser RB (2010b) Toward next-generation sequencing of mitochondrial genomes - focus on parasitic worms of animals and biotechnological implications. *Biotechnology Advances* 28:151-159.

Jiang S, Hong G, Yu M, Li N, Yang Y, Liu Y, Wei Z (2009) Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). *International Journal of Biological Sciences* 5: 351–365.

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 8: 275–282.

Kim I, Lee EM, Seol KY, Yun EY, Lee YB, Hwang JS, Jin BR (2006) The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). *Insect Molecular Biology* 15(2): 217-225.

Kim MJ, Wan X, Kim KG, Hwang JS, Kim I (2010) Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). *African Journal of Biotechnology* 9: 735-754.

Knaus BJ, Cronn R, Liston A, Pilgrim K, Schwartz MK (2011) Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology* 11: 10.

Kristensen NP, Skalski AW (1999) Phylogeny and paleontology. In Kristensen NP (Ed.), *Lepidoptera: Moths and Butterflies, 1: Evolution, Systematics, and Biogeography*. Handbook of Zoology. Berlin and New York: De Gruyter. Volume IV, Part 357-25.

Laslett D, Canbäck B (2008) ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24:172-175.

Lavrov DV, Brown WM, Boore JL (2000) A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proceedings of the National Academy of Sciences USA* 97(25): 13738-42.

Lee E-S, Shin KS, Kim M-S, Park H, Cho S, Kim C-B (2006) The mitochondrial genome of the smaller tea tortrix, *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene* 373: 52-57.

Li W, Zhang X, Fan Z, Yue B, Huang F, King E, Ran J (2011) Structural characteristics and phylogenetic analysis of the mitochondrial genome of the sugarcane borer, *Diatraea saccharalis* (Lepidoptera: Crambidae). *DNA and Cell Biology* 30(1): 3-8.

Liao F, Wang L, Wu S, Li YP, Zhao L (2010) The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae). *International Journal of Biological Sciences* 6: 172–186.

Liu Y, Li Y, Pan M, Dai F, Zhu X, Lu C, Xiang Z (2008) The complete mitochondrial genome of the Chinese oak silkmoth, *Antheraea pernyi* (Lepidoptera: Saturniidae). *Acta Biochimica et Biophysica Sinica (Shanghai)* 40: 693–703.

Ma C, Liu C, Yang P, Kang L (2009) The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmoratus* and *Oedaleus asiaticus*. *BMC Genomics* 10: 1–12.

Margam V, Coates BS, Hellmich RL, Agunbiade T, Seufferheld JM, Sun W, Ba MN, Sanon A, Binso-Dabire C, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011) Mitochondrial genome sequence and expression profiling for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(2): e16444.

Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* 5: e14004.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Nielsen SCA, Rasmussen M, Willerslev E, Gilbert MTP, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* 20: 908-916.

Nardi F, Caeapelli A, Dallai R, Frati F (2003) The mitochondrial genome of the olive fly, *Bactrocera oleae*: Two haplotypes from distant geographic locations. *Insect Molecular Biology* 12: 605-611.

Oliveira DCSG, Raychoudhury R, Lavrov DV, Werren (2008) Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution* 25: 2167-2180.

Pan M, Yu Q, Xia Y, Dai F, Liu Y, Zhang Z, Xiang Z (2008) Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bombyx mandarina* (Lepidoptera: Bombycidae). *Science in China Series C: Life Sciences* 51: 693–701.

Patel RK, Jain M (2012) NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* 7: e30619.

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences USA*. 85(8): 2444-8.

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitogenomes. *Journal of Molecular Evolution* 41: 353–358.

Powell JR (1986) Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. *Proceedings of the National Academy of Sciences USA* 83(23): 9090–9093.

Rambaut A (2012) FigTree v1.4.0. Available from: http://tree.bio.ed.ac.uk/software/figtree/.

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. *Trends in Genetics* 16(6): 276-277.

Ronquist F and JP Huelsenbeck (2003) "MrBayes 3: Bayesian phylogenetic inference under mixed models". *Bioinformatics* 19(12): 1572–1574.

Salvato P, Simonato M, Battisti A, Negrisolo E (2008) The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* 9: 331.

Taanman JW (1999) The mitochondrial genome: Structure, transcription, translation and replication. *Biochimica et Biophysica Acta* 1410: 103-123.

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences USA* 101: 11030-11035.

Taylor MFJ, McKechnie SW, Pierce N, Kreitman M (1993) The Lepidopteran mitochondrial control region: Structure and evolution. *Molecular Biology and Evolution* 10: 1259-1272.

Vasta V, Ng SB, Turner EH, Shendure J, Hahn SH (2009) Next generation sequence analysis for mitochondrial disorders. *Genome Medicine* 1: 100.

Wolstenholme DR (1992) Animal mitochondrial DNA: Structure and evolution. *International Review of Cytology* 141: 173–216.

Yin J, Hong GY, Wang AM, Cao YZ, Wei ZJ (2010) Mitochondrial genome of the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and comparison with other Lepidopterans. *Mitochondrial DNA* 21: 160–169.

Yukuhiro K, Sezutsu H, Itoh M, Shimizu K and Banno Y (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. *Molecular Biology and Evolution* 19 (8): 1385-1389.

Zerbino DR, Birney E (2008) Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.

Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology* 25: 99–120.

Zhang R, Lu Z, Zhao H, Zhang X, Diasio RB, Habus I, Jiang Z, Iyer RP, Yu D, Agrawal S (1995) In vivo stability, disposition and metabolism of a "hybrid" oligonucleotide phosphorothioate in rats. *Biochemical Pharmacology* 50: 545–556.

FIGURES AND TABLES

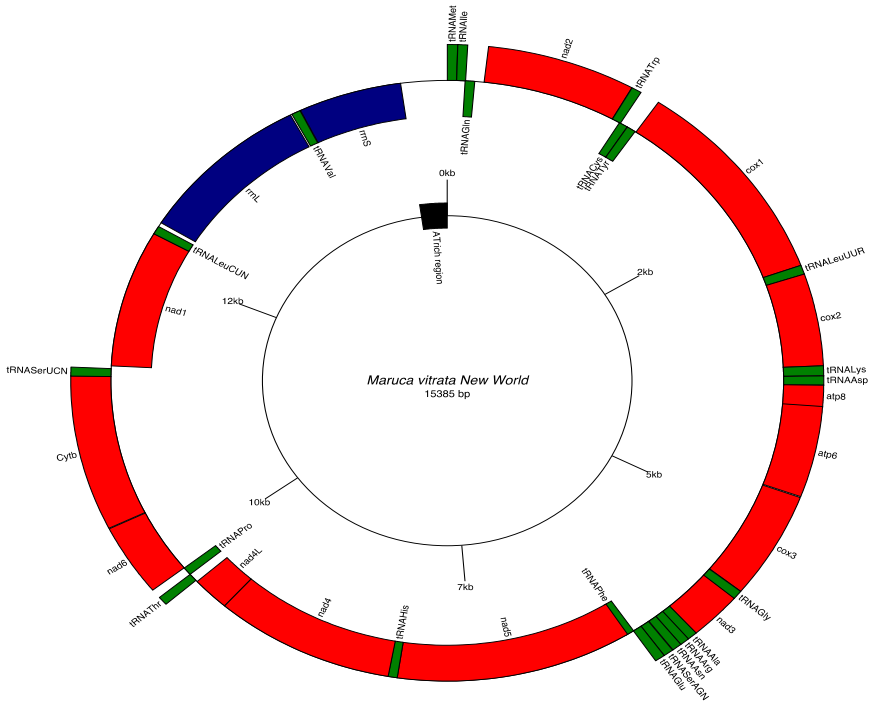
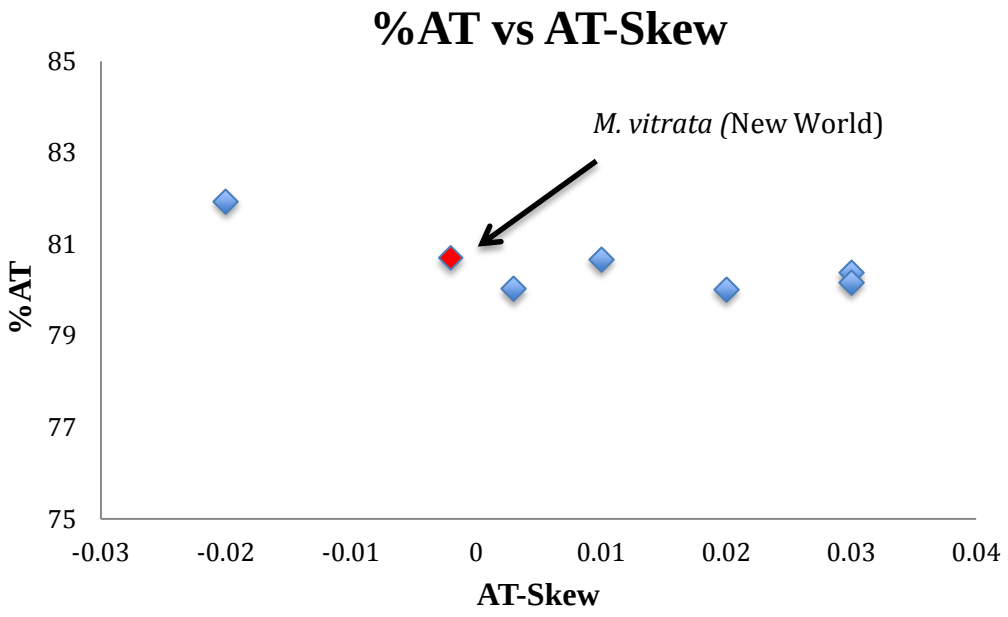


Figure 4.1. Gene arrangements in the circular mitogenome of *M. vitrata* (New World).

a)



b)

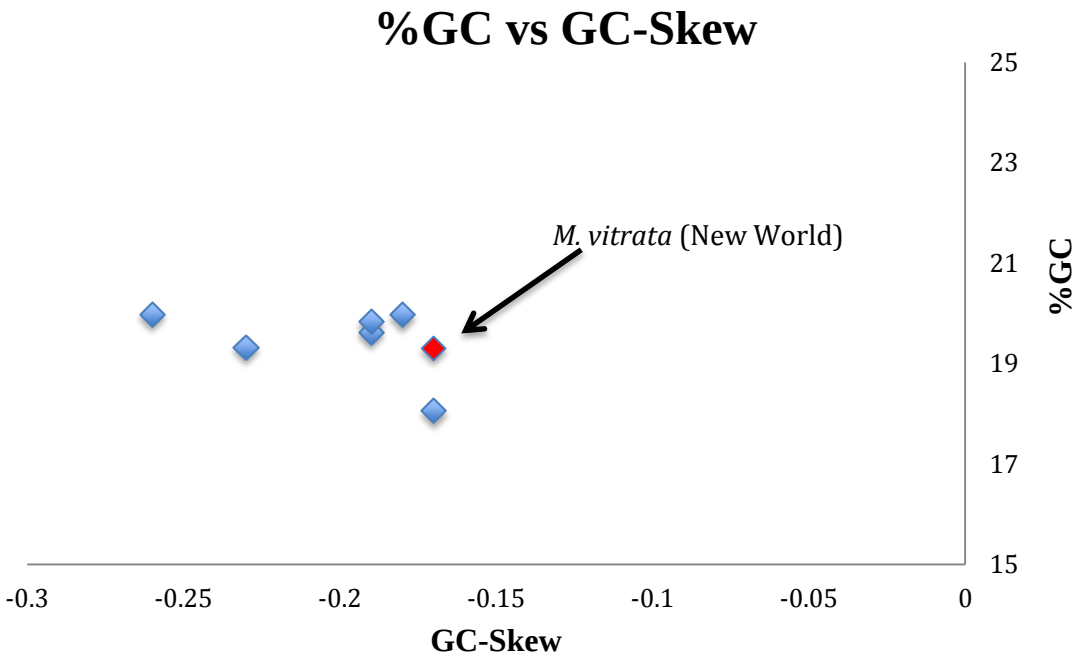


Figure 4.2. Scatter plot of the AT% vs AT-skew and GC% vs GC-skew in Crambid mitogenomes.

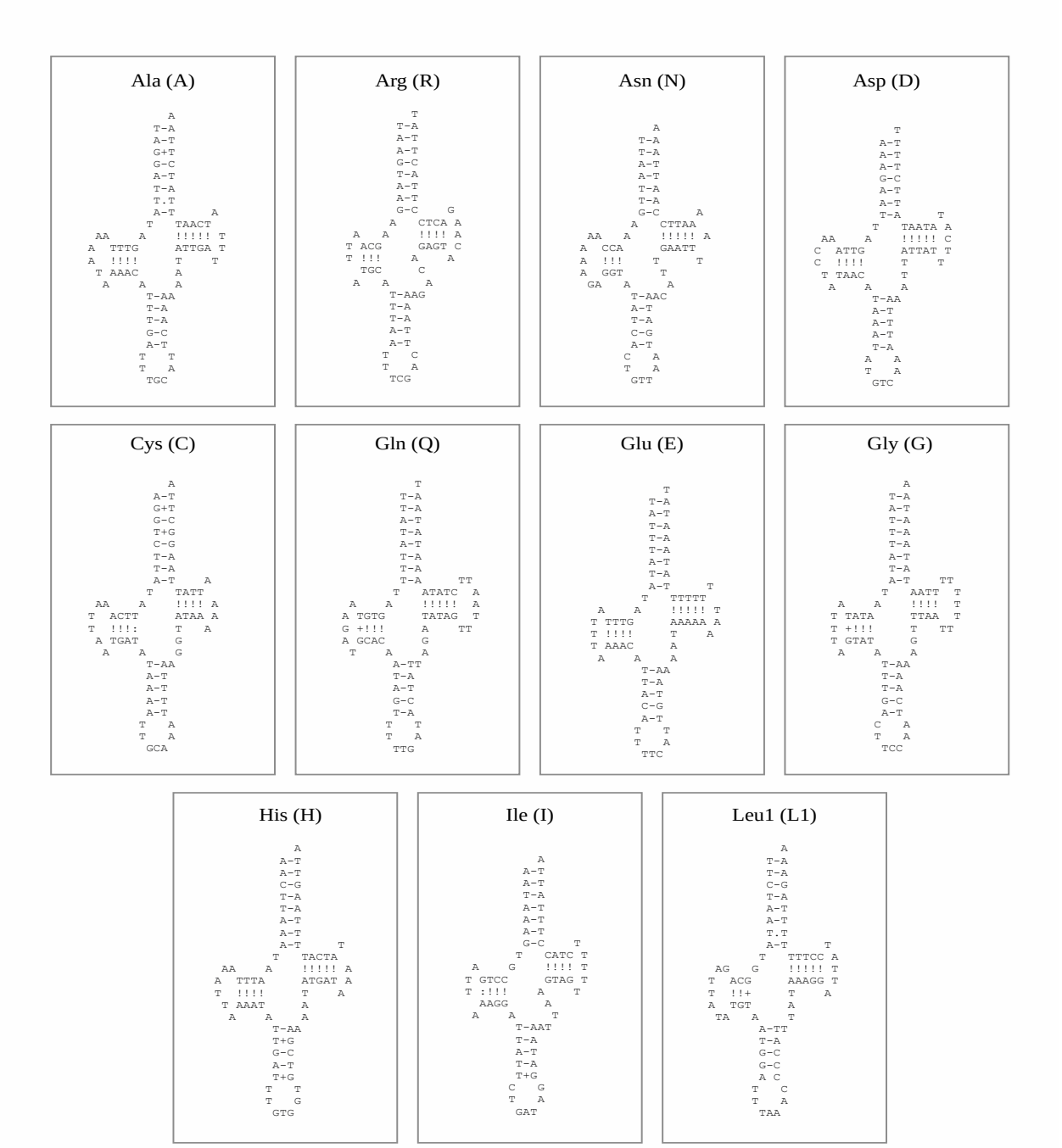


Figure 4.3. Inferred secondary structures for the 22 tRNAs of *M. vitrata* (New World) mitogenome. The tRNAs are labeled with the abbreviations of their corresponding amino acids. Watson-Crick base pairing is indicated as pairs by lines.

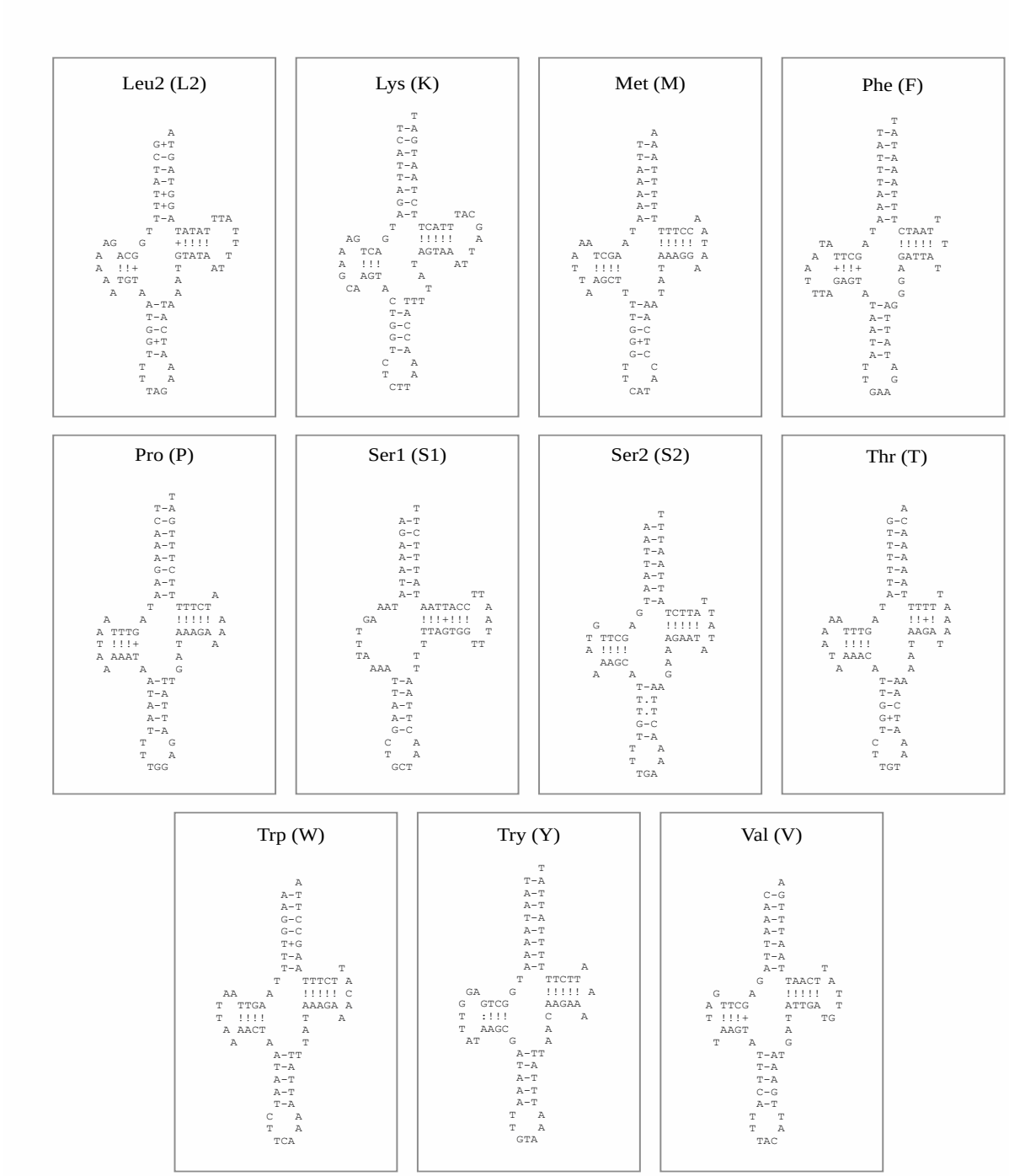


Figure 4.3. (cont.)

M. vitrata (New World)	ATACTAAAAATAATATAATAAT
M. *vitrata* (Old World)	ATACTAAAAATAATATCT
C. medinalis	ATACTAAATAATATATTT
O. furnacalis	ATACTAATAATATTAACTTAAT-
O. nubilalis	ATACTAAAAATATTAACTTACTTACTTAA
C. suppressalis	ATACTAAATATATTAATA
D. saccharalis	ATACTAAATTTATTTATA

Figure 4.4. Alignment of the ATACTAA spacer region (*nad*1 – $tRNA^{SER2}$) across the six available Crambid mitogenomes including *M. vitrata* (New World).

TATTGTAGGATTTTAGACATAGTTTTTTTTTTTTTTTTTTATATATATAAAATTTAATATAAATTATTAAATATTAAATAT

Origin of light strand replication *$(AT)_4$ Tandem repeats*

TTTCTTTCTTTTTCTTCTTTATAACATTAATATTAAAAATTAATACGTAGATTCATCGATTAATAATCATTTAAATAAATAATTAATTAATATATTTTAAAATTAATTAAATTGAAATTTAAAATATTAATTTTACTAAATTAATTAATTAATTTTAATTAA

$(AT)_3$ Tandem repeats

TATTAAAAATATTAATAAATTAAATATTTAATATATATATATATATATAAATATAAACCGTTTTTAATATTTTTTCTATA

$(AT)_9$ Tandem repeats

AATAAAAAAAAA

9 bp Poly-A stretch

Figure 4.5. The structure of the A+T-rich Region of *M. vitrata* (New World) mitogenome.

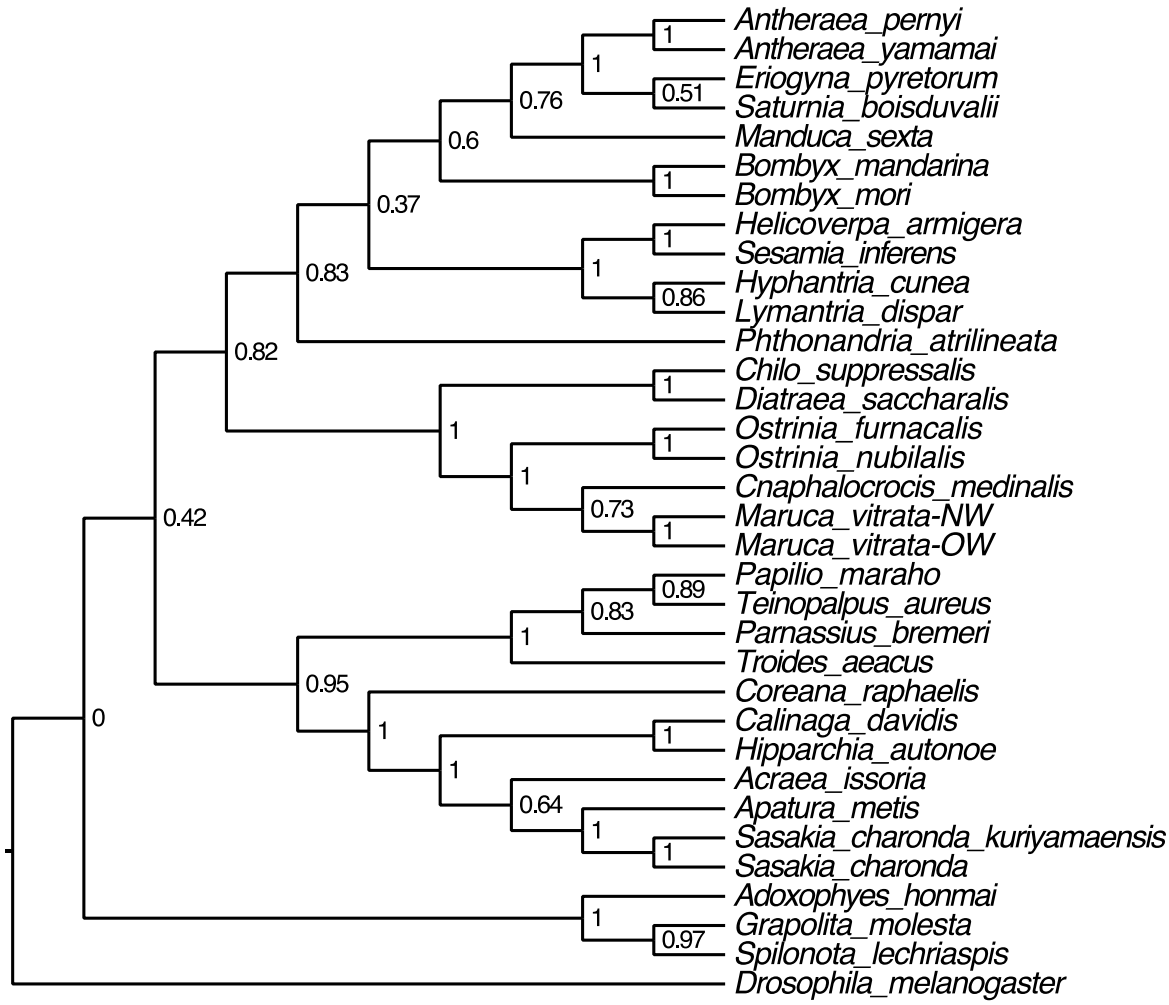


Figure 4.6a. Phylogenetic relationships among species from the insect order Lepidoptera, which is rooted with *D. melanogaster* (Insecta: Diptera) as an out-group. Numbers at each node indicate maximum likelihood bootstrap support. The tree is based on a concatenated sequence of 13 derived proteins from complete or near-complete mitogenomes. GenBank accessions used for all species in the phylogenetic analysis are listed in Appendix C.

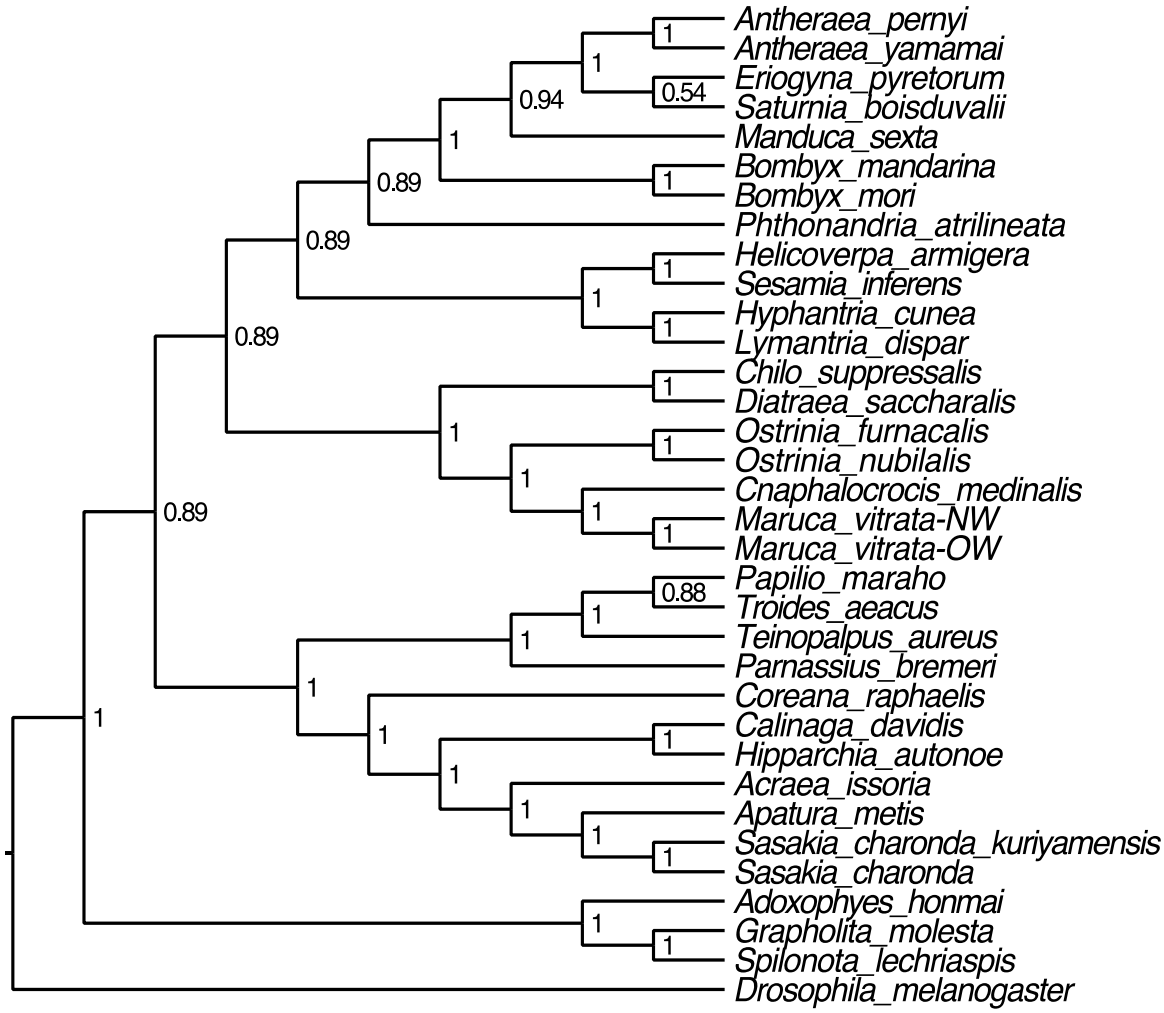


Figure 4.6b. Phylogenetic relationships among species from the insect order Lepidoptera, which is rooted with *D. melanogaster* (Insecta: Diptera) as an out-group. Numbers at each node indicate Bayesian posterior probabilities. The tree is based on a concatenated sequence of 13 derived proteins from complete or near-complete mitogenomes. GenBank accessions used for all species in the phylogenetic analysis are listed in Appendix C.

Table 4.1. Annotation of the full mitogenome of *M. vitrata* (New World).

Gene	Strand	CDS Positions	Size (bp)	Start Codon	Stop Codon	Anticodon
tRNAMet	F	1..68	68			CAT
tRNAIle	F	69..133	65			GAT
tRNAGln	R	130..200	69			TTG
*nad*2	F	269..1255	987	ATA	TAA	
tRNATrp	F	1257..1324	68			TCA
tRNACys	R	1316..1382	65			GCA
tRNATyr	R	1383..1451	67			GTA
*cox*1	F	1458..2988	1531	CGA	T--	
tRNALeuUUR	F	2988..3056	67			TAA
*cox*2	F	3056..3737	682	ATG	T-	
tRNALys	F	3737..3809	71			CTT
tRNAAsp	F	3811..3878	68			GTC
*atp*8	F	3879..4040	162	ATC	TAA	
*atp*6	F	4034..4708	675	ATG	TAA	
*cox*3	F	4715..5503	789	ATG	TAA	
tRNAGly	F	5505..5573	67			TCC
*nad*3	F	5570..5926	357	ATA	TAG	
tRNAAla	F	5924..5992	66			TGC
tRNAArg	F	5990..6054	63			TCG
tRNAAsn	F	6054..6119	66			GTT
tRNASerAGN	F	6122..6189	66			GCT
tRNAGlu	F	6189..6256	66			TTC
tRNAPhe	R	6255..6324	68			GAA
*nad*5	R	6307..8058	1752	ATT	TAA	
tRNAHis	R	8058..8126	67			GTG
*nad*4	R	8125..9465	1341	ATG	TAA	

Table 4.1. (cont.)

*nad*4L	R	9465..9746	282	ATA	TAA	
tRNAThr	F	9758..9823	66			TGT
tRNAPro	R	9823..9890	66			TGG
*nad*6	F	9892..10422	531	ATT	TAA	
Cytb	F	10426..11574	1149	ATG	TAA	
tRNASerUCN	F	11573..11638	66			TGA
*nad*1	R	11660..12790	1131	ATG	TAA	
tRNALeuCUN	R	12791..12859	69			TAG
rrnL	R	12894..14197	1304			
tRNAVal	R	14210..14276	67			TAC
rrnS	R	14280..15044	765			
A+T-rich region		15045..15385	341			

Table 4.2. Nucleotide composition in the whole mitogenomes of *M. vitrata* (New World) and other available mitogenomes from 6 additional species from the Lepidopteran family Crambidae.

Species	**A**	**T**	**A+T**	**AT-Skew**	**G**	**C**	**G+C**	**GC-Skew**
Cnaphalocrocis medinalis	40.36	41.58	81.94	-0.02	7.45	10.61	18.06	-0.17
Chilo suppressalis	40.64	40.03	80.67	0.01	7.39	11.94	19.33	-0.23
Diatraea saccharalis	40.87	39.15	80.02	0.02	7.42	12.56	19.98	-0.26
*Ostrinia furnacalis**	41.46	38.92	80.38	0.03	7.91	11.71	19.62	-0.19
*Ostrinia nubilalis**	41.36	38.81	80.17	0.03	8.02	11.82	19.83	-0.19
Maruca vitrata (Old World)*	40.14	39.89	80.03	0.003	8.21	11.76	19.97	-0.18
Maruca vitrata (New World)	40.27	40.44	80.70	-0.002	7.99	11.31	19.30	-0.17

* partial mitogenome sequences which lacked portions of the AT-rich control region sequence.

Table 4.3. Codon usage and relative synonymous codon usage (RSCU) in *M. vitrata* (New World) mitogenome.

Amino Acid	Codon	Count	RSCU	Amino Acid	Codon	Count	RSCU
Ala (A)	GCU	78	2.22	**Pro (P)**	CCU	75	2.21
	GCC	7	0.21		CCC	9	0.29
	GCA	40	1.21		CCA	43	1.14
	GCG	1	0.05		CCG	2	0.06
Cys (C)	UGU	26	1.31	**Gln (Q)**	CAA	63	2.00
	UGC	4	0.23		CAG	0	0.00
Asp (D)	GAU	65	1.32	**Arg (R)**	CGU	13	0.90
	GAC	3	0.22		CGC	0	0.00
Glu (E)	GAA	72	1.72		CGA	41	2.71
	GAG	6	0.13		CGG	1	0.08
Phe (F)	UUU	344	1.81	**Ser1 (S)**	AGU	32	0.75
	UUC	32	0.19		AGC	3	0.05
Gly (G)	GGU	67	1.23		AGA	81	1.89
	GGC	4	0.09		AGG	0	0.00
	GGA	122	2.15	**Ser2 (S)**	UCU	109	2.77
	GGG	15	0.22		UCC	4	0.06
His (H)	CAU	57	1.34		UCA	87	2.39
	CAC	12	0.35		UCG	4	0.09
Ile (I)	AUU	443	1.88	**Thr (T)**	ACU	89	2.23
	AUC	26	0.12		ACC	9	0.31
Lys (K)	AAA	93	1.77		ACA	58	1.12
	AAG	14	0.23		ACG	2	0.04
Leu1 (L)	CUU	27	0.31	**Val (V)**	GUU	86	2.22
	CUC	3	0.04		GUC	2	0.03
	CUA	23	0.68		GUA	59	1.71
	CUG	0	0.00		GUG	3	0.04
Leu2 (L)	UUA	481	4.86	**Trp (W)**	UGA	94	2.00
	UUG	11	0.12		UGG	0	0.00
Met (M)	AUA	267	1.83	**Try (Y)**	UAU	183	1.84
	AUG	23	0.17		UAC	13	0.16
Asn (N)	AAU	238	1.94				
	AAC	8	0.06				

Table 4.4. Percent nucleotide and amino acid similarity in the 13 protein-coding gene sequences of *M. vitrata* (New World) compared to orthologs in 6 other mitogenomes sequenced from species in the Lepidopteran family Crambidae.

Genes	Mv		Cs		Cm		Ds		Of		On	
	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)	DNA (%)	Protein (%)
***nad*2**	97.9	99.3	84.6	98.5	86.5	97.6	78.7	93.9	83.9	97.3	83.9	97.3
***nad*1**	95.4	99.3	86.6	98.3	88.3	99.7	83.9	95.4	87.4	98.7	87.2	98.3
***cox*1**	94.3	100	88.6	99.2	89.5	97.5	86.5	99	88.5	99.4	87.7	98.6
***cox*2**	94.9	99.6	88	98.7	90	98.2	85.9	98.2	88.1	99.6	87.8	98.7
***cox*3**	94.6	100	86.8	98.9	89	99.2	86.2	98.9	86.4	98.1	86.2	98.1
***atp*8**	93.8	98.1	81.2	88.9	78.6	85.5	75.3	88.7	85.8	94.3	85.2	94.3
***atp*6**	93.9	100	84	96	90.4	99.6	79.9	94.7	85.6	96.9	85.2	96.9
***nad*3**	94.3	100	84.9	97.4	88.4	99.1	81.5	98.3	84.4	98.3	84.2	98.2
***nad*4**	95.7	99.3	87.6	96.6	91.6	98.2	84.3	96	88.1	94.1	87.9	94.1
***nad*4l**	96.4	93.6	89.4	97.8	91.8	97.8	87.9	96.8	89.7	94.6	89	93.5
***nad*5**	96.1	99.8	86	95.4	88.5	97.2	85.4	95.7	87.9	97.4	87.3	97.4
***nad*6**	96.8	96.9	82.5	93.2	85.8	96	78.5	92.6	83.7	96	83.3	96
Cytb	97.5	99.5	86.5	97.9	89.3	98.2	85.9	97.6	87.2	98.2	87.5	98.4

CHAPTER 5

DEVELOPMENT OF REFERENCE TRANSCRIPTOMES FOR THE MAJOR FIELD INSECT PESTS OF COWPEA: A TOOLBOX FOR INSECT PEST MANAGEMENT APPROACHES IN WEST AFRICA[4]

ABSTRACT

Cowpea is a widely cultivated and major nutritional source of protein for many people that live in West Africa. Annual yields and longevity of grain storage is greatly reduced by feeding damage caused by a complex of insect pests that include the pod sucking bugs, *Anoplocnemis curvipes* Fabricius (Hemiptera: Coreidae) and *Clavigralla tomentosicollis* Stål (Hemiptera: Coreidae); as well as phloem-feeding cowpea aphids, *Aphis craccivora* Koch (Hemiptera: Aphididae) and flower thrips, *Megalurothrips sjostedti* Trybom (Thysanoptera: Thripidae). Efforts to control these pests remain a challenge and there is a need to understand the structure and movement of these pest populations in order to facilitate the development of integrated pest management (IPM) strategies. Molecular tools have the potential to help facilitate a better understanding of pest populations. Towards this goal, we used 454 pyrosequencing technology to generate 319,126, 176,262, 320,722 and 227,882 raw reads from *A. curvipes*, *A. craccivora*, *C. tomentosicollis* and *M. sjostedti*, respectively. The reads were *de novo* assembled into 11,687, 7,647, 10,652 and 7,348 transcripts for *A. curvipes*, *A. craccivora*, *C. tomentosicollis* and *M. sjostedti*, respectively. Functional annotation of the resulting

[4] This chapter has been published as Agunbiade et al. 2013. Development of reference transcriptomes for the major field insect pests of cowpea: A toolbox for insect pest management approaches in West Africa. *PLoS ONE* 8(11): e79929. It is an open access article and is available at www.plosone.org and DOI: 10.1371/journal.pone.0079929. I acknowledge the contribution of co-authors to the publication.

transcripts identified genes putatively involved in insecticide resistance, pathogen defense and immunity. Additionally, sequences that matched the primary aphid endosymbiont, *Buchnera aphidicola*, were identified among *A. craccivora* transcripts. Furthermore, 742, 97, 607 and 180 single nucleotide polymorphisms (SNPs) were respectively predicted among *A. curvipes*, *A. craccivora*, *C. tomentosicollis* and *M. sjostedti* transcripts, and will likely be valuable tools for future molecular genetic marker development. These results demonstrate that Roche 454-based transcriptome sequencing could be useful for the development of genomic resources for cowpea pest insects in West Africa.

INTRODUCTION

Crops of cowpea (*Vigna Unguiculata* (L). Walp) provide a major nutritional source of protein for about 200 million people in sub-Saharan Africa (Cork et al. 2009). Cowpea production is highest in the West African countries of Nigeria, Niger and Burkina Faso, where insect feeding damage by over 100 pest species is a major constraint on field production and in grain storage (Cork et al. 2009). Yield is most dramatically affected by insect pests that occur during the flowering and seed pod stages. These include flower and pod feeding insects such as flower thrips, *Megalurothrips sjostedti* Trybom (Thysanoptera: Thripidae); legume pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae); pod sucking insects, *Clavigralla tomentosicollis* Stål (Hemiptera: Coreidae) (Adati et al. 2007) and *Anoplocnemis curvipes* Fabricius (Hemiptera: Coreidae); and phloem-feeding cowpea aphids, *Aphis craccivora* Koch (Hemiptera: Aphididae). Crop damage by these insect pests can be as high as 60 to 100% in the field (Singh and Allen 1980; Jackai and Daoust 1986; Dugje et al. 2009). *Aphis craccivora* can cause significant

damage even at low population densities due to its ability to transmit at least 14 viruses including the potyviruses, the cowpea aphid-borne mosaic virus (CABMV) (Thottappilly and Rossel 1985; Atiri et al. 1986) and the blackeye cowpea mosaic virus (BICMV) (Dijkstra et al. 1987). These viruses produce severe cowpea mottling, chlorosis, and seed shriveling (Brunt et al. 1990) which severely reduce yields (Bock and Conti 1974; Singh and van Emden 1979). In contrast to most plant viruses, which fail to cross into developing embryos from infected maternal tissues (Bennett 1969), CABMV and BICMV appear to propagate via vertical transmission from parent to progeny seed (Gillaspie et al. 1993; Bashir and Hampton 1996) and is exacerbated by horizontal transmissions by aphid vectors.

Much research has been directed towards developing strategies to control the feeding of the legume pod borer, *M. vitrata* on cowpea crops. Past use of chemical insecticides has resulted in increased frequencies of resistance in *M. vitrata* to three classes of insecticides in Nigeria (Ekesi 1999), and unfortunately fits within the paradigm where selection pressures imposed by widespread application of a chemical control agent can oftentimes lead to the evolution of insecticide resistance within targeted pest insect populations (Champ and Dyte 1976; Georghiou and Lagunes-Tejeda 1991). Additionally, chemical insecticides are often financially inaccessible to smallholder farmers in West Africa, and pose serious health and environmental risks when used indiscriminately by untrained applicators (Kamara et al. 2007; Oparaeke 2007). Therefore, recent shifts toward the use of affordable and sustainable biocontrol measures have been initiated within West Africa (Amevoin et al. 2007; Adati et al. 2007). Some of these potential *M. vitrata* control strategies have included the deployment of bio-pesticides (Oparaeke 2006;

Tamò et al. 2012) and the development of a transgenic cowpea that expresses *Bacillus thuringiensis* (*Bt*) toxins (Bean/Cowpea CRSP 2001). Additionally, traps baited with the female *M. vitrata* sex pheromone blends, in a 100:5:5 ratio of (*E*, *E*)-10,12-hexadecadienal, (*E*, *E*)-10,12-hexadecadienol and (*E*)-10-hexadecenal, have been distributed to farmers in Benin and used as a successful early warning tool (Downham et al. 2003). These baited traps have the potential for monitoring seasonal northward *M. vitrata* migrations during the rainy season, but the fidelity of pheromone blend, and trap position and design can affect the accuracy of resulting estimates of population size and route of migration (Bartels et al. 1997; Bradshaw et al. 1983; Lewis and Macaulay 1976). The population genetic structure of *M. vitrata* has been described through the application of next generation sequencing (NGS) and high throughput single nucleotide polymorphism (SNP) genotyping technologies (Margam et al. 2011) as well as microsatellite loci (Agunbiade et al. 2012a). In these aforementioned studies, *M. vitrata* population structure and estimates of gene flow (migration) within West African cowpea production area were assessed. The results of these studies have the potential to enhance integrated pest management (IPM) programs to determine the logical locations of natural enemy releases. This prior research on *M. vitrata* serves as a model for the application of genome-based approaches to increase the effectiveness of strategies used to control pest insect populations.

Since transgenic cowpea that express *Bacillus thuringiensis* (*Bt*) Cry1Ab toxin shows no toxicity towards non-Lepidopteran insects and a majority of biocontrol strategies for *M. vitrata* are species-specific, cowpea crops remain susceptible to continued feeding and plant disease transmission by thrips, aphids and pod sucking pests.

Thus, the only method widely available to date for the control of *A. craccivora*, *A. curvipes*, *C. tomentosicollis* and *M. sjostedti* by indigenous farmers in West Africa has been the application of chemical insecticides. Since insecticide resistance had evolved within *M. vitrata* populations (Ekesi 1999), the establishment of effective insect resistance management (IRM) plans for *A. craccivora*, *A. curvipes*, *C. tomentosicollis* and *M. sjostedti* may be critical for delaying the evolution of resistance. The absence of population genetic data for these species hinders the estimation of movement patterns of these insects within their endemic range such that the regional scales necessary for IRM programs to remain effective are difficult to devise. NGS technologies that include Roche 454 GS FLX, Solexa/Illumina Genome Analyzer, ABI/SOLiD Gene Sequencers and Helicos Genetic Analysis System platforms use massively parallel pyrosequencing technologies to collect millions of nucleotide sequences in very short time frames (Margulies et al. 2005; Moore et al. 2006; Wicker et al. 2006; Huse et al. 2007; Weber et al. 2007). Moreover, NGS technologies provide a rapid and cost-effective way to obtain large amounts of DNA sequence data from organisms where no prior information had existed (Margam et al. 2011). *De novo* transcriptome analysis has proven to be a valuable first step to obtaining sequence information and expression levels of genes involved in developmental and metabolic pathways, insecticide resistance, and to discover single nucleotide polymorphisms (SNPs) in all kinds of model and non-model organisms (Sloan et al. 2012; Xue et al. 2010; Mittapalli et al. 2010; Poelchau et al. 2011). SNPs are changes of a single nucleotide at a specific location within the genome of a species, and high-throughput assays have been developed for their detection and application as genetic markers (Tang et al. 1999; Vignal et al. 2002; Brumfield et al. 2003; Morin et al. 2004;

Schlötterer 2004). Estimation of allelic frequency variation at SNP loci are effective for describing population demographics (Coates et al. 2011) and are increasingly becoming the marker of choice in population genetic analysis.

In this study, we applied Roche 454 sequencing technology to generate and subsequently assemble contigs from DNA sequencing reads from independent normalized cDNA libraries for *A. curvipes*, *A. craccivora*, *C. tomentosicollis* and *M. sjostedti*. Annotations of individual gene transcripts were used to identify candidate genes putatively involved in insecticide resistance, regulation of insect growth and response to disease transmission. This is the first report of genomic data for these insect pests and provides valuable tool for understanding molecular gene functions of several major field insect pests in cowpea cropping systems of West Africa. The application of this genomics data might ultimately lead to a better understanding of the pest populations, with the long-term potential to improve the effectiveness of IPM programs by better defining pest-pathogen interactions, and pest population dynamics prior to deployment of biocontrol agents.

MATERIALS AND METHODS

Development of reference transcriptome sequence assemblies

Insect samples were collected during the summer through fall of 2011 at 7, 11 and 9 locations in Benin, Burkina Faso and Niger respectively for *A. craccivora*, *A. curvipes, C. tomentosicollis* and *M. sjostedti* (Figure 5.1). A total of 79, 1,920, 364 and 740 individual insect samples were collected respectively for *A. craccivora, A. curvipes, C. tomentosicollis,* and *M. sjostedti* from these locations. Both the larval and adult life stages

were sampled for all species and stored in RNAlater (Ambion, TX, USA) immediately after collection in the field. All samples from each species, from a single location, were pooled and total RNA was extracted from the insect samples at IITA Benin and INERA Burkina Faso using QIAGEN RNeasy RNA extraction kits (CA, USA) and following the manufacturer's instructions. The RNA was shipped to University of Illinois at Urbana-Champaign (UIUC), USA in 70% ethanol where it was resuspended in water and quantified by measuring the absorbance at 260 nm using a NanoDrop spectrophotometer (Thermo Scientific, DE, USA). The samples were then stored in an ultra-low temperature freezer (-80°C).

Four normalized cDNA libraries were constructed and sequenced on a Roche 454 GS-FLX at the W.M. Keck Center for Comparative and Functional Genomics, Roy J. Carver Biotechnology Center, UIUC. Briefly, messenger RNA (mRNA) was isolated from 10µg of total RNA with the Oligotex kit (Qiagen, Valencia, CA). The mRNA-enriched fraction was converted to 454 barcoded cDNA libraries and normalized (Lambert et al. 2010). The barcoded libraries were pooled in equimolar concentration based on average fragment length and concentration. After library construction, the pooled libraries were quantified using a Qubit fluorometer (Invitrogen, CA, USA) and average fragment sizes were determined by analyzing 1µl of the samples on the Bioanalyzer (Agilent, CA, USA) using a DNA 7500 chip. The pooled library was diluted to 1 x 10^6 molecules/µl. Emulsion-based clonal amplification and sequencing on a full plate on the 454 Genome Sequencer FLX+ system was performed according to the manufacturer's instructions (454 Life Sciences, CT, USA). Signal processing and base calling were performed using the bundled 454 Data Analysis Software v2.6.

The raw sequence read data from the four insect pests were analyzed using the CLC Genomics Workbench 6.0.1 (Cambridge, MA, USA). Pre-processing of the raw reads from each of the four insect samples involved trimming each 454 read using a Phred quality score of 20 and also removing nucleotides < 50 bp from the ends. The adapter sequences were also trimmed from the raw reads. The processed read data from each of the four insect samples were assembled into contiguous sequences using parameters: mismatch cost = 2, insertion and deletion cost = 3, length fraction = 60% and similarity = 90%. After assembly, the vector contamination were removed using the UniVec database and also after assembly, human, bacterial, fish (*Danio rerio*), mouse (*Mus musculus*), *Salmonella enterica*, archeal and viral contamination were removed using a web-based version of DeConSeq (Schmieder and Edwards 2011) using a coverage of 90% and a sequence identity threshold of 94%. The clean transcriptomes, with the contaminations removed, were then deposited at DDBJ/EMBL/GenBank for each of the four insect species.

Functional gene annotation

Open reading frames (ORFs) were predicted from assembled contigs using the ORF-Predictor server (Min et al. 2005) using all 6 possible reading frames for prediction. The assembled transcripts were used as queries to search against NCBI's non-redundant (nr) database using the Blastx algorithm (Altschul et al. 1990), with a cut-off *E*-value of $\leq 10^{-6}$ and a high scoring segment pairs (HSP) length cut-off of 33. The Blast2GO software package 2.6.5 (Ashburner et al. 2000) was used for automating Blastx searches as well as to retrieve associated gene ontology (GO) terms that allowed the prediction of transcript

functions (Ashburner et al. 2000; Shaw et al. 1999). The contigs with significant GO terms were determined with an *E*-value hit filter of $\leq 1 \times 10^{-6}$ and an annotation cut off of 55. Gene ontologies were categorized with respect to molecular function, biological process, and cellular component.

Annotations for *A. craccivora*, *A. curvipes*, *C. tomentosicollis*, and *M. sjostedti* gene functions were manually searched for those putatively involved in the expression of insecticide resistance traits, and pathogen defense and immunity, using each gene function as keywords to search GO terms. Prediction of candidate gene function was also obtained using InterProScan (Hunter et al. 2009; Quevillon et al. 2005) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses (Kanehisa and Goto 2000) using Blast2GO 2.6.5 (Ashburner et al. 2000).

Prediction of putative SNPs

The associated SNP detection software on the CLC Workbench 6.0.1 was used for putative SNP discovery among Roche 454 reads for all species. Attempts to reduce the rate of false SNP discovery included applying a read coverage cut-off of ≥ 35-fold and reporting SNPs that were present in ≥ 35% of the aligned reads. These criteria might reduce the false SNP discovery rate by potentially eliminating sequencing errors from the prediction. However, such stringent criteria likely increases type II error, therefore we also performed a prediction of putative SNPs using a reduced coverage cut off of ≥ 10%. All putative indels and nucleotide variants involving > 2 nucleotides were excluded. Lastly, only SNPs located in an ORF were extracted and reported in this study. We checked whether SNPs introduced an amino acid change to differentiate non-synonymous and

synonymous SNPs by using the open reading frames of each of the contigs with SNPs, identifying the codons containing the SNPs and then translating and comparing the amino acids for each allele on CLC Workbench 6.0.1. We also checked the type of substitution, whether transition or transversion, using the CLC Workbench 6.0.1.

Metagenomic identification of endosymbiont and pathogen transcripts

The bacterial endosymbiont, *Buchnera aphidicola*, was identified from a Blastn search against *Buchnera* (Taxid: 32199) in NCBI using the assembled *A. craccivora* contigs as queries. The *A. craccivora* contigs with relevant hits to *Buchnera* were extracted and further confirmatory analysis was performed using InterProScan on Blast2GO 2.6.5, and the associated KEGG pathways were investigated also using Blast2GO 2.6.5.

Data deposition

The raw Roche 454 sequence data were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) with accession numbers of SRR768514, SRR768515, SRR768524 and SRR768525 for *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti*, respectively. The Transcriptome Shotgun Assembly project for the four insect species were submitted to DDBJ/EMBL/GenBank under the accession numbers of GAJV00000000, GAJW00000000, GAJX00000000, and GAJY00000000 for *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti*, respectively. The version of the TSA accession numbers described in this paper is the first version for each of the four species.

RESULTS

Development of reference transcriptome sequence assemblies

Normalized species-specific libraries were successfully constructed from mRNA isolated from pooled samples of all tissues and pooled adult and larval life stages. A total of 319,126, 176,262, 320,722 and 227,882 raw reads were respectively obtained from these *A. curvipes, A. craccivora, C. tomentosicollis,* and *M. sjostedti* libraries. The bacterial and human contaminants discovered by DeConSeq were negligible across the four insect species. Seventeen, 2, 11 and 37 contigs were identified by DeConSeq as contaminants in *A. curvipes, A. craccivora, C. tomentosicollis,* and *M. sjostedti,* respectively and were also subsequently removed from the transcripts. The remaining statistics for sequence and contig assemblies are reported in Table 5.1. After assembly and decontamination, the mean contig length ranged from 669.8 to 688.1 bp, from which ORFs with a mean length of between 498.5 and 524.5 bp was predicted (Table 5.1).

Functional gene annotation

ORFs were predicted from ≥ 98% of *A. curvipes, A. craccivora, C. tomentosicollis,* and *M. sjostedti* transcripts, and respectively showed mean lengths of 498.46, 514.28, 524.54 and 508.78 bp (Table 5.1). Blast2GO output indicated that Blastx hits were obtained for 6,430 (55.02%), 7,647 (79.94%), 6,839 (64.19%) and 4,292 (58.41%) contigs in *A. curvipes, A. craccivora, C. tomentosicollis,* and *M. sjostedti,* respectively. The contigs with significant Blastx matches were assigned GO terms into molecular function, biological process, and cellular components (Appendix F1 to F4). In the molecular function category, the most highly represented were assigned to binding

(13.66% for *A. curvipes*, 13.64% for *A. craccivora*, 13.37% for *C. tomentosicollis*, and 13.47% for *M. sjostedti*), and catalytic activity (14.37% for *A. curvipes*, 13.09% for *A. craccivora*, 14.93% for *C. tomentosicollis*, and 13.47% for *M. sjostedti*). In the biological process category, the most highly represented were assigned to cellular process (4.65% for *A. curvipes*, 4.68% for *A. craccivora*, 4.72% for *C. tomentosicollis*, and 4.74% for *M. sjostedti*), and metabolic process (5.36% for *A. curvipes*, 5.40% for *A. craccivora*, 5.80% for *C. tomentosicollis*, and 5.63% for *M. sjostedti*) while in the cellular component category, the most highly represented were assigned to cell (9.27% for *A. curvipes*, 9.14% for *A. craccivora*, 9.45% for *C. tomentosicollis*, and 9.02% for *M. sjostedti*), and cell part (8.20% for *A. curvipes*, 8.03% for *A. craccivora*, 8.31% for *C. tomentosicollis*, and 8.15% for *M. sjostedti*) (Appendix F1 to F4).

The majority of the top Blastx hits in the four insect species were from insects. In *A. curvipes*, the most common were from Hemiptera [*Riptortus pedestris* (23.61%)], Coleoptera [*Tribolium castaneum* (7.05%)], Phthiraptera [*Pediculus humanus* (6.72%)], and another Hemiptera [*Acryrthosiphon pisum* (5.99%)] (Appendix G1). The most frequent hits from *A. craccivora* were from Hemiptera [*A. pisum* (83.67%)], two fungi species [*Rhizopus delemar* (1.52%); *Batrachochytrium dendrobatidis* (0.65%)], and Phthiraptera [*P. humanus* (0.54%)] (Appendix G2). The most frequent in *C. tomentosicollis* were from Hemiptera [*R. pedestris* (27.72%)], Coleoptera [*T. castaneum* (7.72%)], another Hemiptera [*A. pisum* (7.00%)], and Phthiraptera [*P. humanus* (6.65%)] (Appendix G3) while the most frequent in *M. sjostedti* transcripts were from Coleoptera [*T. castaneum* (9.74%)], Phthiraptera [*P. humanus* (8.36%)], Lepidoptera [*Danaus plexippus* (6.45%)], and Hymenoptera [*Nasonia vitripennis* (5.62%)] (Appendix G4).

Blastx hits also revealed matches to other fungi species, bacteria, and a plant among *A. craccivora* contigs (Appendix G2), but these were not analogously observed within libraries from the other three insects.

Within the libraries constructed from *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti* cDNA, a combined total of 23 candidate genes for detoxification, immunity and pathogen defense, development and communication were identified including cytochrome P450, glutathione s-transferase, esterase, cathepsin, heat shock protein, chitinase, defensin, c-Jun NH (2)-terminal kinase (jnk) stimulatory phosphatase, down syndrome critical region protein, epidermal growth factor, lysozyme, nimrod, nitric acid synthase, prophenol oxidase, ubiquinol cytochrome c reductase, peptidoglycan recognition protein, toll protein, chemosensory protein, juvenile hormone inducible protein, juvenile hormone esterase and juvenile hormone epoxide hydrolase, chemosensory binding protein as well as odorant binding protein (Figures 5.3a to 5.3d).

Metagenomic identification of endosymbiont- and pathogen-derived transcripts

Thirty six Blastn hits to the NCBI nr protein database were identical to the primary endosymbiont of aphids, *B. aphidicola*, when transcripts from *A. craccivora* were used as queries of which 23 unique transcripts retrieved InterProScan annotations (Appendix H) and nine were predicted to be involved in nine different bacterial biochemical pathways (Table 5.2). Transcripts from six different fungi species were also predicted among *A. craccivora* transcripts, including *Rhizopus delemar*, and *Batrachochytrium dendrobatidis* (Appendix G2).

Prediction of putative SNPs

All *A. curvipes*, *A. craccivora*, *C. tomentosicollis,* and *M. sjostedti* contigs, that contained a putative ORF, were included in the SNP prediction pipeline (Appendix I1 to I4, Appendix J1 to J4). From these predictions, 258, 30, 225, and 63 contigs respectively from *A. curvipes*, *A. craccivora*, *C. tomentosicollis,* and *M. sjostedti* had putative SNPs, which respectively contained a total of 742, 97, 607, and 180 putative SNPs (Appendix I1 to I4, Appendix J1 to J4). The mean depth of reads aligned to the reference transcripts depth for all putative SNPs was > 60 across all species (Table 5.3). The density of SNPs within transcripts was measured by estimates of mean number of putative SNPs per kilobase, and were 0.09, 0.02, 0.08, and 0.04 respectively among *A. curvipes*, *A. craccivora*, *C. tomentosicollis* and *M. sjostedti* contigs (Table 5.3). The alternate ≥ 10% coverage cut off we used comparatively predicted 2,703, 340, 2,087, and 780 putative SNPs for *A. curvipes*, *A. craccivora*, *C. tomentosicollis,* and *M. sjostedti*, respectively. As a consequence of 454-based sequence by synthesis methods used, resulting reads are prone to sequencing errors known as homopolymers, which comprise imprecise nucleotide numbers in long arrays of the same nucleotide. These errors can cause misalignment within contig assemblies such that incorrect SNP predictions can result in sequence regions flanking the homopolymer stretch. To compensate for these errors, the variants were also filtered for 454/Ion homopolymer INDELS in the SNP detection software in the CLC Genomics Workbench. To understand the effects of SNP mutations and associate them to the different transcripts obtained in our study, we differentiated synonymous SNPs from non-synonymous SNPs. Of the total number of SNPs obtained in

the four insect species, there were 425, 72, 419, and 97 synonymous SNPs respectively in *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti* and respectively 317, 25, 188, and 83 non-synonymous SNPs. In all four species, transitions were more frequent than transversions (Ts/Tv > 1) (Table 5.3).

DISCUSSION

Development of reference transcriptome sequence assemblies

Next generation sequencing technologies offer a rapid entry point into genomic research (Coates et al. 2011) and can generate valuable molecular resources for non-model species (Sloan et al. 2012; Harismendy et al. 2009) that are a foundation from which a diversity of research questions can be addressed (Harismendy et al. 2009; Rothberg and Leamon 2008). In the absence of complete genome sequences, transcriptome sequencing remain a useful molecular resource that can be applied to the identification of candidate insecticide resistance genes and mutations that can be developed into genetic markers for population genetic studies (Margam et al. 2011), as well as the identification of potential targets for RNAi knockdown. The Roche 454 platform provides long sequence read lengths that may better allow the assembly of *de novo* transcriptomes (Harismendy et al. 2009), but remain susceptible to sequencing errors in homopolymer regions. The Roche 454 transcriptome data presented in this study from *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti* had a high median length (≥ 680 bp), and a majority of resulting contigs encoded a predicted ORF (= protein coding sequence, CDS). Despite this, a high number of contigs (> 35%) had ORFs with no amino acid similarity to known proteins within the NCBI nr database and was especially the case for *A. curvipes*, *C.*

tomentosicollis, and *M. sjostedti* transcripts. However, only 20.06% of the *A. craccivora* contigs had no similarity to known proteins in the NCBI nr database. Annotation of previous transcriptome assemblies have similarly revealed a high number of contigs with genes of unknown function (Karatolos et al. 2011; Bai et al. 2011; Wang et al. 2010; Shen et al. 2011) which may represent novel uncharacterized genes and reflect the limitation of inferring transcript functions by comparison to model species that have long evolutionary distances to the non-model species in question (Coates et al. 2008). Even within whole genome sequence assemblies, species-specific genes can comprise a high percentage of predicted ORFs (Tribolium Genome Sequencing Consortium 2008). The presence of these genes of unknown function could similarly suggest these proteins may be species-specific and, that *de novo* transcriptome assemblies from non-model insect pest species are useful for phylogenetic novel gene discovery. Furthermore, the resulting assembly of sequence data allow for the identification of novel gene pathways that have potentials for RNAi targeting within a suite of species-specific control tactics.

Functional annotation

Functional annotations of assembled *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti* transcripts allowed for the identification of candidate genes encoding proteins putatively involved in insecticide resistance, and pathogen defense and immunity. Transcriptomic approaches are powerful tools to identify new genes and gene functions and have been successfully applied to many organisms. In this study, we have identified genes putatively involved in the response to and the detoxification of xenobiotics in the four insect pests. Some of these xenobiotic response/detoxification genes will likely be

useful for the study of chemical insecticide resistance traits as well as their role in detoxification following exposure to plant allelochemicals. For example, strains of *A. craccivora* have elevated esterase activities that were linked to increased resistance to the nicotinic acetylcholine receptor agonist, dinotefuran, which belongs to the third generation of neonicotinoids (Mokbel and Mohamed 2009). Our results will provide a foundation that makes the future study of the involvement of these candidate genes in field-observed insecticide resistance traits in these insects more likely, and may also represent genetic markers that can be used to screen field populations (and compare resistant vs. susceptible individuals) to determine linkage (or not) of the locus to the resistant phenotype trait.

Our current understanding of insect immunity and stress responses comes from holometabolous insects and includes flies, butterflies, beetles, and bees (Gerardo et al. 2010). The four insect pests under study in this paper are all hemimetabolous insects with three of them (*A. curvipes*, *A. craccivora,* and *C. tomentosicollis*) falling into the same insect order of Hemiptera and the fourth insect *M. sjostedti* falling into the insect order, Thysanoptera. Because all studied species exhibit incomplete development, comparison with the genome of a hemimetabolous insect (i.e., pea aphid, *A. pisum)* may provide insights into immunity and defense mechanisms in these pest insects. It is also interesting to note that while the four insect species included in this study were not intentionally immunologically challenged, we still observed some transcripts putatively involved in insect defense and immunity based on studies conducted on other insects such as *A. pisum*. We did not observe as many immunity and defense transcripts in both *A. craccivora* and *M. sjostedti* as we observed in *A. curvipes* and *C. tomentosicollis*. The immune genes observed in this study include most genes involved in the IMD pathway in

insects and includes chitinase, defensin, down syndrome critical region protein, epidermal growth factor receptor, jnk stimulatory phosphatase, lysozyme, nimrod, nitric oxide synthase, odorant binding protein, peptidoglycan-recognition protein and pro-phenol oxidase. We also observed genes involved in toll signaling pathway. It is interesting to note that none of these genes are represented across all the four insect pests. Some insects have particular genes that others lack, and vice versa. For example, *A. craccivora* appears to be missing the defensin gene, however, a lack of such a gene would have to be verified in the future if a genomic project were to occur for this species. This is consistent with studies conducted on *A. pisum,* which shows the pea aphid is lacking many of the antimicrobial peptides, such as defensin, common to other insects (Zou et al. 2007). The reduced humoral immune system in *A. pisum*, including an apparently non-functional IMD signaling pathway and absence of PGRPs, has been suggested to be an adaptation for the symbiosis with the bacterium *B. aphidicola* (Douglas 1998). The presence of defensin in the human louse, *P. humanus* and in the ancient apterygote insect, the fire brat, *Thermobia domestica* (Altincicek and Vilcinskas 2007), suggests that defensins may have been lost during aphid evolution.

Prediction of putative SNPs

Single nucleotide polymorphisms are rapidly becoming the marker of choice for many applications in population ecology, evolution and conservation genetics, because of the potential for high genotyping efficiency, data quality, genome-wide coverage and analytical simplicity (e.g. in modeling mutational dynamics) (Morin et al. 2004). Transcriptome-derived SNPs have several advantages over those developed from genomic

sequences (Hayes et al. 2007; Akey et al. 2003; Picoult-Newberg et al. 1999), including acquisition of actual gene sequences that allow for direct mapping and comparative genome studies among organisms (Wang et al. 2008, and references therein). SNPs derived from transcriptomes are also a source of candidate polymorphisms underlying important traits that can lead to the identification of quantitative trait nucleotides (QTN) (Jalving et al. 2004) linked to ecologically relevant genes. The applicability of SNPs from sequence data for marker development has been previously reported (Novaes et al. 2008; Wiedmann et al. 2008; Williams et al. 2010) and has been applied for the genetic mapping of insect orders such as Lepidoptera [*Bombyx mori*; Yamamoto et al. 2006)] and for population genetics of the Glanville fritillary butterfly, *Melitaea cinxia* (Orsini et al. 2008). The current study provides a set of at least 742, 97, 607, and 180 putative SNPs respectively for *A. curvipes, A. craccivora, C. tomentosicollis,* and *M. sjostedti* (predicted using the criteria that SNPs be present in ≥ 35% of aligned reads), and the segregating mutations that can be developed into molecular genetic markers for the study of the population genetic structure of these insect pests. Although a greater number of putative SNPs were predicted using a more lenient coverage cut off value of 10%, these loci may be prone to type I error and secondary validation methods may likely be required to distinguish these from sequencing errors.

The frequency of SNPs in laboratory strains of *Drosophila* was reported at 5 SNP per kilobase (Berger et al. 2001) and at 1.3 SNPs per kilobase in the inbred Dazao strain of *Bombyx mori* (Cheng et al. 2004). Similarly, laboratory strains of the malaria mosquito, *Anopheles funestus,* were reported to have 7.2 SNPs per kilobase (Wondji et al. 2007), and 8.0 SNPs per kilobase in *An. gambiae* (Morlais et al. 2004). Compared to the

results obtained from the above studies, we did observe a lower amount of SNPs per kilobase in the present study. Although laboratory strains were used in those studies, we used field-collected insects in our study and usually a reduced SNP frequency is reported in laboratory strains because homozygosity may be increased by the effects of inbreeding or random genetic drift. Non-synonymous SNPs have commonly been reported to occur less frequently than synonymous SNPs, and is presumably due to the evolutionary constraints of negative selection that may eliminate deleterious substitutions from the population (Cargill et al. 1999). Non-synonymous SNPs are of particular interest because they are more likely to affect the function of the encoded protein and may influence phenotype. It has been estimated that 20–30% of non-synonymous SNPs affect protein function (Sunyaev et al. 2001; Chasman and Adams 2001). In our study, we did observe a higher number of synonymous SNPs than non-synonymous SNPs across transcripts from all the four insect species. In metazoan DNA sequences, an excess of transition vs. transversion mutations is often observed. This may be partly due to the relatively high rate of change of methylated cytosines to thymine, as well as post-mutation processes of selection on codon-usage bias within coding regions (Keller et al. 2007). The role of population genetic and biochemical effects on the rate and direction of nucleotide changes remains unknown, but are likely factors that affect the observed level of SNP allele frequencies within natural populations.

Metagenomic identification of endosymbiont- and pathogen-derived transcripts

Aphids are sap-feeding insects that infest a wide range of plant species. Although sap fluids from plant phloem contain high concentrations of carbohydrates, they are

deficient in nitrogenous nutrients such as specific amino acids (Sandström and Moran 1999; Houk and Griffiths 1980). To overcome these nutritional deficiencies, species within Aphidoidea have established mutualistic relationships with the obligate intracellular endosymbiont, *B. aphidicola* (Lai et al. 1994; Munson et al. 1991). *Buchnera* endosymbionts produce essential amino acids that cannot be synthesized by aphids or obtained in sufficient quantities from plant saps (Lai et al. 1994; Douglas 1998). In return, aphids provide *Buchnera* with other nutrients required to survive (Wilkinson and Douglas 1995). Relationships between these two groups have existed for approximately 150 to 200 million years (Moran et al. 1993) resulting in drastic *Buchnera* genome reductions due to the loss of many genes needed for independent life and has led to the inability to survive outside host cells (Shigenobu et al. 2000). Therefore, aphids and associated *Buchnera* symbionts may be inseparable mutualistic partners. We observed 36 *B. aphidicola* transcripts among Blastn hits to our *A. craccivora* transcripts. These 36 *B. aphidicola* transcripts were annotated from twelve different strains of *B. aphidicola*. We also observed ubiquinone in the Blastx search of *A. craccivora*. Symbiotic *B. aphidicola* are aerobic bacteria, which due to gene reduction in metabolic pathways, cannot carry out respiration without obtaining gene products from the host (Zientz et al. 2004). The electron transport chain consists of a primary dehydrogenase and a terminal reductase, which are linked by ubiquinone (Unden and Bongaerts 1997). Ubiquinone is an essential redox component of the aerobic respiration of bacteria and mitochondria (Søballe and Poole 2000), and participates in the transfer of electrons and hydrogen between flavoproteins and cytochrome b in the respiratory chain. Also, one of the 36 *A. craccivora* contigs with hits to *B. aphidicola* was annotated as the gene symbionin, which has been

reported to increase the transmission of plant viruses by binding to the read-through domain of the viral coat protein (Banerjee et al. 2004). Further study of these genes may likely lead to a better understanding of symbiosis and plant disease transmission by *A. craccivora*, and may lead to potential tactics to reduce or eliminate the disease vectoring capacity of *A. craccivora*.

Additionally, we observed Blastx hits to six fungi species among *A. craccivora* sequences. *Rhizopus oryzae* (*R. oryzae* has been reclassified to include *R. oryzae* and *R. delemar* (Abe et al. 2007). *Rhizopus delemar* was observed in the Blastx hits in this study and was previously reported to be an entomopathogenic fungal species (Sharma et al. 2012). *Batrachochytrium dendrobatidis* causes chytridiomycosis and is a major cause of amphibian population decline worldwide (Friesen and Kuhn 2012) and the sequences within our aphid transcripts may be derived from a related fungal species that is capable of infecting *A. craccivora* in West Africa. In contrast, *Melampsora larici-populina* is a cause of rust in poplar trees (Yu et al. 2009). These may have resulted from environmental contamination or were present within gut contents of whole aphids that were used for library preparations. These results suggest that application of NGS may be used for the metagenomic identification of putative pathogen species, which in turn may be useful for the biological control of pest insect species.

CONCLUSION

With the exception of prior studies that focused on *M. vitrata*, this study represents the first attempt to develop transcriptomic and molecular marker data for field insect pests of cowpea in West Africa. Although the sequence data, and biological functions of these

genes, may be of interest and importance to molecular biologists, the molecular markers are potentially of much greater near-term pragmatic importance for those interested in controlling these pests. Previous studies have already demonstrated that such molecular markers can give us important insights into pest movement patterns that ultimately will impact how pest control strategies for *M. vitrata* need to be developed in different agro-ecological zones in West Africa (Margam et al. 2011; Agunbiade et al. 2012a). For example, *M. vitrata* is an endemic pest in the southern part of the selected West African countries and migratory in the northern part; thus biocontrol agents need to be released in the south and spraying of pesticides or biopesticides (Tamò et al 2012) may be a better solution in areas where these insects are not endemic. This understanding has emerged from a combination of studies on the biology of this pest and through the use of molecular markers. We term this approach, combining traditional IPM strategies with knowledge that emerges from population genetics/genomics tools, IPM-omics (Agunbiade et al. 2012b). This study lays the foundation for research in other pest species, with the long-term goal to develop a comprehensive program that integrates genomics datasets into IPM and IRM programs in order to minimize the crop damage inflicted by pest insect species of cowpea in West Africa.

REFERENCES

Abe A, Oda Y, Asano K, Sone T (2007) *Rhizopus delemar* is the proper name for *Rhizopus oryzae* fumaric-malic acid producers. *Mycologia* 99: 714–722.

Adati T, Tamò M, Yusuf SR, Downham MCA, Singh BB, Hammond W (2007) Integrated pest management for cowpea–cereal cropping systems in the West African savannah. *International Journal of Tropical Insect Science* 27: 123-137.

Agunbiade TA, Kim KS, Forgacs D, Margam VM, Murdock LL, Ba NM, Binso-Dabire CL, Baoua I, Ishiyaku MF, Tamò, Pittendrigh BR (2012a) The spatial genetic differentiation of the legume pod borer, *Maruca vitrata* F. (Lepidoptera: Crambidae) populations in West Africa. *Bulletin of Entomological Research* 102(5): 589-599.

Agunbiade T, Steele L, Coates B, Gassmann A, Margam V, Ba M, Dabire-Binso C, Baoua I, Bello-Bravo J, Seufferheld F, Sun W, Tamò, Pittendrigh BR (2012b) IPM-omics: From genomics to extension for integrated pest management of cowpea. In Boukar O, Coulibaly O, Fatokun C, Lopez K, Tamò M (Eds.), *Improving livelihoods in the cowpea value chain through advancements in science.* Proceedings of the 5th World Cowpea Research Conference, pp. 231 – 248.

Akey JM, Zhang K, Xiong M, Jin L (2003) The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium. *Molecular Biology and Evolution* 20 (2): 232-242.

Altincicek B, Vilcinskas A (2007) Identification of immune-related genes from an apterygote insect, the firebrat *Thermobia domestica*. *Insect Biochemistry and Molecular Biology* 37: 726-731.

Altschul SF, Warren G, Webb M, Eugene WM, David JL (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.

Amevoin K, Sanon A, Apossaba M, Glitho IA (2007) Biological control of bruchids infesting cowpea by the introduction of *Dinarmus basalis* (Rondani) (Hymenoptera: Pteromalidae) adults into farmers' stores in West Africa. *Journal of Stored Products Research* 43(3): 240–247.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: Tool for the unification of biology. *Nature Genetics* 25: 25–29.

Atiri GI, Enobakhare DA, Thottappilly A (1986) The importance of colonizing and non-colonizing aphid vectors in the spread of cowpea aphid-borne mosaic virus in cowpea. *Crop Protection* 5: 406-410.

Bai X, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O (2011) Transcriptomics of the Bed Bug (*Cimex lectularius*). *PLoS ONE* 6(1): e16336.

Banerjee S, Hess D, Majumder P, Roy D, Das S (2004) The interactions of *Allium sativum* leaf agglutinin with a chaperonin group of unique receptor protein isolated from a bacterial endosymbiont of the mustard Aphid. *Journal of Biological Chemistry* 279(22): 23782–23789.

Bartels DW, Hutchison WD, Udayagiri S (1997) Pheromone trap monitoring of Z-strain European corn borer (Lepidoptera: Pyralidae): Optimum pheromone blend, comparison with black light trap, and trap number requirements. *Journal of Economic Entomology* 90: 449-457.

Bashir M, Hampton RO (1996) Detection and identification of seed-borne viruses from cowpea [*Vigna unguiculata* (L.) Walp.] germplasm. *Plant Pathology* 4: 54-58.

Bean/Cowpea CRSP (2001) The Dakar symposium/workshop on the genetic improvement of cowpea. Jan. 8-12.

Bennett CW (1969) Seed transmission of plant viruses. *Advances in Virus Research* 14: 221–262.

Berger J, Suzuki T, Senti KA, Stubbs J, Schaffner G, Dickson BJ (2001) Genetic mapping with SNP markers in *Drosophila*. *Nature Genetics* 29(4): 475-481.

Bock KR, Conti M (1974) Cowpea aphid-borne mosaic virus, descriptions of plant viruses, No. 134. Commonwealth Mycological Institute and Association of Applied Biologists. Kew, Surrey, UK, 4pp.

Bradshaw JWS, Baker R, Longhurst C, Edwards JC, Lisk JC (1983) Optimization of a monitoring system for the pine beauty moth, *Panolis flammea* (Denis & Schiffermüller), using sex attractants. *Crop Protection* 2: 63–73.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18 (5): 249-256.

Brunt A, Crabtree K, Gibbs A (1990) Viruses of tropical plants. C.A.B. International, Wallingford, UK, pp. 97-102.

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* 22(3): 231-238.

Champ BR, Dyte CE (1976) Report of the FAO global survey of pesticide susceptibility of stored grain pests. FAO Plant Production Project. Paper No. 5. FAO/UN, Rome.

Chasman D, Adams RM (2001) Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. *Journal of Molecular Biology* 307(2): 683-706.

Cheng TC, Xia QY, Qian JF, Liu C, Lin Y, Zha XF, Xiang ZH (2004) Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain Dazao. *Insect Biochemistry and Molecular Biology* 34: 523-530.

Coates BS, Bayles DO, Wanner KW, Robertson HM, Hellmich RL, Sappington TW (2011) The application and performance of single nucleotide polymorphism markers for population genetic analyses of Lepidoptera. *Frontiers in Research* 2 (38): 1 -10.

Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2008) Mining an *Ostrinia nubilalis* midgut expressed sequence tag (EST) library for candidate genes and single nucleotide polymorphisms (SNPs). *Insect Molecular Biology* 17: 607–620.

Cork A, Dobson H, Grzywacz D, Hodges R, Orr A, Stevenson P (2009) Review of pre- and post-harvest pest management for pulses with special reference to Eastern and Southern Africa. Natural Resources Institute, University of Greenwich, Central Avenue, Chatham Maritime, Kent ME4 4TB, UK. Prepared for and funded by McKnight Foundation. Collaborative Crops Research Program.

Dijkstra J, Bos L, Bouwmeester HJ, Hadiastono T, Lohuis H (1987) Identification of blackeye cowpea mosaic virus from germplasm of yard-long bean and from soybean, and the relationships between blackeye cowpea mosaic virus and cowpea aphid-borne mosaic virus. *Netherlands Journal of Plant Pathology* 93(3): 115-133.

Douglas AE (1998) Nutritional interactions in insect-microbe symbioses: Aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology* 43: 17–37.

Downham MCA, Tamò M, Hall DR, Datinon B, Adetonah S, Farman DI (2003) Developing pheromone traps and lures for *Maruca vitrata* in Benin, West Africa. *Entomologia Experimentalis et Applicata* 110(2): 151-158.

Dugje IY, Omoigui LO, Ekeleme F, Kamara AY, Ajeigbe H (2009) Farmers' guide to cowpea production in West Africa. IITA Ibadan, Nigeria, 20pp.

Ekesi S (1999) Insecticide resistance in field populations of the legume pod-borer, *Maruca vitrata* Fabricius (Lepidoptera: Pyralidae), on cowpea, *Vigna unguiculata* (L.) Walp in Nigeria. *International Journal of Pest Management* 45: 57-59.

Friesen LR, Kuhn RE (2012) Fluorescent microscopy of viable *Batrachochytrium dendrobatidis*. *Journal of Parasitology* 98(3): 509-512.

Georghiou GP, Lagunes-Tejeda A (1991) The occurrence of resistance to pesticides in arthropods. FAO, Rome.

Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, Duncan EJ, Evans JD, Gabaldón T, Ghanim M, Heddi A, Kaloshian I, Latorre A, Moya A, Nakabachi A, Parker BJ, Pérez-Brocal V, Pignatelli M, Rahbé Y, Ramsey JS, Spragg CJ, Tamames J, Tamarit D, Tamborindeguy C, Vincent-Monegat C, Vilcinskas A (2010) Immunity and other defenses in pea aphids, *Acyrthosiphon*

pisum. *Genome Biology* 11: R21.

Gillaspie AG Jr, Hopkins MS, Pinnow DL (1993) Relationship of cowpea seed-part infection and seed transmission of blackeye cowpea mosaic potyvirus in cowpea. *Plant Disease* 77(9): 875-877.

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Lecy S, Frazer KA (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology* 10: R32.

Hayes J, Laerdahl JK, Lien S, Moen T, Berg P, Hindar K, Davidson WS, Koop BF, Adzhubei A, Høyheim B(2007) An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. *Aquaculture* 265 (1-4): 82-90.

Houk EJ, Griffiths GW (1980) Intracellular symbiotes of the Homoptera. *Annual Review of Entomology* 25: 161–187.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2009) InterPro: The integrative protein signature database. *Nucleic Acids Research* 37: D211–D215.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively-parallel DNA pyrosequencing. *Genome Biology* 8: R143.

Jackai LEN, Daoust RA (1986) Insect pests of cowpeas. *Annual Review of Entomology* 31: 95–119.

Jalving R, van't Slot R, van Oost BA (2004) Chicken single nucleotide polymorphism identification and selection for genetic mapping. *Poultry Science* 83(12): 1925-1931.

Kamara AY, Chikoye D, Omoigui LO, Dugje IY (2007) Cultivar and insecticide spraying regimes effects on insect pests and grain yield of cowpea in the dry savannas of north-eastern Nigeria. *African Crop Science Conference Proceedings* 8: 179-184.

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28: 27–30.

Karatolos N, Pauchet Y, Wilkinson P, Chauhan R, Denholm I, Gorman K, Nelson DR, Bass C, ffrench-Constant RH, Williamson MS (2011) Pyrosequencing the transcriptome of the greenhouse whitefly, *Trialeurodes vaporariorum* reveals multiple transcripts encoding insecticide targets and detoxifying enzymes. *BMC Genomics* 12: 56.

Keller A, Zhuang H, Chi Q, Vosshall LB, Matsunami H (2007) Genetic variation in a human odorant receptor alters odour perception. *Nature* 449: 468-472.

Lai CY, Baumann L, Baumann P (1994) Amplification of trpEG: Adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proceedings of the National Academy of Sciences USA* 91(9): 3819-3823.

Lambert DJ, Yi Chan Xin, Spiecker B, Sweet HC (2010) Characterizing the embryonic transcriptome of the snail, Ilyanassa. *Integrative and Comparative Biology* 50(5): 768-777.

Lewis T, Macaulay EDM (1976) Design and elevation of sex attractant traps for pea moth, *Cydia nigricana* (Steph.) and the effect of plume shape on catches. *Ecological Entomology* 1: 175–187.

Margam VM, Coates BS, Bayles DO, Hellmich RL, Agunbiade T, Seufferheld MJ, Sun W, Kroemer JA, Ba MN, Binso-Dabire CL, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR (2011) Transcriptome sequencing, and rapid development and application of SNP markers for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE* 6(7): e21388.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 326–327.

Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research* 33: W677–W680.

Mittapalli O, Bai X, Mamidala P, Rajarapu SP, Bonello P, Herms DA (2010) Tissue-specific transcriptomics of the exotic invasive insect pest Emerald Ash Borer (*Agrilus planipennis*). *PLoS ONE* 5: e13708.

Mokbel ES, Mohamed AI (2009) Development of resistance in field strain of *Aphis craccivora* to the dinotefuran insecticides from the new class neonicotinoids and its effect on some enzymes. *Egyptian Academic Journal of Biological Sciences* 1(1): 65-69.

Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, Soltis DE (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology* 6: 17.

Moran NA, Munson MA, Baumann P, Ishikawa H (1993) A molecular clock in

endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society London B* 253: 167–171.

Morin PA, Luikart G, Wayne RK, SNP-Workshop_Grp (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* 19: 208-216.

Morlais I, Ponçon N, Simard F, Cohuet A, Fontenille D (2004) Intraspecific nucleotide variation in *Anopheles gambiae*: New insights into the biology of malaria vectors. *American Journal of Tropical Medicine and Hygiene* 71(6): 795-802.

Munson MA, Baumann P, Kinsey MG (1991) *Buchnera* gen. nov. and *Buchnera aphidicola* sp. Nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *International Journal of Systematic Bacteriology* 41(4): 566-568.

Novaes ED, Drost R, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9: 312.

Oparaeke AM (2006) The potential for controlling *Maruca vitrata* Fab. and *Clavigralla tomentosicollis* Stal. using different concentrations and spraying schedules of *Syzigium aromaticum* (L.) Merr and Perr on cowpea plants. *Journal of Plant Science* 1: 132–137.

Oparaeke AM (2007) Toxicity and spraying schedules of a biopesticide prepared from *Piper guineense* against two cowpea pests. *Plant Protection Science* 43: 103–108.

Orsini L, Corander J, Alasentie A, Hanski I (2008) Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. *Molecular Ecology* 17: 2629–2642.

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. *Genome Research* 9: 167–174.

Poelchau MF, Reynolds JA, Denlinger DL, Elsik CG, Armbruster PA (2011) A *de novo* transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. *BMC Genomics* 12: 619.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiller R, Lopez R (2005) InterProScan: Protein domains identifier. *Nucleic Acids Research* 33: W116–W120.

Rothberg JM, Leamon JH (2008) The development and impact of 454 sequencing. *Nature Biotechnology* 26: 1117–1124.

Sandström J, Moran NA (1999) How nutritionally imbalanced is phloem sap for aphids? *Entomologia Experimentalis et Applicata* 91(1): 203-210.

Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion?

Nature Reviews Genetics 5(1): 63-69.

Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* 6: e17288.

Sharma A, Chandla VK, Thakur DR (2012) Biodiversity and pathogenicity potential of mycoflora associated with *Brahmina coriacea* in potato fields of North-Western Indian Hills. *Journal of Entomology* 9: 319-331.

Shaw DR, Ashbumer M, Blake JA, Baldarelli RM, Botstein D, Davis AP, Cherry JM, Lewis S, Lutz CM, Richardson JE, Eppig JT (1999) Gene ontology: A controlled vocabulary to describe the function, biological process and cellular location of gene products in genome databases. *American Journal of Human Genetics* 65(4): A419-1419.

Shen GM, Dou W, Niu JZ, Jiang HB, Yang WJ, Jia FX, Hu F, Cong L, Wang J-J (2011) Transcriptome analysis of the Oriental fruit fly (*Bactrocera dorsalis*). *PLoS ONE* 6: e29127.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407(6800): 81-86.

Singh SR, Allen DJ (1980) Pests, diseases, resistance and protection in cowpeas. In Summerfield RJ, Bunting AH. *Advances in Legume Science*, Royal Botanical Gardens, Kew, pp. 419–443.

Singh SR, van Emden HF (1979) Insect pests of grain legumes. *Annual Review of Entomology* 24: 255–278.

Sloan DB, Keller SR, Berardi AE, Sanderson BJ, Karpovich JF, Taylor DR (2012) *De novo* transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources* 12: 333–343.

Søballe B, Poole RK (2000) Ubiquinone limits oxidative stress in *Escherichia coli*. *Microbiology* 146: 787–796.

Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P (2001) Prediction of deleterious human alleles. *Human Molecular Genetics* 10(6): 591-597.

Tamò M, Srinivasan R, Dannon E, Agboton C, Datinon B, Dabire C, Baoua I, Ba M, Haruna B, Pittendrigh BR (2012) Biological control: A major component for the long-term cowpea pest management strategy. In Boukar O, Coulibaly O, Fatokun C, Lopez K, Tamò M (Eds.), *Improving livelihoods in the cowpea value chain through advancements in science*. Proceedings of the 5th World Cowpea Research Conference, pp. 249-259.

Tang K, Fu D-J, Julien D, Braun A, Cantor CR, Köster H (1999) Chip-based genotyping by mass spectrometry. *Proceedings of the National Academy of Sciences USA* 96:10016–10020.

Thottappilly G, Rossel HW (1985) Virus diseases of important food crops in tropical Africa. Ibadan, Nigeria, International Institute of Tropical Agriculture pp 35.

Tribolium Genome Sequencing Consortium (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452: 949-955.

Unden G, Bongaerts J (1997) Alternative respiratory pathways of *Escherichia coli*: Energetics and transcriptional regulation in response to electron acceptors. *Biochimica et Biophysica Acta* 1320: 217–234.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 34: 275-305.

Wang S, Sha Z, Sonstegard TS, Liu H, Xu P, Somridhivej B, Peatman E, Kucuktas H, Liu Z (2008) Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics* 9: 450.

Wang X-W, Luan J-B, Li J-M, Bao Y-Y, Zhang C-X, Liu S-S (2010) *De novo* characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics* 11: 400.

Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiology*144: 32-42.

Wicker T, Schlagenhauf E, Graner A, Close TJ, Keller B, Stein N (2006) 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* 7: 275.

Wiedmann RT, Smith TPL, Nonneman DJ (2008) SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genetics* 9: 81.

Wilkinson TL, Douglas AE (1995) Why pea aphids (*Acyrthosiphon pisum*) lacking symbiotic bacteria have elevated levels of the amino acid glutamine. *Journal of Insect Physiology* 41(11): 921-927.

Williams LM, Ma X, Boyko AR, Bustamante CD, Oleksiak MF (2010) SNP identification, verification, and utility for population genetics in a non-model genus. *BMC Genetics* 11: 32.

Wondji CS, Morgan J, Coetzee M, Hunt RH, Steen K, Black WC, Hemingway J, Ranson H (2007) Mapping a quantitative trait locus (QTL) conferring pyrethroid resistance in the African malaria vector *Anopheles funestus*. *BMC Genomics* 8: 34.

Xue J, Bao YY, Li BL, Cheng YB, Peng ZY, Liu H, Xu HJ, Zhu ZR, Lou YG, Cheng JA, Zhang CX (2010) Transcriptome analysis of the brown planthopper, *Nilaparvata lugens*. *PLoS ONE* 5: e14233.

Yamamoto K, Zhang P, Banno Y, Fujii H (2006) Identification of a sigma-class glutathione-S-transferase from the silkworm, *Bombyx mori*. *Journal of Applied*

Entomology 130: 515–522.

Yu Z, Liang J, Cao Z, Guo Z, Dan J, Zhao G (2009) Nuclear behavior in the life cycle of *Melampsora larici-populina* Kleb. *Journal of Food, Agriculture and Environment* 7 (3 and 4): 791-794.

Zientz E, Dandekar T, Gross R (2004) Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiology and Molecular Biology Reviews* 68(4): 745-770.

Zou Z, Evans JD, Lu Z, Zhao P, Williams M, Sumathipala N, Hetru C, Hultmark D, Jiang H (2007) Comparative genomic analysis of the *Tribolium* immune system. *Genome Biology* 8: R177.

FIGURES AND TABLES

Figure 5.1. Map of Benin, Burkina Faso, and Niger showing the sites from which *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti* were collected.

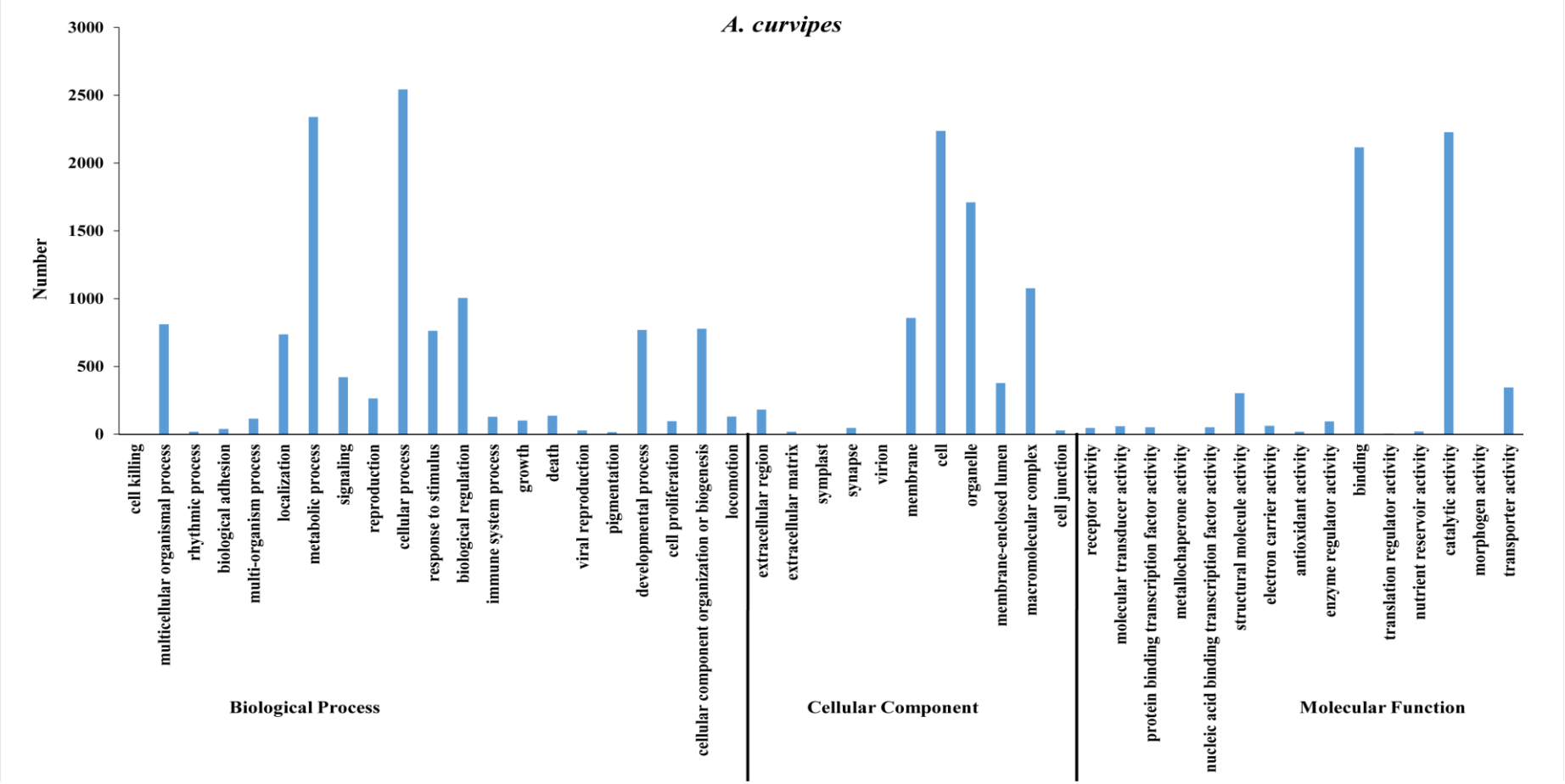


Figure 5.2a. Gene ontology classification into biological process, cellular component, and molecular function. Gene ontology terms were determined using an e-value of ≤ 1.0 e^{-6} and sorted based on level 2 classifications in all the contigs of *A. curvipes.*

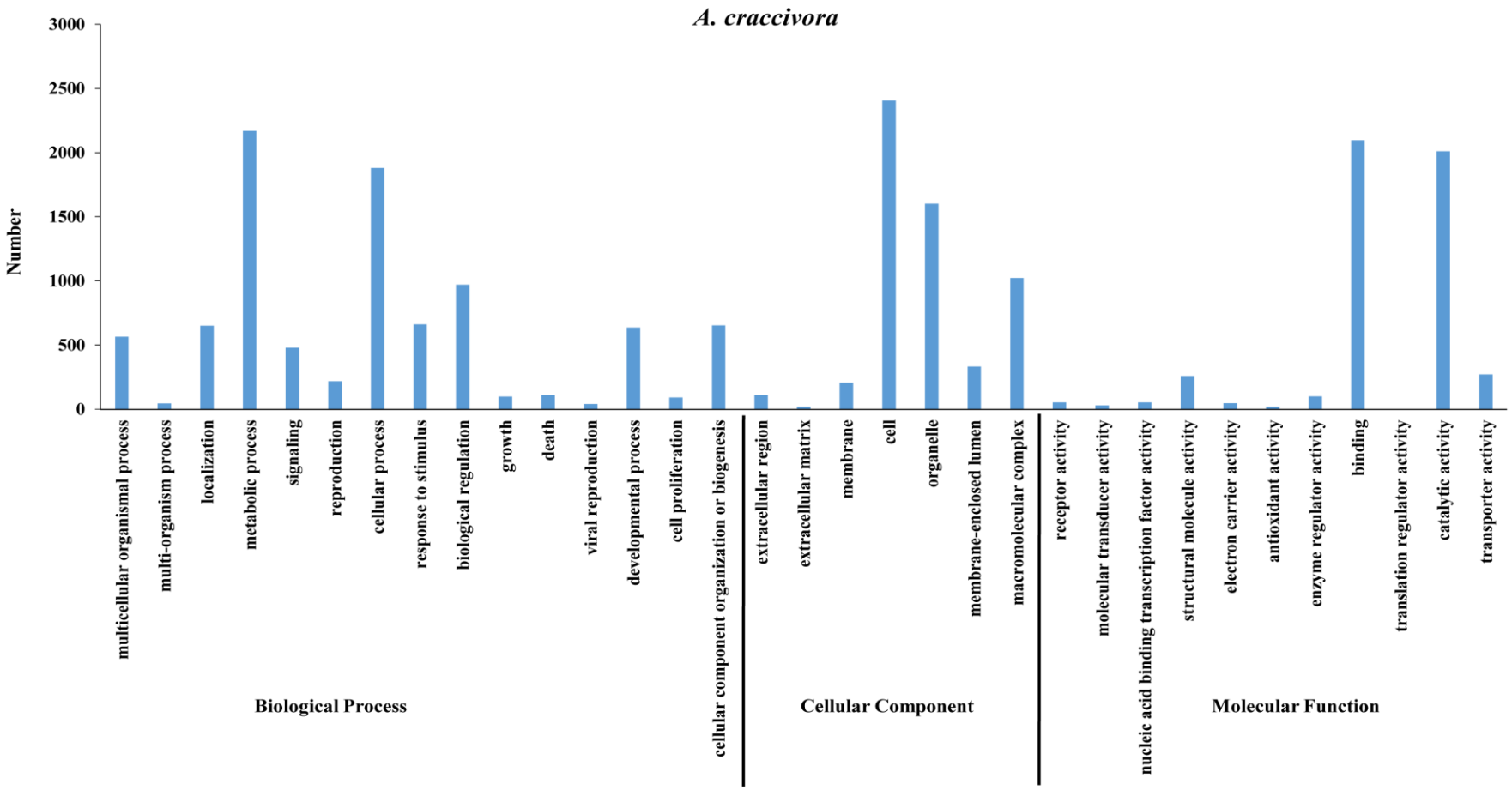


Figure 5.2b. Gene ontology classification into biological process, cellular component, and molecular function. Gene ontology terms were determined using an e-value of ≤ 1.0 e^{-6} and sorted based on level 2 classifications in all the contigs of *A. craccivora.*

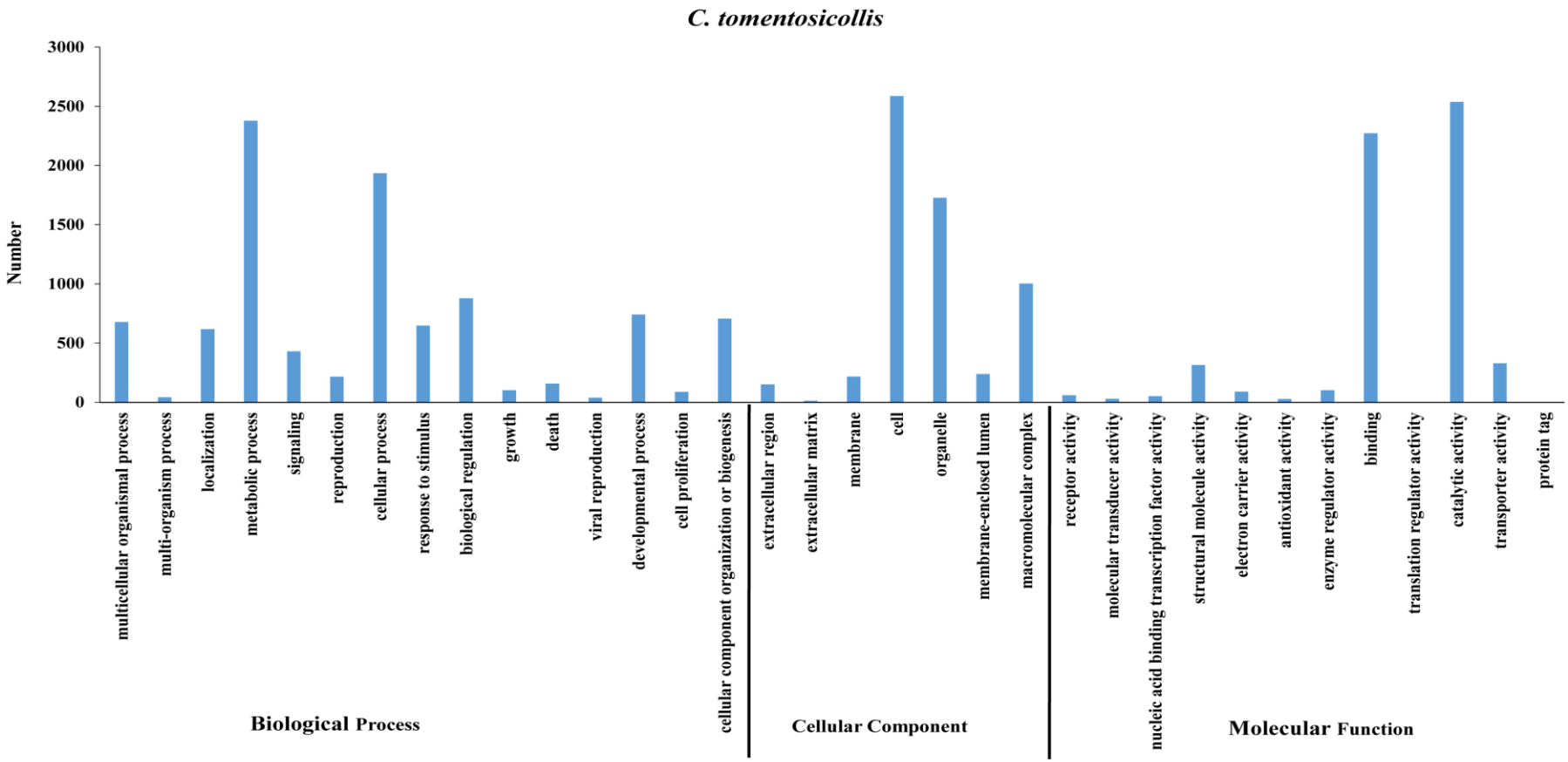


Figure 5.2c. Gene ontology classification into biological process, cellular component, and molecular function. Gene ontology terms were determined using an e-value of ≤ 1.0 e^{-6} and sorted based on level 2 classifications in all the contigs of *C. tomentosicollis.*

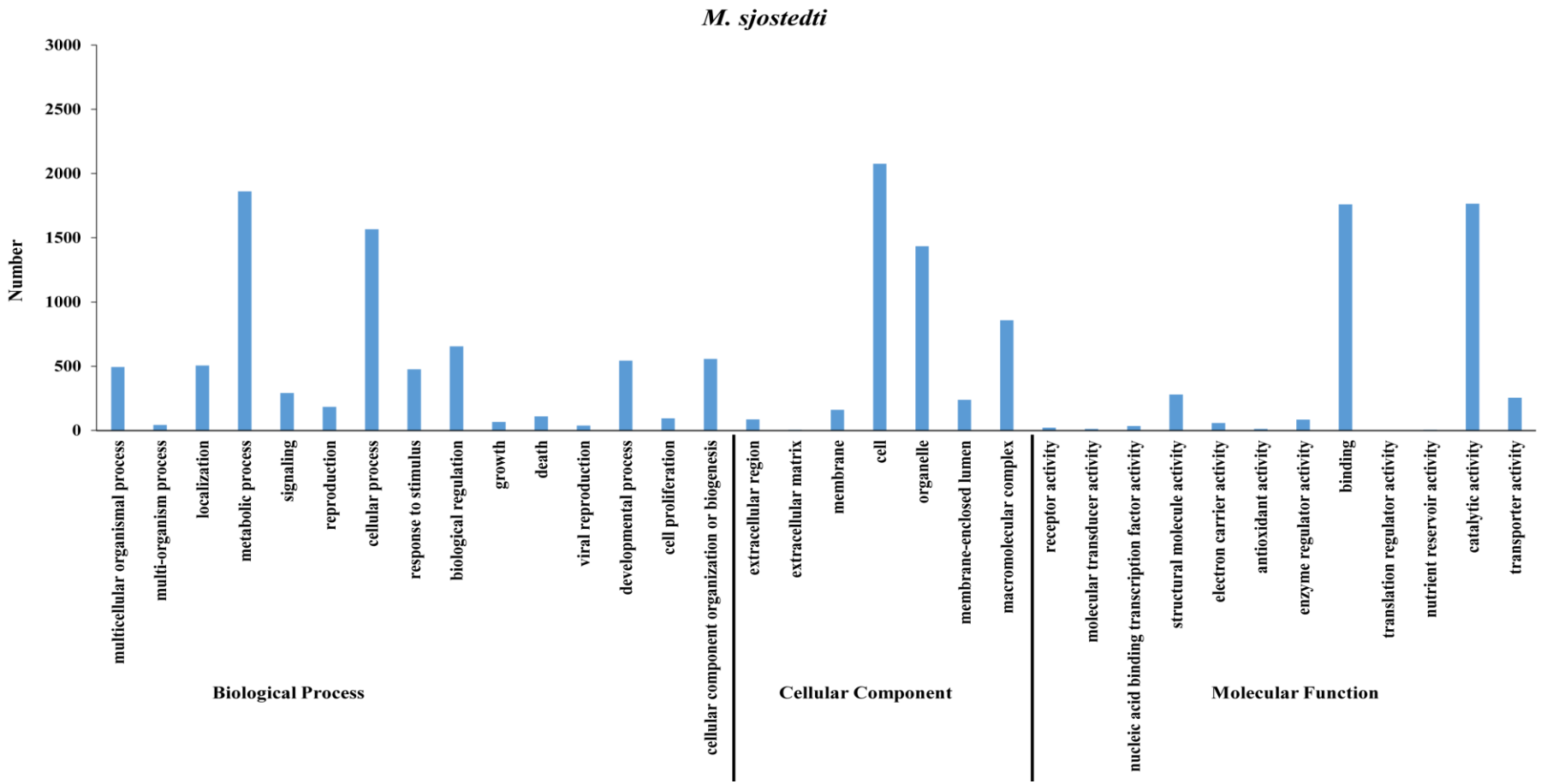


Figure 5.2d. Gene ontology classification into biological process, cellular component, and molecular function. Gene ontology terms were determined using an e-value of ≤ 1.0 e^{-6} and sorted based on level 2 classifications in all the contigs of *M. sjostedti.*

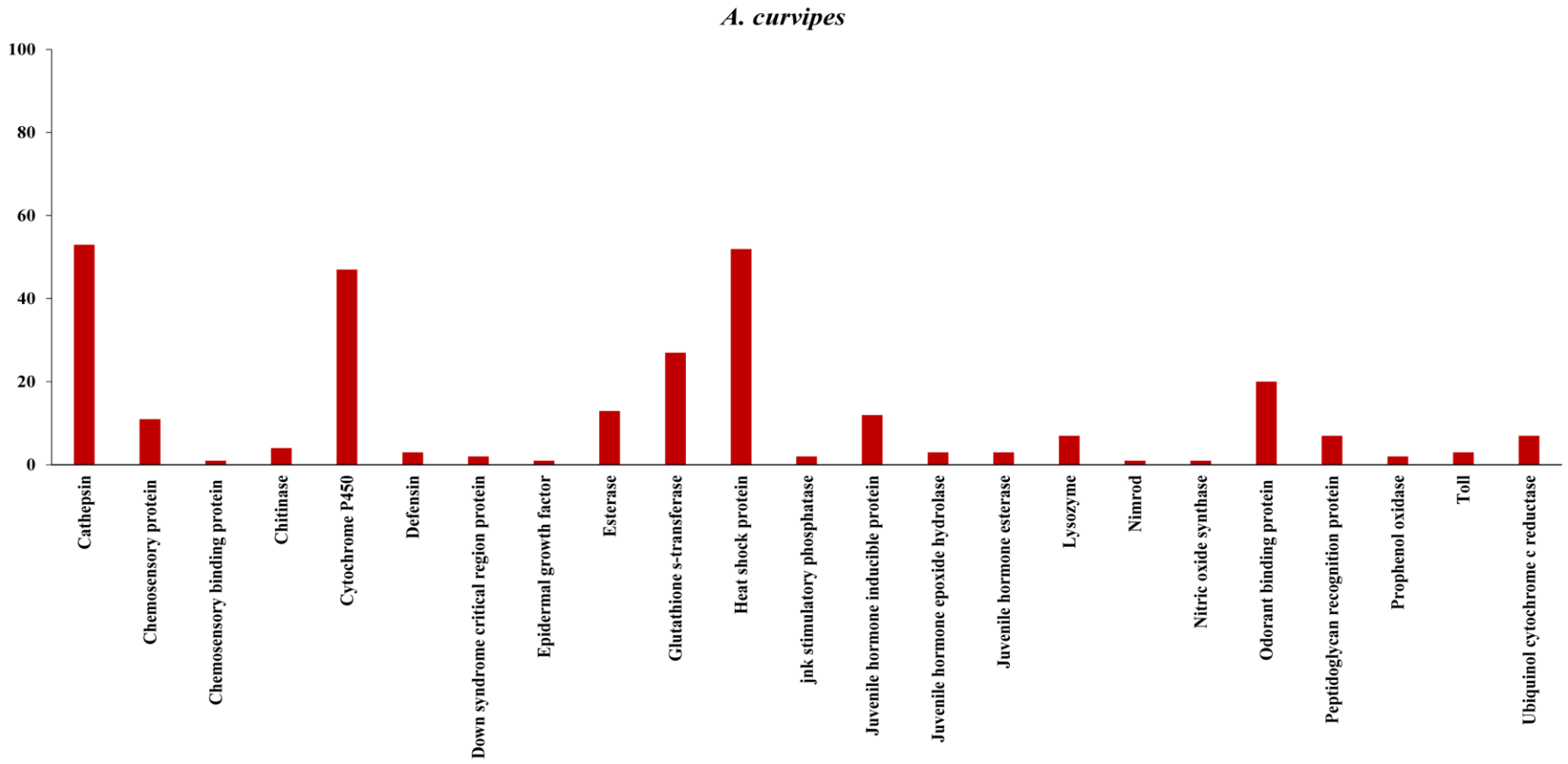


Figure 5.3a. Transcripts putatively involved in responses to xenobiotic (e.g., insecticide resistance), and disease transmission in *A. curvipes.*

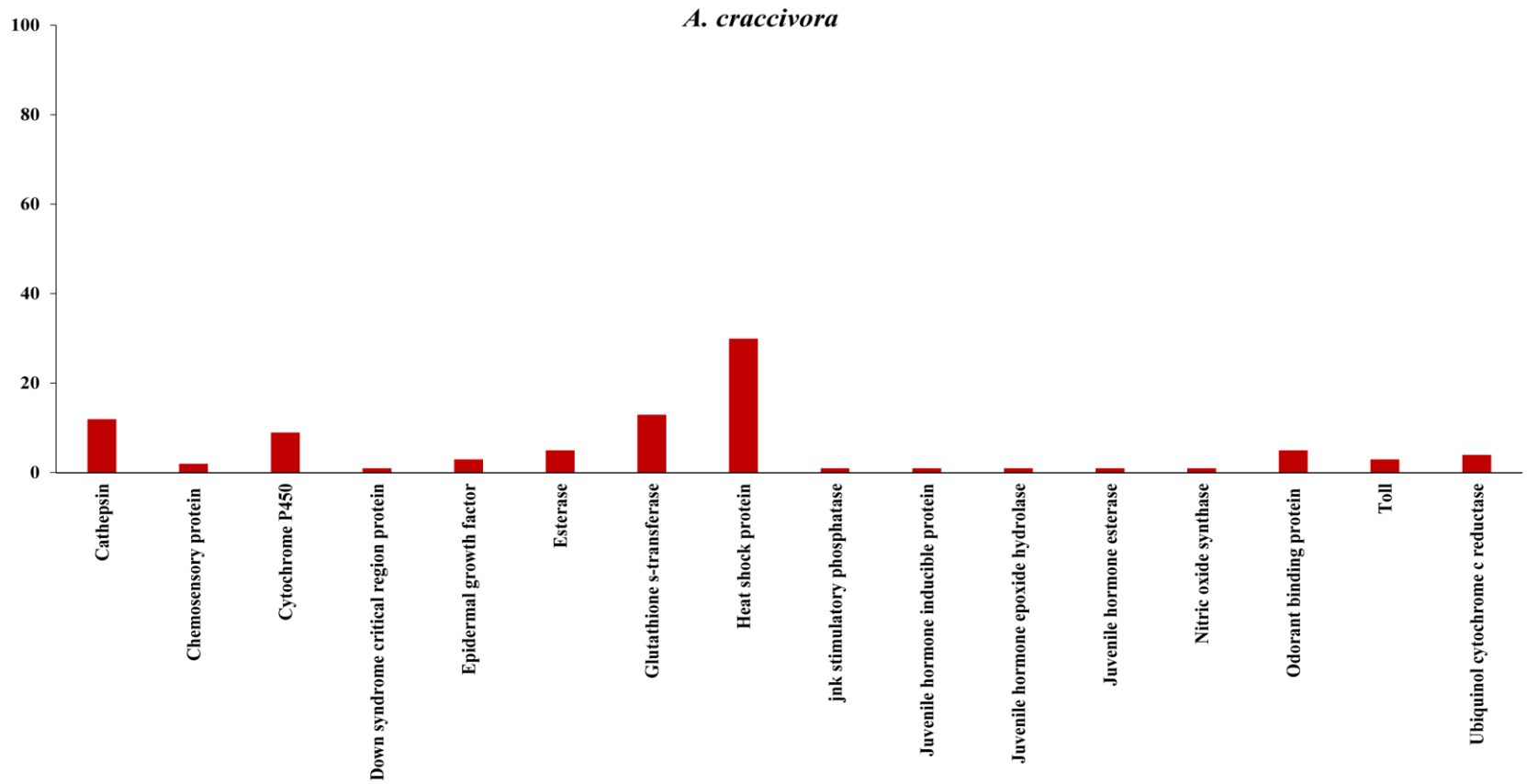


Figure 5.3b. Transcripts putatively involved in responses to xenobiotic (e.g., insecticide resistance), and disease transmission in *A. craccivora.*

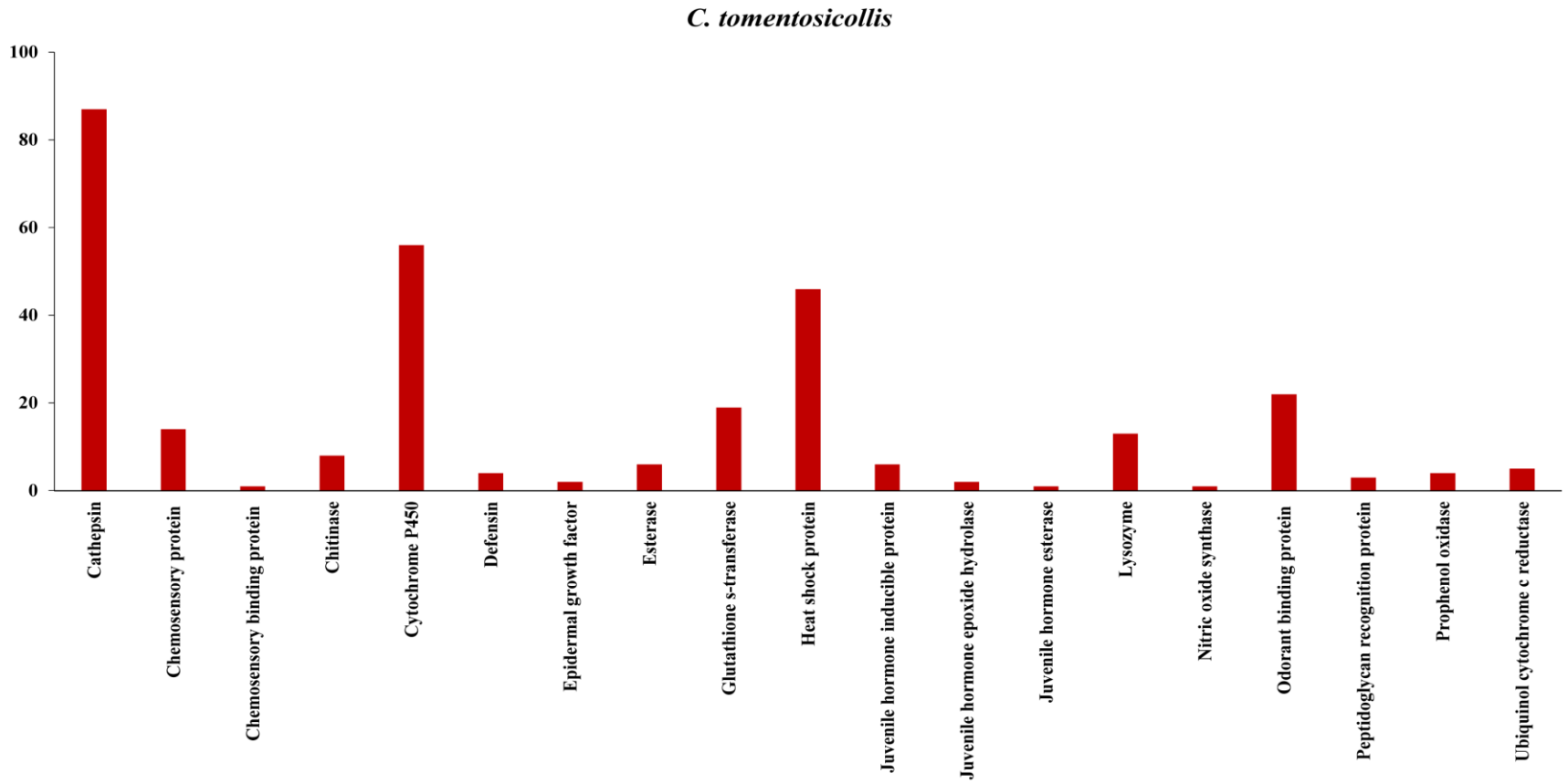


Figure 5.3c. Transcripts putatively involved in responses to xenobiotic (e.g., insecticide resistance), and disease transmission in *C. tomentosicollis.*

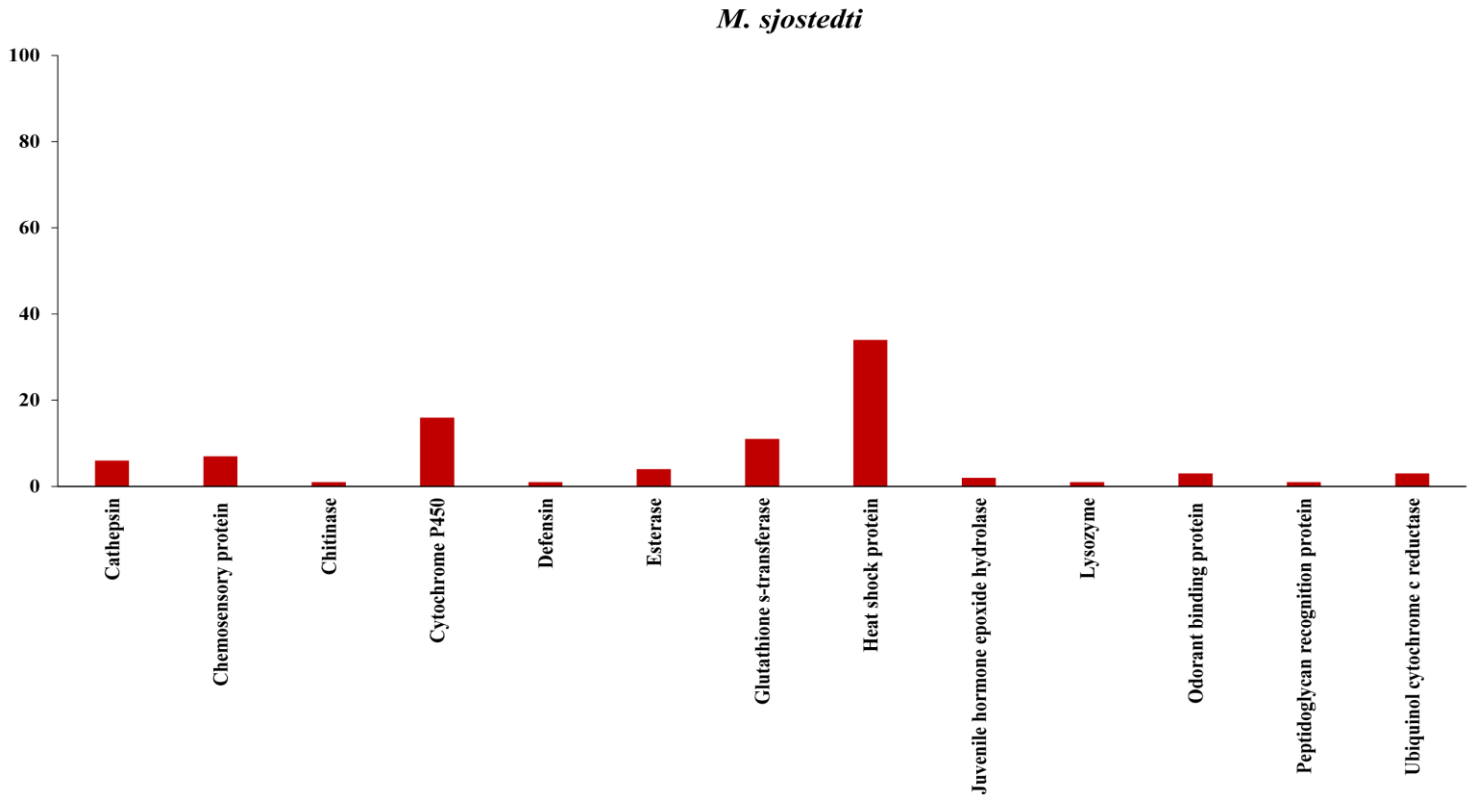


Figure 5.3d. Transcripts putatively involved in responses to xenobiotic (e.g., insecticide resistance), and disease transmission in *M. sjostedti.*

Table 5.1. Statistics from Roche 454 sequencing of *A. curvipes, A. craccivora, C. tomentosicollis* and *M. sjostedti* cDNA libraries generated from pools of all tissues and the larval and adult life stages.

	A. curvipes	*A. craccivora*	*C. tomentosicollis*	*M. sjostedti*
Putative SNPs identified	742	97	607	180
Contigs with SNPs	256	30	225	63
Transition	505	65	423	116
Transversion	237	32	184	64
Transition/Transversion Ratio	2.1	2	2.3	1.8
SNPs per Kilobase	0.09	0.02	0.08	0.04
Synonymous SNPs	425	72	419	97
Non-synonymous SNPs	317	25	188	83
Mean read depth of SNPs	97.5	66.6	115.5	74.9

Table 5.2. The orthologs of *A. craccivora* contigs derived from the genome of the primary aphid endosymbiont, *B. aphidicola* (identified in GenBank accession BA000003.2.). Information regarding protein function can be retrieved from SwissProt database (http://enzyme.expasy.org/) by searches for EC number (not available for all *B. aphidicola* genes).

Contig ID	Orthologous *B. aphidicola* gene	*B. aphidicola* protein (EC)
Aphis 1561	D-fructose-6-phosphate amidotransferase (glmS)	BAB12753.1 (EC 2.6.1.16)
Aphis 5691	UDP-N-acetylglucosamine pyrophosphorylase (glmU)	BAB12754.1 (EC 2.7.7.23)
Aphis 5225	S-adenosylmethionine synthetase (metK)	BAB13109.1 (EC 2.5.1.6)
Aphis 7020	acetolactate synthase small subunit (ilvH)	BAB12941.1 (EC 2.2.1.6)
Aphis 6159	2-oxoglutarate dehydrogenase e1 component (sucA)	BAB13011.1 (EC 1.2.4.2)
Aphis 5021	ABC transporter ATP-binding protein (uup)	BAB13068.1
Aphis 376	Spermidine synthase (speE)	BAB12926.1 (EC 2.5.1.16)
Aphis 3768	6-phsphoglucanate dehydrogenase (gnd)	BAB12826.1 (EC 1.1.1.44)
Aphis 3870	Hypothetical GTP-binding protein (yfgK)	BAB13291.1

Table 5.3. Summary of the putative single nucleotide polymorphism (SNP) predictions from reads mapped to reference *A. curvipes*, *A. craccivora*, *C. tomentosicollis*, and *M. sjostedti* transcripts.

	A. curvipes	*A. craccivora*	*C. tomentosicollis*	*M. sjostedti*
Normalization	Normalized	Normalized	Normalized	Normalized
Total number of raw reads	319,126	176,262	320,722	227,882
Mean raw read lengths (bp)	382.7	402.1	389.6	391.5
Total number of processed reads after trimming	304,110	166,565	306,666	211,626
Mean trimmed read length (bp)	315.5	356.5	327.1	340.9
Final processed number of assembled contigs	11,687	7,647	10,652	7,348
Mean length of assembled contig (bp)	688.1	669.8	685.8	683.7
Total number of singletons	219	211	180	115
Mean ORF length (bp)	498.5	514.3	524.5	508.8
Total number of contigs with Blastx hits	6,430	6,113	6,839	4,292

APPENDIX A
Characteristics of the *M. vitrata* individuals in the three West African countries

Population	Locus	N	Na	Ho	He	*F*	Probability
Niger	**444**	72	2.0	0.1	0.1	-0.0	0.1ns
	32008	72	8.0	0.5	0.5	-0.0	0.1ns
	3393	58	1.0	0.0	0.0	#N/A	
	7_02K06	66	4.0	0.0	0.2	0.9	0.0***
	CO241	72	7.0	0.5	0.4	-0.2	0.0***
	CO325	72	3.0	0.2	0.2	0.0	0.1ns
Nigeria	**444**	52	4.0	0.2	0.3	0.2	0.0**
	32008	51	8.0	0.5	0.5	-0.0	0.0***
	3393	41	3.0	0.0	0.1	1.0	0.0***
	7_02K06	52	5.0	0.1	0.3	0.6	0.0***
	CO241	53	4.0	0.2	0.3	0.0	0.0**
	CO325	52	3.0	0.3	0.3	-0.0	0.1ns
Fada	**444**	40	4.0	0.1	0.1	-0.0	0.1ns
	32008	39	4.0	0.5	0.6	0.1	0.0**
	3393	26	2.0	0.0	0.1	1.0	0.0***
	7_02K06	39	3.0	0.1	0.4	0.8	0.0***
	CO241	40	3.0	0.3	0.2	-0.1	0.1ns
	CO325	40	2.0	0.8	0.5	-0.6	0.0***
Farakoba	**444**	86	4.0	0.1	0.1	-0.0	0.1ns
	32008	70	5.0	0.4	0.5	0.3	0.0***
	3393	64	2.0	0.0	0.0	1.0	0.0***
	7_02K06	86	4.0	0.2	0.4	0.6	0.0***
	CO241	86	4.0	0.2	0.2	-0.1	0.1ns
	CO325	86	4.0	0.4	0.4	-0.1	0.1ns
Kamboinse	**444**	49	3.0	0.1	0.1	-0.0	0.1ns
	32008	41	6.0	0.4	0.5	0.2	0.0**
	3393	31	2.0	0.0	0.0	-0.0	0.1ns
	7_02K06	49	3.0	0.1	0.5	0.8	0.0***
	CO241	44	2.0	0.2	0.2	-0.1	0.0**
	CO325	49	2.0	0.4	0.5	0.3	0.0**

Sample size (N), number of alleles (Na), observed heterozygosity (Ho), expected heterozygosity (He), F_{IS} per sample and loci (after Bonferroni adjusted threshold with corresponding *P*-values from ≤0.005 to 0.05 and the probability ($P < 0.001$ - ***, $P < 0.01$ - ** and $P < 0.05$ - *).

APPENDIX B
Characteristics of the *M. vitrata* individuals

Population	Locus	Na	Ne	Ho	He	F_{IS}	Probability	Significance
Oueme-Plateau (*V. unguiculata*)	**C0241**	5	1.16	0.10	0.14	0.25	0.02	*
	7_02K06	2	1.93	0.33	0.48	0.32	0.02	*
	01_B12	2	1.35	0.22	0.26	0.13	0.35	Ns
	32008	6	2.20	0.69	0.54	-0.27	0.31	Ns
	C0444	2	1.08	0.08	0.08	-0.04	0.77	Ns
Zou-Collines (*V. unguiculata*)	**C0241**	5	1.29	0.16	0.22	0.27	0.20	Ns
	7_02K06	2	1.15	0.14	0.13	-0.08	0.59	Ns
	01_B12	2	1.52	0.44	0.34	-0.28	0.05	*
	32008	5	3.00	0.80	0.67	-0.19	0.43	Ns
	C0444	2	1.08	0.08	0.08	-0.04	0.77	Ns
Mono-Couffo (*V. unguiculata*)	**C0241**	3	1.09	0.08	0.08	-0.03	0.99	Ns
	7_02K06	2	1.51	0.10	0.34	0.70	0.00	***
	01_B12	2	1.56	0.43	0.36	-0.19	0.18	Ns
	32008	6	2.66	0.65	0.62	-0.05	0.00	***
	C0444	2	1.14	0.13	0.12	-0.07	0.64	Ns
Oueme-Plateau (*L. sericeus*)	**C0241**	4	1.15	0.09	0.13	0.30	0.00	***
	7_02K06	2	1.02	0.02	0.02	-0.01	0.94	Ns
	01_B12	2	1.97	0.37	0.49	0.25	0.14	Ns
	32008	3	2.37	0.75	0.58	-0.30	0.01	**
	C0444	2	1.02	0.02	0.02	-0.01	0.94	Ns

Number of alleles (*N*a), number of effective alleles (*Ne*), observed heterozygosity (H_O), expected heterozygosity (H_E), fixation index (F_{IS}), and probability per sample site.

APPENDIX B. (cont.)

Zou-Collines (*L. sericeus*)	**C0241**	1	Monomorphic					
	7_02K06	2	1.02	0.02	0.02	-0.01	0.94	Ns
	01_B12	2	2.00	0.40	0.50	0.21	0.17	Ns
	32008	4	2.25	0.89	0.56	-0.60	0.00	***
	C0444	2	1.09	0.09	0.08	-0.05	0.76	Ns
Oueme-Plateau (*P. phaseoloides*)	**C0241**	4	1.09	0.08	0.08	-0.03	1.00	Ns
	7_02K06	1	Monomorphic					
	01_B12	2	1.16	0.11	0.14	0.23	0.12	Ns
	32008	5	2.82	0.71	0.65	-0.11	0.00	***
	C0444	2	1.13	0.13	0.12	-0.07	0.64	Ns
Zou-Collines (*P. phaseoloides*)	**C0241**	2	1.08	0.08	0.08	-0.04	0.77	Ns
	7_02K06	2	1.11	0.02	0.10	0.79	0.00	***
	01_B12	2	1.30	0.18	0.23	0.23	0.12	Ns
	32008	4	2.45	0.75	0.59	-0.27	0.31	Ns
	C0444	2	1.02	0.02	0.02	-0.01	0.94	Ns
Mono-Couffo (*P. phaseoloides*)	**C0241**	3	1.04	0.04	0.04	-0.02	1.00	Ns
	7_02K06	2	1.34	0.13	0.25	0.50	0.00	***
	01_B12	2	1.26	0.19	0.21	0.07	0.61	Ns
	32008	4	2.17	0.77	0.54	-0.43	0.03	*
	C0444	2	1.09	0.04	0.08	0.48	0.00	***
Mono-Couffo (*T. candida*)	**C0241**	2	1.04	0.04	0.03	-0.02	0.89	Ns
	7_02K06	2	1.75	0.21	0.43	0.52	0.00	***
	01_B12	2	1.27	0.18	0.22	0.19	0.16	Ns
	32008	4	2.30	0.67	0.57	-0.18	0.00	***

APPENDIX B. (cont.)

	C0444	2	1.02	0.02	0.02	-0.01	0.95	Ns
Zou-Collines (*T. candida*)	**C0241**	4	1.19	0.10	0.16	0.41	0.00	***
	7_02K06	2	1.19	0.18	0.16	-0.10	0.49	Ns
	01_B12	2	1.49	0.29	0.33	0.12	0.42	Ns
	32008	5	2.84	0.79	0.65	-0.22	0.00	**
	C0444	2	1.14	0.13	0.13	-0.07	0.60	Ns
Oueme-Plateau (*T. candida*)	**C0241**	2	1.04	0.04	0.04	-0.02	0.89	Ns
	7_02K06	2	1.10	0.02	0.09	0.79	0.00	***
	01_B12	2	1.60	0.29	0.38	0.22	0.12	Ns
	32008	4	2.73	0.82	0.63	-0.29	0.02	*
	C0444	2	1.02	0.02	0.02	-0.01	0.94	Ns

APPENDIX C

Information on the insect mitogenomes included in the phylogenetic analysis

Species	Family	Length (bp)	Accession Number	Reference
Antheraea pernyi	Saturniidae	15566	AY242996	Liu et al. 2008
Antheraea yamamai	Saturniidae	15338	EU726630	Kim et al. 2009
Eriogyna pyretorum	Saturniidae	15327	FJ685653	Jiang et al. 2009
Saturnia boisduvalii	Saturniidae	15360	EF622227	Hong et al. 2008
Bombyx mori strain *Xiafang*	Bombycidae	15664	AY048187	Lu et al. 2002
Bombyx mandarina	Bombycidae	15928	AB070263	Yukuhiro et al. 2002
Manduca sexta	Sphingidae	15516	EU286785	Cameron and Whiting 2008
Phthonandria atrilineata	Geometridae	15499	EU569764	Yang et al. 2009
Helicoverpa armigera	Noctuidae	15347	GU188273	Yin et al. 2010
Sesamia inferens	Noctuidae	15413	JN039362	Unpublished
Lymantria dispar	Erebidae	15569	FJ617240	Unpublished
Hyphantria cunea	Arctiidae	15481	GU592049	Liao et al. 2010
Acraea issoria	Nymphalidae	15245	GQ376195	Hu et al. 2010
Apatura metis	Nymphalidae	15236	JF801742	Zhang et al. 2012
Calinaga davidis	Nymphalidae	15267	HQ658143	Unpublished
Hipparchia autonoe	Nymphalidae	15489	GQ868707	Kim et al. 2006
Sasakia charonda	Nymphalidae	15244	AP011824	Unpublished
Sasakia charonda kuriyamaensis	Nymphalidae	15222	AP011825	Unpublished
Coreana raphaelis	Lycaenidae	15314	DQ102703	Kim et al. 2006
Papilio maraho	Papilionidae	16094	FJ810212	Unpublished
Teinopalpus aureus	Papilionidae	15242	HM563681	Unpublished

APPENDIX C. (cont.)

Troides aeacus	Papilionidae	15263	EU625344	Unpublished
Parnassius bremeri	Papilionidae	15389	FJ871125	Kim et al. 2009
Adoxophyes honmai	Tortricidae	15680	DQ073916	Lee et al. 2006
Grapholita molesta	Tortricidae	15717	HQ392511	Buckley et al. 2000
Spilonota lechriaspis	Tortricidae	15368	HM204705	Zhao et al. 2011
Chilo suppressalis	Crambidae	15395	JF339041	Chat et al. 2012
Cnaphalocrocis medinalis	Crambidae	15388	JN246082	Chai et al. 2012
Diatraea saccharalis	Crambidae	15490	FJ240227	Li et al. 2011
Maruca vitrata (Old World)	Crambidae	14054	HM751150	Margam et al. 2011
Ostrinia furnacalis	Crambidae	14536	AF467260	Coates et al. 2005
Ostrinia nubilalis	Crambidae	14535	AF442957	Coates et al. 2005
Maruca vitrata (New World)	Crambidae	15385	KJ466365	This study
Drosophila melanogaster	Drosophilidae	19517	DMU37541	Lewis et al. 1995

APPENDIX D
Tandem repeats predicted in the mitogenome of *M. vitrata* (New World)

Position	Length	Repeats	Sequence
22	2	9	TTTTTTTTTTTTTTTTT
40	4	2	ATATATAT
62	3	2	ATTATT
81	4	2	TTTCTTTC
92	3	2	TTCTTC
141	3	2	TAATAA
152	4	2	TAAATAAA
161	4	2	AATTAATT
170	2	3	ATATAT
181	4	2	AATTAATT
219	4	4	AATTAATTAATTAATT
254	3	2	TAATAA
272	2	9	ATATATATATATATATAT
311	2	3	TTTTTT
319	4	2	ATAAATAA
328	2	3	AAAAAA

APPENDIX E

Multiple sequence alignment of mitogenome sequences from *M. vitrata* (Old World) (MvOW), and *M. vitrata* (New World) (MvNW)

```
MvNW               1 TTAAAAATAAGCTAAATTAAGCTTTTGGGTTCATACCTCAAATATAAAGG     50
MvOW               1 ---------------------------------------------------      0

MvNW              51 AATAACCTTTTTTTTAAAAATAAAGTGCCTGATTAAAGGATTATTCTGAT    100
MvOW               1 ---------------------------------------------------      0

MvNW             101 AGGATAAATTAAGTAGTTTTTCTACCTTTATTATATTTTATAGAATTAAA    150
MvOW               1 ---------------------------------------------------      0

MvNW             151 CTATATCTAATAGTATCAAAAACTATTGTGCATCTTACACTAAAATATAA    200
MvOW               1 ---------------------------------------------------      0

MvNW             201 TTATAAATTTTTATTTATAAAAAGAATTTCTTTTATTTTAAATTTTTTTC    250
MvOW               1 ---------------------------------------------------      0

MvNW             251 AATTTTAATTCTAATAAAATATTTTTCTTATTTATTATTTTTTTCAGAAC    300
                        ||.||||||||||||||||||||||||||||||||||||||||||||
MvOW               1 ---TTAAATTCTAATAAAATATTTTTCTTATTTATTATTTTTTTCAGAAC     47

MvNW             301 ATTAATCTCTATTTCTTCTAATTCTTGATTTGGTTGCTGAATTGGATTAG    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW              48 ATTAATCTCTATTTCTTCTAATTCTTGATTTGGTTGCTGAATTGGATTAG     97

MvNW             351 AAATTAATTTATTAAGTTTTATCCCCCTAATTAATAATTCTAATAATATT    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW              98 AAATTAATTTATTAAGTTTTATCCCCCTAATTAATAATTCTAATAATATT    147
```

```
MvNW            401 TTATCTACAGAAGCCTCATTAAAATATTTTCTAGTACAATCAATTGCTTC    450
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            148 TTATCTACAGAAGCCTCATTAAAATATTTTCTAGTACAATCAATTGCTTC    197

MvNW            451 TATTAATCTATTATTTTGTATTATTTTTAAAATAATCTTATTAAAAAATT    500
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            198 TATTAATCTATTATTTTGTATTATTTTTAAAATAATCTTATTAAAAAATT    247

MvNW            501 TTGAAATAAATAATATTTTATCAATCTTAATTAATTCATCACTATTAATA    550
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            248 TTGAAATAAATAATATTTTATCAATCTTAATTAATTCATCACTATTAATA    297

MvNW            551 AAAATGGGATCAACCCCTTTTCACTTTTGATTCCCTAATATTGTAGAAGG    600
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            298 AAAATGGGATCAACCCCTTTTCACTTTTGATTCCCTAATATTGTAGAAGG    347

MvNW            601 ATTATCCTGATTTAATAATTTTATTTTAATAACTTGACAAAAAATTACCC    650
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            348 ATTATCCTGATTTAATAATTTTATTTTAATAACTTGACAAAAAATTACCC    397

MvNW            651 CCATAATTTTATTATCATATTATTTTAATAAAAATTTTTTAATTATTATT    700
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            398 CCATAATTTTATTATCATATTATTTTAATAAAAATTTTTTAATTATTATT    447

MvNW            701 ATTATTATAAATTCTATTATTGGTGCTATTGGAGGATTAAATCAAACTTC    750
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            448 ATTATTATAAATTCTATTATTGGTGCTATTGGAGGATTAAATCAAACTTC    497

MvNW            751 TCTACGAAAATTAATGGCTTTTTCATCAATTAATAATTTAAGATGAATAA    800
                    |||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            498 TCTACGAAAATTAATGGCTTTTTCATCAATTAATAATTTAAGATGAATAA    547

MvNW            801 TTTCTTCTTTAATAATCAGAGAAAATTTATGAATAATATATTTTTTTTTT    850
                    |||||||||||||||||||||||||||||||||||||||    ||||||
MvOW            548 TTTCTTCTTTAATAATCAGAGAAAATTTATGAATAATATA----TTTTTT    593
```

```
MvNW            851 TATA-GTTTTTTAATTAGAATTATATGTTTATTATTTTACTTGACTAATA    899
                    ||||  ||||||||||||||||||||||||||||||||||||||||||||
MvOW            594 TATATGTTTTTTAATTAGAATTATATGTTTATTATTTTACTTGACTAATA    643

MvNW            900 TATATTTTATTAATCAATTATTTTTTTTTAATATAAATTACATAATTAAA    949
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            644 TATATTTTATTAATCAATTATTTTTTTTTAATATAAATTACATAATTAAA    693

MvNW            950 TTGTCTTTATTAATTAATTTTTTATCTTTAGGGGGTTTACCTCCATTTAT    999
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            694 TTGTCTTTATTAATTAATTTTTTATCTTTAGGGGGTTTACCTCCATTTAT    743

MvNW           1000 TGGATTCTTTCCTAAATGAATCATTATTAATTTCTTACTAAAAAATAATT   1049
                    ||||||.||||||||.||||.||||||||||..||.||||||||||||
MvOW            744 TGGATTTTTTCCTAAGTGAATTATTATTAATTTTCTATTAAAAAATAATT    793

MvNW           1050 ATTTTTTTATAACTTTTATCTTAATTATAATAAGATTAATTTTATTATTT   1099
                    .|||||||.||||||||||.||||||||||||||||.||||||||||||
MvOW            794 TTTTTTTTTTAACTTTTATTTTAATTATAATAAGATTAGTTTTATTATTT    843

MvNW           1100 TTTTATATTCGAATTTTATATTCATCATTCATATTTAATTACTTAAAATT   1149
                    ||||||||||||||||||||||||||||||.||||||||||.||||||||
MvOW            844 TTTTATATTCGAATTTTATATTCATCATTTATATTTAATTATTTAAAATT    893

MvNW           1150 AAAATGAATAAAAATTTTTATTAAAAATAAAATAATATATTTTATTAATT   1199
                    ||||||||||||||||||.||.||||||||||||||||||||||||||||
MvOW            894 AAAATGAATAAAAATTTTCATCAAAAATAAAATAATATATTTTATTAATT    943

MvNW           1200 TACTTTCACTTATTTCTTCTATAGGCTTAATTTTAAGTAATTTTTTTTAT   1249
                    |.||||||||||||||||||||||.||||||||||||||||||||||||||
MvOW            944 TTCTTTCACTTATTTCTTCTATAGGTTTAATTTTAAGTAATTTTTTTTAT    993

MvNW           1250 TTA------TAAGAAGGTTTTAAGTTAATTAAAACTAATAATCTTCAAAA   1293
                    |||      ||||||||||||||||||||.|||||||||||||||.|||
MvOW            994 TTATAATTTTAAGAAGGTTTTAAGTTAATTTAAACTAATAATCTTCCAAA   1043
```

```
MvNW          1294 TTATTTATAAAGAAACATTCTTTAAGCCTTAATAATTTTTTATACCTTAA   1343
                   |||..||.||.||||.||||||||||||||||||||||||||||||||
MvOW          1044 TTACGTACAATGAAATATTCTTTAAGCCTTAATAATTTTTTATACCTTAA   1093

MvNW          1344 AATTTGCAATTTTATATCATTAATTT-GAATATAAGACCTATAATAAAAA   1392
                   ||||||||||||||||||.|||||||  ||||||||||.||||||.|||||
MvOW          1094 AATTTGCAATTTTATATCTTTAATTTCGAATATAAGACGTATAACAAAAA   1143

MvNW          1393 AGAATTTTTCTTGTTAATAAATTTACAATTTATCGCTTATAACCTCAGCC   1442
                   |||||||||...||.||||||||||||||||||||||||||.|||||||
MvOW          1144 AGAATTTTTGGCGTCAATAAATTTACAATTTATCGCTTATAACGTCAGCC   1193

MvNW          1443 ATTTTATTATTATAGCGAAAATGAATTTACTCAACAAATCATAAAGATAT   1492
                   |||||||||  ||||||||||||||||||||||||||||||||||||||||
MvOW          1194 ATTTTATTA--ATAGCGAAAATGAATTTACTCAACAAATCATAAAGATAT   1241

MvNW          1493 TGGAACATTATATTTTATTTTTGGAATTTGAGCAGGAATAGTAGGAACAT   1542
                   |||.||||||||||||||||||||||||||||||||||||||||||||||
MvOW          1242 TGGTACATTATATTTTATTTTTGGAATTTGAGCAGGAATAGTAGGAACAT   1291

MvNW          1543 CTTTAAGTTTATTAATTCGAGCAGAATTAGGTAATCCAGGATCTTTAATT   1592
                   |||||||||||||||||||||||||||||||||||.||||||||||||
MvOW          1292 CTTTAAGTTTATTAATTCGAGCAGAATTAGGTAATCCTGGATCTTTAATT   1341

MvNW          1593 GGAGATGATCAAATTTATAATACTATTGTAACAGCTCACGCATTTATTAT   1642
                   |||||||||||||||||||||||||||||||||||.||||||||||||
MvOW          1342 GGAGATGATCAAATTTATAATACTATTGTAACAGCTCATGCATTTATTAT   1391

MvNW          1643 AATTTTTTTTATGGTTATACCTATTATAATTGGAGGATTTGGAAATTGAT   1692
                   ||||||||||||.||.||||||||||||||||||||||||||.||||
MvOW          1392 AATTTTTTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAACTGAT   1441

MvNW          1693 TAGTTCCTTTAATATTAGGAGCCCCAGATATAGCTTTTCCACGAATAAAT   1742
                   ||||||||||||||||||||||.|||||||||||||.||||||||||||
MvOW          1442 TAGTTCCTTTAATATTAGGAGCTCCAGATATAGCTTTCCCACGAATAAAT   1491
```

```
MvNW          1743 AATATAAGATTCTGAATATTACCCCCATCATTAACTTTATTAATTTCGAG   1792
                   ||||||||||.||||||||||||||||||||||||||||||||||||
MvOW          1492 AATATAAGATTTTGAATATTACCCCCATCATTAACTTTATTAATTTCGAG   1541

MvNW          1793 AAGAATTGTAGAAAATGGAGCAGGTACTGGATGAACAGTATACCCCCCTC   1842
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          1542 AAGAATTGTAGAAAATGGAGCAGGTACTGGATGAACAGTATACCCCCCTC   1591

MvNW          1843 TTTCATCTAATATTGCTCATGGAGGAAGATCAGTTGATTTAGCTATTTTT   1892
                   |.||||||.||||||||.||.||||||.|||||||||||||||||||||||
MvOW          1592 TCTCATCAAATATTGCCCACGGAGGTAGATCAGTTGATTTAGCTATTTTT   1641

MvNW          1893 TCTTTACATTTAGCTGGAATTTCATCAATTTTAGGGGCAATCAATTTTAT   1942
                   ||||||||||||||||.||||||||||||||||.||||||.||||||||
MvOW          1642 TCTTTACATTTAGCTGGTATTTCATCAATTTTAGGAGCAATTAATTTTAT   1691

MvNW          1943 TACTACGATTATTAATATACGAGTAAATGGATTAACTTTTGATCAAATAC   1992
                   |||.||.|||||||||||||||||||||||.||.|.||||||||||||||
MvOW          1692 TACCACAATTATTAATATACGAGTAAATGGACTATCCTTTGATCAAATAC   1741

MvNW          1993 CTTTATTTGTTTGATCTGTTGGAATTACAGCTCTTTTATTATTACTTTCT   2042
                   ||.||||||||||.|||||||||||||||||||||||.|.|||||||||
MvOW          1742 CTCTATTTGTTTGGTCTGTTGGAATTACAGCTCTTTTACTTTTACTTTCT   1791

MvNW          2043 CTACCAGTTTTAGCAGGTGCTATTACTATACTTTTAACAGACCGAAATCT   2092
                   .|||||||||||||||||||||||||||||||||||||||.|||||||.|
MvOW          1792 TTACCAGTTTTAGCAGGTGCTATTACTATACTTTTAACAGATCGAAATTT   1841

MvNW          2093 TAATACTTCTTTTTTTGATCCAGCTGGAGGAGGAGATCCAATTTTATATC   2142
                   .||||||||||||||||.||||||||||||||||||||||||||||||||
MvOW          1842 AAATACTTCTTTTTTTGACCCAGCTGGAGGAGGAGATCCAATTTTATATC   1891

MvNW          2143 AACATTTATTTTGATTTTTTGGTCACCCTGAAGTTTATATTTTAATTCTT   2192
                   ||||||||||||||||||||||.||.||.||||||||.|||||||||.|.
MvOW          1892 AACATTTATTTTGATTTTTTGGACATCCAGAAGTTTACATTTTAATTTTA   1941
```

```
MvNW          2193 CCTGGATTTGGAATAATTTCCCATATTATTTCACAAGAAAGTGGAAAAAA   2242
                   ||||||||||.||||||||||||||||||||.||||||||.||.||||||
MvOW          1942 CCTGGATTTGGTATAATTTCCCATATTATTTCCCAAGAAAGAGGTAAAAA   1991

MvNW          2243 AGAAACATTTGGATCTTTAGGAATAATTTATGCTATAATAGCAATTGGAT   2292
                   .||||||||||||||||||||.||||||||||||||||||||||||||||
MvOW          1992 GGAAACATTTGGATCTTTAGGGATAATTTATGCTATAATAGCAATTGGAT   2041

MvNW          2293 TATTAGGATTTGTAGTATGAGCTCATCATATATTTACAGTAGGTATAGAT   2342
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          2042 TATTAGGATTTGTAGTATGAGCTCATCATATATTTACAGTAGGTATAGAT   2091

MvNW          2343 ATTGATACACGAGCTTATTTTACTTCCGCAACTATAATTATTGCTGTACC   2392
                   ||||||||||||||||.||||||||||||||||||||||.||||||
MvOW          2092 ATTGATACACGAGCTTACTTTACTTCCGCAACTATAATTATTGCAGTACC   2141

MvNW          2393 AACAGGAATTAAAATTTTTAGATGATTGGCTACTTTTCACGGAACTCAAA   2442
                   ||||||||||||||||||||||||||.||.||||||.||.||||||.||||
MvOW          2142 AACAGGAATTAAAATTTTTAGATGATTAGCAACTTTCCATGGAACACAAA   2191

MvNW          2443 TTAATTATAGTCCTTCAATTTTATGAAGATTAGGATTTGTTTTTTTATTC   2492
                   ||||||||||||||.||||||.||.||||||||||||||||||||||||||.
MvOW          2192 TTAATTATAGTCCCTCAATCTTTTGAAGATTAGGATTTGTTTTTTTATTT   2241

MvNW          2493 ACTGTTGGAGGATTAACAGGAGTAATTTTAGCTAATTCCTCAATTGATGT   2542
                   ||||||||||||||||||||||||||||||||||.||||||.||.||||||||
MvOW          2242 ACTGTTGGAGGATTAACAGGAGTAATTTTAGCCAATTCTTCTATTGATGT   2291

MvNW          2543 AGCTCTACATGATACTTATTATGTAGTAGCCCATTTTCACTATGTTTTAT   2592
                   ||||||.||||||.||||||||||||||.||||||||||.||||||||||
MvOW          2292 AGCTCTTCATGACACTTATTATGTAGTGGCCCATTTTCATTATGTTTTAT   2341

MvNW          2593 CTATAGGAGCAGTTTTTGCAATTATAGCAGGATTTATTCACTGATATCCT   2642
                   ||||||||||||||.||.||||||||||||||||||||||.|||||||||
MvOW          2342 CTATAGGAGCAGTATTCGCAATTATAGCAGGATTTATTCATTGATATCCT   2391
```

```
MvNW        2643 TTATTTACTGGTCTTACTCTTAACCCTTTTATATTAAAGATTCAATTTTT   2692
                 ||||||||.||.||||||||||.||||.|||.||||.||||||||.||
MvOW        2392 TTATTTACAGGACTTACTCTTAATCCTTTCATACTAAAAATTCAATTCTT   2441

MvNW        2693 TACTATATTTATTGGAGTAAATTTAACATTTTTCCCACAACATTTCTTAG   2742
                 ||||||||||||||||||||||||||||||.||.||.||||||.|||||||
MvOW        2442 TACTATATTTATTGGAGTAAATTTAACATTCTTTCCTCAACACTTCTTAG   2491

MvNW        2743 GATTAGCAGGAATACCCCGACGATATTCTGATTATCCTGATGTTTATACT   2792
                 |||||||.||||||||.||||||||||.||||||||||||||.||||||
MvOW        2492 GATTAGCTGGAATACCTCGACGATATTCAGATTATCCTGATGTCTATACT   2541

MvNW        2793 TCATGAAATATTGCTTCTTCTTTAGGATCTTATATCTCTCTATTAGCAGT   2842
                 ||||||||||||||||||||.||||||||||||||.|||.|||||||.||
MvOW        2542 TCATGAAATATTGCTTCTTCATTAGGATCTTATATTTCTTTATTAGCTGT   2591

MvNW        2843 ATTATTATTTTTAATTATTATTTGAGAATCTATAATTAGTCAACGAATAA   2892
                 ..||||||||.||||||||||||||||||.||||||||||||||||||||
MvOW        2592 GATATTATTTCTAATTATTATTTGAGAATCAATAATTAGTCAACGAATAA   2641

MvNW        2893 TTTTATTTTCATTAAATTTATCATCTTCAATTGAATGATATCAAAATTTA   2942
                 ||||||||||||||||||||||||||||.||||||||||||||||||||||
MvOW        2642 TTTTATTTTCATTAAATTTATCATCTTCTATTGAATGATATCAAAATTTA   2691

MvNW        2943 CCACCTGCAGAACATTCATATAATGAACTTCCAATTTTAAGAAATTTCTA   2992
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW        2692 CCACCTGCAGAACATTCATATAATGAACTTCCAATTTTAAGAAATTTCTA   2741

MvNW        2993 ATATGGCAGATTATATGTAATGGATTTAAACCCCATTTATAAAGGATTAT   3042
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW        2742 ATATGGCAGATTATATGTAATGGATTTAAACCCCATTTATAAAGGATTAT   2791

MvNW        3043 CCTTTTTTTAGAAATGGCAACATGATCTAATTTTAACTTACAAAACGGAG   3092
                 ||||||||||||||||||||||||||||||||||||||||||||||.||||
MvOW        2792 CCTTTTTTTAGAAATGGCAACATGATCTAATTTTAACTTACAAAATGGAG   2841
```

```
MvNW            3093 CATCTCCACTTATAGAACAAATTATTTTTTTCCATGATCATACTTTAATT   3142
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            2842 CATCTCCACTTATAGAACAAATTATTTTTTTCCATGATCATACTTTAATT   2891

MvNW            3143 ATTTTAATTATAACTACTATTTTAGTAGGATATTTAATATTAAGTTTATT   3192
                     |||||||||||||.||||.|.|||||||||||||||||||||.|||||
MvOW            2892 ATTTTAATTATAATTACTGTATTAGTAGGATATTTAATATTAAGATTATT   2941

MvNW            3193 TTTTAATAAATATATTAATCGATTTTTATTAGAAGGTCAAATAATTGAGT   3242
                     ||||||||||||||||||||||||||||||||||||||||||||||||.|
MvOW            2942 TTTTAATAAATATATTAATCGATTTTTATTAGAAGGTCAAATAATTGAAT   2991

MvNW            3243 TAATTTGAACAATTTTACCAGCTATTACTTTAATTTTTATTGCTTTACCT   3292
                     ||||||||||.|||||||||||||||||||||||||||||||.||||.
MvOW            2992 TAATTTGAACTATTTTACCAGCTATTACTTTAATTTTTATTGCTCTACCC   3041

MvNW            3293 TCATTACGTTTACTTTATTTATTAGATGAACTTAATAATCCATTAATTAC   3342
                     ||.|||||||||||||||||||||||||||||||.||||||||||||
MvOW            3042 TCTTTACGTTTACTTTATTTATTAGATGAACTTAATAACCCATTAATTAC   3091

MvNW            3343 TTTAAAATCTATTGGCCATCAATGATATTGAAGATACGAATATTCAGATT   3392
                     .||||.|||||||||.|||||||||||||||||.||.||||||||||
MvOW            3092 CTTAAGATCTATTGGACATCAATGATATTGAAGATATGAGTATTCAGATT   3141

MvNW            3393 TTAATAATATTGAATTTGATTCATATATAACCCCTATGAATGAAATAAAT   3442
                     ||||.|||||||||||||||||||||||||||.|.|||||||.|||
MvOW            3142 TTAACAATATTGAATTTGATTCATATATAACCCCTGTAAATGAAATGAAT   3191

MvNW            3443 AATAATAATTTTCGATTATTAGATGTTGATAATCGAATTGTTTTACCGAT   3492
                     ||.||||.|||.||||||||||||||||||||||||||||||||.||
MvOW            3192 AACAATAGTTTCCGATTATTAGATGTTGATAATCGAATTGTTTTACCAAT   3241

MvNW            3493 GGGGAATCAAATTCGAATTATAGTAACTGCTACAGATGTTATTCACTCAT   3542
                     |||.|||||||||||||||||||||||||||||||||||.||.||||
MvOW            3242 GGGTAATCAAATTCGAATTATAGTAACTGCTACAGATGTTATCCATTCAT   3291
```

```
MvNW            3543 GAACTATCCCATCTTTAGGTGTAAAAGTAGATGCTAATCCAGGACGATTA   3592
                     ||||.||||||||.||||||||||||||||||||||||||||||||||
MvOW            3292 GAACCATCCCATCCTTAGGAGTAAAAGTAGATGCTAATCCAGGACGATTA   3341

MvNW            3593 AATCAAACAAATTTCTTTATTAATCGTCCTGGAATTTTTTATGGACAATG   3642
                     ||||||||||||||||||||||||||.||||||||||||||||||||||||
MvOW            3342 AATCAAACAAATTTCTTTATTAATCGACCTGGAATTTTTTATGGACAATG   3391

MvNW            3643 TTCTGAAATTTGTGGAGCAAACCACAGTTTTATACCTATTGTTATTGAAA   3692
                     ||||||||||||||||.||||||.||.||||||||||||||||||||||||
MvOW            3392 TTCTGAAATTTGTGGTGCAAATCATAGTTTTATACCTATTGTTATTGAAA   3441

MvNW            3693 GAATTTCAATTAAAAATTTTATTAATTGAATTAATAATTATTCATCATTA   3742
                     ||||.||||||||||||||||||||||||||||||||||||||||||||||
MvOW            3442 GAATCTCAATTAAAAATTTTATTAATTGAATTAATAATTATTCATCATTA   3491

MvNW            3743 GATGACTGAAAGCAAGTACTGGTCTCTTAAACCATTTTATAGTAAATTAG   3792
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            3492 GATGACTGAAAGCAAGTACTGGTCTCTTAAACCATTTTATAGTAAATTAG   3541

MvNW            3793 CATTTACTTCTAATGATTAAAGAATTAGTTAAACCTATAACATAAATATG   3842
                     ||||||||||||||||.||||||||||||||..||||||||||.||||
MvOW            3542 CATTTACTTCTAATGAATAAAGAATTAGTTAAATTTATAACATAAGTATG   3591

MvNW            3843 TCAAATTTAAATTATTATTTCAT-ATAATATTCTTTTATCCCTCAAATAA   3891
                     |||||.||||||||||||  |.|| ||||||||||||||.||||||||||
MvOW            3592 TCAAACTTAAATTATTA--TAATAATAATATTCTTTTATTCCTCAAATAA   3639

MvNW            3892 TACCAATTAATTGAATTTTTTCTTTCTTTTTTTTTTGTTATTATTTTTATT   3941
                     ||||||||||||||||||||||||||||||||||||||||.||||||||
MvOW            3640 TACCAATTAATTGAATTTTTTCTTTCTTTTTTTTTTGTTATTGTTTTTATT   3689

MvNW            3942 ATTTTTAATATCATAAATTATTTTATTTTTATTAATAAAAATAATAGAAA   3991
                     ||||||||||||.||||||||||.||||||||||||||||||||||.|||
MvOW            3690 ATTTTTAATATTATAAATTATTATATTTTTATTAATAAAAATAATAAAAA   3739
```

```
MvNW        3992 TAATATTTTTTTTCAAAAAAAAAATAAAAACCTATTTTGAAAATGATAAC   4041
                 ||||||.|||||||..||||||||||.|||||.||||||||||||||||||
MvOW        3740 TAATATCTTTTTTTTTAAAAAAAAATCAAAACTTATTTTGAAAATGATAAC   3789

MvNW        4042 TAATCTTTTTTCAATTTTTGACCCATCTACTAATTTATTATATTTACCTT   4091
                 |||||||||||||||||||||||||||||||||||||||||||||||||.
MvOW        3790 TAATCTTTTTTCAATTTTTGACCCATCTACTAATTTATTATACTTACCTC   3839

MvNW        4092 TAAATTGAATTAGAACTTTATTAGGAATTATATTTATTCCTTATTCATTT   4141
                 |||||||||||||||||||||||||.||.||||||||.||||||||||||
MvOW        3840 TAAATTGAATTAGAACTTTATTAGGGATCATATTTATCCCTTATTCATTT   3889

MvNW        4142 TGATTAATCCCTAATCGATACTACTTATTTTGAAATTTTATTTTAAATAA   4191
                 ||||||||.||||||||||.||.|||||||||||||||.||||||||||||
MvOW        3890 TGATTAATTCCTAATCGATATTATTTATTTTGAAATTTCATTTTAAATAA   3939

MvNW        4192 ACTCCATAAAGAATTTAAAACTTTATTAGGAAATAATTCAAATGGATCGA   4241
                 |||.|||||||||||||||||||||||||||||||||||||||||||||.|
MvOW        3940 ACTTCATAAAGAATTTAAAACTTTATTAGGAAATAATTCAAATGGATCAA   3989

MvNW        4242 CTTTTATTTTTATTTCGATATTTACTTTTGTTCTATTTAATAATTTTTTA   4291
                 ||||||||||||||||.|||||||||||||..||||||||||||||||||
MvOW        3990 CTTTTATTTTTATTTCAATATTTACTTTTGTATTATTTAATAATTTTTTA   4039

MvNW        4292 GGATTATTCCCATATATTTTTACTAGAACAAGCCACTTAACCTTATCACT   4341
                 ||||||||.||.||||||||||.||||||||.||.|||||.||||||.|
MvOW        4040 GGATTATTTCCTTATATTTTTACAAGAACAAGTCATTTAACTTTATCATT   4089

MvNW        4342 ATCAATTTCACTACCATTGTGATTAAGATTTATATTTTATGGATGAATTA   4391
                 ||||||||||.|||||||.||||||||||||||||||||||||||||||||
MvOW        4090 ATCAATTTCATTACCATTATGATTAAGATTTATATTTTATGGATGAATTA   4139

MvNW        4392 ATAATACTCAACATATATTTATTCACATAATTCCACAAGGTACTCCTGGT   4441
                 |||||||||||||||||||||||||.||||||||.||||||.||||||.|||
MvOW        4140 ATAATACTCAACATATATTTATTCATATAATTCCTCAAGGAACTCCAGGT   4189
```

```
MvNW          4442 ATTTTAATACCTTTTATAGTTTTAATTGAAACAATCAGAAATATTATTCG   4491
                   |||||||||||||||||||||||||||||||||||.|||||||||||||||
MvOW          4190 ATTTTAATACCTTTTATAGTTTTAATTGAAACAATTAGAAATATTATTCG   4239

MvNW          4492 ACCAGGAACATTAGCTGTTCGACTTACAGCCAATATAATTGCAGGACATT   4541
                   ||||||.|||.||||||||||||||||||.||||||||||||||||||||
MvOW          4240 ACCAGGGACACTAGCTGTTCGACTTACAGCTAATATAATTGCAGGACATT   4289

MvNW          4542 TATTAATAACTTTATTAAGAGGGACAGGACCTAACTTACCTATTTATTTT   4591
                   |||||||||||||||.|||||||.||||||||||.|||||||.|.|||||||
MvOW          4290 TATTAATAACTTTACTAAGAGGAACAGGACCTAATTTACCTGTATATTTT   4339

MvNW          4592 ATTATTGTATTAATTATTATTCAAATTTTATTATTAGTTTTAGAATCAGC   4641
                   |||.|.|||||||.||||||||||||||||.||.||||||||||||||||
MvOW          4340 ATTGTAGTATTAGTTATTATTCAAATTTTACTACTAGTTTTAGAATCAGC   4389

MvNW          4642 AGTTGCGGTTATTCAATCTTATGTTATTGCTATTTTAAGAACATTATATT   4691
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          4390 AGTTGCGGTTATTCAATCTTATGTTATTGCTATTTTAAGAACATTATATT   4439

MvNW          4692 CTAGAGAAGTAAATTAATAT-CTAATGAAAAATAATTTTAATCACCCCTA   4740
                   ||||||||||||||||||.| ||||||||||||||||||||||||||||||
MvOW          4440 CTAGAGAAGTAAATTAATCTACTAATGAAAAATAATTTTAATCACCCCTA   4489

MvNW          4741 TCATTTAGTTGATTATAGTCCATGACCCTTAACTGGTGCTGTTGGAGTTT   4790
                   |||.|||||||||||||||||||||.||||||.||.|||.|||||||||||
MvOW          4490 TCACTTAGTTGATTATAGTCCATGACCTTTAACAGGAGCTATTGGAGTTT   4539

MvNW          4791 TAACCTTAGTAACTGGAATAGTAAAATGATTCCATAATTTTAATATAAAT   4840
                   |||||.||||||||||||||||||||||||||||||||||||||||||||||
MvOW          4540 TAACTTTAGTAACTGGAATAGTAAAATGATTCCATAATTTTAATATAAAT   4589

MvNW          4841 TTATTAATTCTAGGATATTTTATTGTTATTTTAACTATATATCAATGATG   4890
                   ||||||||||||||||||||||||||.||.||||||||||||||||||||
MvOW          4590 TTATTAATTCTAGGATATTTTATTGTCATCTTAACTATATATCAATGATG   4639
```

```
MvNW        4891 ACGAGATATTTGCCGAGAAGGAACTCTTCAAGGAAAACATACAATTTTAG   4940
                 ||||||||||||.||||||||||||||||||||.||||||||.||||||||
MvOW        4640 ACGAGATATTTGTCGAGAAGGAACTCTTCAAGGTAAACATACGATTTTAG   4689

MvNW        4941 TAACTAAAGGATTACGATGAGGAATAATTTTATTTATTATTTCAGAAATT   4990
                 ||||||||||||||||||||||||||||||||||||||||||||||||.||
MvOW        4690 TAACTAAAGGATTACGATGAGGAATAATTTTATTTATTATTTCAGAAGTT   4739

MvNW        4991 TTCTTTTTTGTATCTTTTTTTTGAGCTTTTTTTCATAGAAGTTTATCACC   5040
                 ||.||||||.||||||||||||||||||||||||||||||||||||||||||
MvOW        4740 TTTTTTTTCGTATCTTTTTTTTGAGCTTTTTTTCATAGAAGTTTATCACC   4789

MvNW        5041 AAATATTGAAATTGGTGCTTTATGACCCCCTATAAGTATCACCCCATTTA   5090
                 ||||||||||||||||||||||||||.||.||||||||.||.||||||||
MvOW        4790 AAATATTGAAATTGGTGCTTTATGACCTCCAATAAGTATTACTCCATTTA   4839

MvNW        5091 ATCCTTTCCAAATTCCTCTCTTAAATACTATTATTTTAATTACTTCAGGA   5140
                 |||||||.||||||||||.||||||||||||||||||||||.||.|||
MvOW        4840 ATCCTTTTCAAATTCCTCTTTTAAATACTATTATTTTAATTACATCCGGA   4889

MvNW        5141 ATTACAGTTACATGAGCTCATCATGCAATTATAGAAAATAATCATTCACA   5190
                 ||||||||||||||||.||||||||.|||||||||||||||.||||||||
MvOW        4890 ATTACAGTTACATGAGCACATCATGCCATTATAGAAAATAACCATTCACA   4939

MvNW        5191 AATAACCCAAGGACTTTTTTTTACTATTATTTTAGGAATTTATTTTACAA   5240
                 ||||||.||||||||.||||||||||||.||||||||||||||||||.|
MvOW        4940 AATAACTCAAGGACTCTTTTTTACTATTGTTTTAGGAATTTATTTTACTA   4989

MvNW        5241 TTTTACAAGCATATGAATATATTGAAGCACCTTTTTCTATTGCTGATAGT   5290
                 ||||||||||.||||||||||||||||||||||.|||||||.||||||.
MvOW        4990 TTTTACAAGCTTATGAATATATTGAAGCACCTTTTACTATTGCAGATAGA   5039

MvNW        5291 ATTTACGGATCAACTTTTTTTATAGCTACTGGATTTCATGGATTACATGT   5340
                 |||||.||||||||||||||||||||||||||.|||||||||||||||
MvOW        5040 ATTTATGGATCAACTTTTTTTATAGCTACTGGATTCCATGGATTACATGT   5089
```

```
MvNW        5341 AATAATCGGAACATTATTTTTATTAATTTGTTTAATTCGGCATATTTATA   5390
                 ||||||.|||||.||||||||||||||||.|||||||.||||||||||
MvOW        5090 AATAATTGGAACTTTATTTTTATTAATTTGCTTAATTCGACATATTTATA   5139

MvNW        5391 ATCATTTTTCTAATAATCATCATTTTGGCTTTGAAGCTGCTGCTTGATAT   5440
                 ||||||||||||||||.||||||||||.|||||||||||||||.||||||
MvOW        5140 ATCATTTTTCTAATAACCATCATTTTGGATTTGAAGCTGCTGCCTGATAT   5189

MvNW        5441 TGACATTTCGTAGATGTAGTTTGATTATTCCTTTATATTTCTATTTATTG   5490
                 |||||||||||||||.||||||||||||||||||||||||||||||||||
MvOW        5190 TGACATTTCGTAGACGTAGTTTGATTATTCCTTTATATTTCTATTTATTG   5239

MvNW        5491 ATGAGGAAATTAATTATTTATATAATATATTTAGTATATTTGACTTCCAA   5540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW        5240 ATGAGGAAATTAATTATTTATATAATATATTTAGTATATTTGACTTCCAA   5289

MvNW        5541 TCAAAAAGTTTAATTTTTTTTTAATATAAATAATTTTATTAATTTTATAT   5590
                 |||||||||||||  ||.||||||||||||||||||||||||||.|||
MvOW        5290 TCAAAAAGTTTAA-TTATTTTTAATATAAATAATTTTATTAATTTTTTAT   5338

MvNW        5591 ATGACCTTAATTTTAATTTTAATTTCTAATATTATAATATTTTTATCAAT   5640
                 ||.||.||||||||||||||||||||||||||||||||||||||||||||
MvOW        5339 ATAACTTTAATTTTAATTTTAATTTCTAATATTATAATATTTTTATCAAT   5388

MvNW        5641 TTTACTATCAAAAAAATCTTTTTCTGATCGAGAAAAATGTTCACCTTTCG   5690
                 ||||.|||||||||||||||||||.||.||||||||||||||||||.|
MvOW        5389 TTTATTATCAAAAAAATCTTTTTCTGACCGTGAAAAATGTTCACCTTTTG   5438

MvNW        5691 AATGTGGATTTGACCCAAAATCTTCTGCTCGAATCCCCTTTTCAATACAT   5740
                 ||||||||||||||||.||||||||||||||||.||.||.||.||||||
MvOW        5439 AATGTGGATTTGACCCTAAATCTTCTGCTCGAATTCCTTTCTCTATACAT   5488

MvNW        5741 TTTTTTTTAATTACTGTAATTTTTTTGATTTTTGATGTTGAAATTGCATT   5790
                 ||||||||||||||.||||||||||.||||||||||.||||||||||||
MvOW        5489 TTTTTTTTAATTACCGTAATTTTTTTAATTTTTGATGTAGAAATTGCATT   5538
```

```
MvNW            5791 AATCTTCCCAATTATTAATTTATTTAAAATTACTAATTTTATTATTTGAT   5840
                     |||.||.|||||||||||||||||||||||||||||||||||||||||||
MvOW            5539 AATTTTTCCAATTATTAATTTATTTAAAATTACTAATTTTATTATTTGAT   5588

MvNW            5841 CTAAAATTAGTTTTTTTTTTATTATTATTTTACTTTTAGGTTTATTTCAT   5890
                     |||||||||||||||||||||||||||||||||||||||.||||||||.
MvOW            5589 CTAAAATTAGTTTTTTTTTTATTATTATTTTACTTTTAGGATTATTTCAC   5638

MvNW            5891 GAATGAAATCAAAATATACTTAATTGAATTAATT---AGGATTATAGTTT   5937
                     |||||||||||||||.||||||||||||||||||   |||||||||||||
MvOW            5639 GAATGAAATCAAAACATACTTAATTGAATTAATTAAAAGGATTATAGTTT   5688

MvNW            5938 AAAATAAAACATTTGATTTGCATTCAAAAAATATTGATTTATCAATTTAT   5987
                      |||||||||.||||||||||||||||||||||||||||||||||||||||
MvOW            5689 -AAATAAAACTTTTGATTTGCATTCAAAAAATATTGATTTATCAATTTAT   5737

MvNW            5988 CTTAAGTAAGAAGCAATTATGCATTTAATTTCGACTTAAAAGACAGAGTA   6037
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            5738 CTTAAGTAAGAAGCAATTATGCATTTAATTTCGACTTAAAAGACAGAGTA   5787

MvNW            6038 CAAGACTCCTTACTTATTAATTGAAACCAAAAAAGAGGTATATCACTGTTA   6087
                     ....||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            5788 TTTAACTCCTTACTTATTAATTGAAACCAAAAAAGAGGTATATCACTGTTA   5837

MvNW            6088 ATGATAACATTGAATTTAAAATTCCAATTAAATTAGAAATATAAAGTTTA   6137
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            5838 ATGATAACATTGAATTTAAAATTCCAATTAAATTAGAAATATAAAGTTTA   5887

MvNW            6138 AAATTAAGCTGCTAACTTAATTTTTAGTGGTTTAATTCCATTAATATTTC   6187
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            5888 AAATTAAGCTGCTAACTTAATTTTTAGTGGTTTAATTCCATTAATATTTC   5937

MvNW            6188 TTATTTATATAGTTTATTTAAAACATTACATTTTCATTGTAAAAATAAAA   6237
                     |||||||||||||||| ||||||||||||||||||||||||||||||||
MvOW            5938 TTATTTATATAGTTTA-TTAAAACATTACATTTTCATTGTAAAAATAAAA   5986
```

```
MvNW          6238 AAATT-TTTTTTATAAATATATTTAAAGATTAAAATAATCTCCCTAATAT   6286
                   ||||| ||||||||||||||||||||||||||||||||||||||||||||
MvOW          5987 AAATTATTTTTTATAAATATATTTAAAGATTAAAATAATCTCCCTAATAT   6036

MvNW          6287 CTTCAATATTATACTCTAAATTATAAGCTATTTAAATAAAATATAATTAT   6336
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          6037 CTTCAATATTATACTCTAAATTATAAGCTATTTAAATAAAATATAATTAT   6086

MvNW          6337 TATTATAAAATAAATTATTATTCATAAAATAAAACTAAATAAATAAATTT   6386
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          6087 TATTATAAAATAAATTATTATTCATAAAATAAAACTAAATAAATAAATTT   6136

MvNW          6387 TATAATTAGTTAATTGAAATAAATTATATAAAACTGAATACTTTTTTAAA   6436
                   ||||||||||.||||||||||||||||||||||||||||||||||||||||
MvOW          6137 TATAATTAGTCAATTGAAATAAATTATATAAAACTGAATACTTTTTTAAA   6186

MvNW          6437 ATTATATAAATTCCATAACCGCTATAAACTTCTCTTCAACCTATATCAAT   6486
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          6187 ATTATATAAATTCCATAACCGCTATAAACTTCTCTTCAACCTATATCAAT   6236

MvNW          6487 ATTTTTCAATAAATCATAACCAAAATTTAAAAAATGATAATTTAAACCAT   6536
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          6237 ATTTTTCAATAAATCATAACCAAAATTTAAAAAATGATAATTTAAACCAT   6286

MvNW          6537 AAGTAGAAAGTCTAGGTATAAATCATATTATTCTTAAAAAATTTCTAACT   6586
                   ||||||||||||||||||||||||||||||||||||||.|||||||.|
MvOW          6287 AAGTAGAAAGTCTAGGTATAAATCATATTATTCTTAAAAAACTTCTAATT   6336

MvNW          6587 TCATAACTAATAAGAAACTTATTAATTGAATAAATATTTATATTTCTAAC   6636
                   |||||||||||||.|||.||||||.|||||||||||||||||||||||||.
MvOW          6337 TCATAACTAATAATAAATTTATTTATTGAATAAATATTTATATTTCTAAT   6386

MvNW          6637 TAAATAACCTAATAGTAAACCTAAAATTCTAACATAAATTACTATTATTT   6686
                   |||.||||||||||.||||||||||.||||||||||||||||||||||.|
MvOW          6387 TAAGTAACCTAATAATAAACCTAAAATACTAACATAAATTACTATTATCT   6436
```

```
MvNW          6687 TTAAATTAAAAGGCAAATAAATTATATAAGGATAAGGAAAAATTATTCAC   6736
                   |.|.|||||||||.|||||||||||||||||||||||||||||||||.
MvOW          6437 TCATATTAAAAGGTAAATAAATTATATAAGGATAAGGAAAAATTATTCAT   6486

MvNW          6737 ATTAATATACTTCCTCTAATAATTCTTATAAATAATAAAATAAATATACT   6786
                   ||||||.||.||.|||||||||||||||||||||||||||||||||||||
MvOW          6487 ATTAACATTCTACCTCTAATAATTCTTATAAATAATAAAATAAATATACT   6536

MvNW          6787 TTTTAATATAGTAAAATCTTCATCATATAAATTATAAATAGATAATAAAT   6836
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          6537 TTTTAATATAGTAAAATCTTCATCATATAAATTATAAATAGATAATAAAT   6586

MvNW          6837 TAAAATCATTTACTATTAAATATATTGTTAAACGAAATCTATAAAATATT   6886
                   ||||||||||.|||||||||||||||||||||||||||||||||||||||.
MvOW          6587 TAAAATCATTCACTATTAAATATATTGTTAAACGAAATCTATAAAATATA   6636

MvNW          6887 GTTAATCCTGTAGAAATATAATATAATAAAAAAATAAAAAAATTTAAATT   6936
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          6637 GTTAATCCTGTAGAAATATAATATAATAAAAAAATAAAAAAATTTAAATT   6686

MvNW          6937 TCTTATTCTAACTATTTCTAAAATTAAATCCTTAGAATAAAAACCAGCTA   6986
                   |||||||||.|||||||||||||||||||||||||||||||.|||||||
MvOW          6687 TCTTATTCTTACTATTTCTAAAATTAAATCCTTAGAATAAAATCCAGCTA   6736

MvNW          6987 AAAAAGGAATACCACATAAAGCTATATTAGAAATATTTATACATAAAGAA   7036
                   |.|||||.|||||||||||||||||||||||||||||||||||||||||||
MvOW          6737 AGAAAGGGATACCACATAAAGCTATATTAGAAATATTTATACATAAAGAA   6786

MvNW          7037 GTTAAAGGAATAAATCTACCAATTCCCCCTATAAAACGAATATCTTGAAT   7086
                   ||||||||||||||||||||||||||.||||||||||||||||||||||||
MvOW          6787 GTTAAAGGAATAAATCTACCAATTCCACCTATAAAACGAATATCTTGAAT   6836

MvNW          7087 ATCTAATATTATATGAATAATAACCCCAGCACATATAAATAATAAAGCTT   7136
                   ||||.||||||||||||||||.|||||||||||||||||||||||||||||
MvOW          6837 ATCTGATATTATATGAATAATAACTCCAGCACATATAAATAATAAAGCTT   6886
```

```
MvNW          7137 TAAATATAGCATGAGTTAGTAAATGAAAAAAAGCTAAATCAGGTAATCCC   7186
                   ||||||||||||||||||.||||||||||||||||||.|||||||||
MvOW          6887 TAAATATAGCATGAGTTAATAAATGAAAAAAAGCTAAATCGGGTAATCCC   6936

MvNW          7187 ATTCTTAAAATTCTTATTATTAAACCTAATTGTCTTAAAGTAGATAAAGC   7236
                   ||||||||||||||||||||||||||||||||.||||||||||||||||
MvOW          6937 ATTCTTAAAATTCTTATTATTAAACCTAATTGACTTAAAGTAGATAAAGC   6986

MvNW          7237 AATAATTTTTTTTAAATCAAATTCATAATTAGCAGAAATTCCAGCTATAA   7286
                   ||||||||||||||||||||||||||||||||||||||||.|||||||
MvOW          6987 AATAATTTTTTTTAAATCAAATTCATAATTAGCAGAAATTCCTGCTATAA   7036

MvNW          7287 ATATAGTTAAACCAGATAATAATATTAAAAATTTTATAAACATTATATCA   7336
                   ||||||||||.||||||||||||||||||||||.|||||||||
MvOW          7037 ATATAGTTAATCCAGATAATAATATTAAAAATTTTATAAATATTATATCA   7086

MvNW          7337 ACTAATAATAAATTAAAACGAATTAATAAATAAACTCCTGCTGTTACTAA   7386
                   ||||||||||||||||||||||||||||||||||.|||||||||||||||
MvOW          7087 ACTAATAATAAATTAAAACGAATTAATAAATAAACACCTGCTGTTACTAA   7136

MvNW          7387 AGTAGAAGAATGCACTAAAGCAGAAACTGGAGTTGGTGCTGCTATTGCAG   7436
                   ||||||||||||.||||||||||.||||||||||||||||||.||||
MvOW          7137 AGTAGAAGAATGAACTAAAGCAGATACTGGAGTTGGTGCTGCTATAGCAG   7186

MvNW          7437 CAGGTAATCAAGATCTAAAAGGAATTTGAGCACTTTTAGTTATAGCTGCT   7486
                   |||||||||||||||||||||||||||||||||||||||.||||||||.||.
MvOW          7187 CAGGTAATCAAGATCTAAAAGGAATTTGAGCACTTTTTGTTATAGCAGCA   7236

MvNW          7487 ATAATAATTATTCTACCGACTATTATTATATAATAATCATTCTTTATAAA   7536
                   ||||||||||||||||.|||||||||||||||||||||.||.||||||
MvOW          7237 ATAATAATTATTCTACCAACTATTATTATATAATAATCATTTTTCATAAA   7286

MvNW          7537 TTCTAAATAAAAAATATAATTTCATCTTCCATAATTTATTATTCAAGAAA   7586
                   |||||||||||||||||||||||.||.||||||||||||||||||||||
MvOW          7287 TTCTAAATAAAAAATATAATTTCAACTCCCATAATTTATTATTCAAGAAA   7336
```

```
MvNW          7587 TAACTATTAAAATAAATACATCCCCAATTCGATTAGATAAAGCAGTTAAT   7636
                   |||||||||||||||||||||.||||||||||||||..||.|||||||
MvOW          7337 TAACTATTAAAATAAATACATCACCAATTCGATTAGATAGGGCTGTTAAT   7386

MvNW          7637 ATCCCAGCATTATATGATTTAATATTTTGATAATAAATTACCAAACAATA   7686
                   ||.|||||||||||||||||||||||||||||||||.||.||||||||
MvOW          7387 ATTCCAGCATTATATGATTTAATATTTTGATAATAAATGACTAAACAATA   7436

MvNW          7687 AGATACTAAACCTAAACCATCTCAACCTAATAAAATTCTAATAATATTAG   7736
                   |||||||||.||||||.||||||.||||||||||||||||||||||||
MvOW          7437 AGATACTAATCCTAATCCATCCCAACCTAATAAAATTCTAATAATATTAG   7486

MvNW          7737 GACTAATAATTAATAAAATCATTGAAAATACGAATAGTAAAACTAATATA   7786
                   ||||||||||||||||||.||||||||.||.||||.|||.|||||||||
MvOW          7487 GACTAATAATTAATAAAATTATTGAAAAAACAAATAATAATACTAATATA   7536

MvNW          7787 ATAAAACGATTTAAATTTAATTCTGAACTTATATATCTTTTTCTATAAAA   7836
                   ||||||||||||||||||||||||||||||||||.|||||||||||||||
MvOW          7537 ATAAAACGATTTAAATTTAATTCTGAACTTATATAACTTTTTCTATAAAA   7586

MvNW          7837 AATAACCACTGAAGAAATTAATATAACAAATATTATAAATAATAAAGATA   7886
                   ||||||.|||||||||||||||||||||||||||||||||||||||||.|
MvOW          7587 AATAACTACTGAAGAAATTAATATAACAAATATTATAAATAATAAAGACA   7636

MvNW          7887 TTCAATCTAATAAAATTCTTATTACAATTCTTATTGAATTGAATGAAATT   7936
                   |||||||||||||||||||||||||||||||||||||.|||||||||||
MvOW          7637 TTCAATCTAATAAAATTCTTATTACAATTCTTATTGAATTAAATGAAATT   7686

MvNW          7937 AATTCTCACTCTAAAAAAATAACTAAATTATTTATAATAAAATAAATTAT   7986
                   ||||||||.|||||||||||||.|||||||||||||||||||||||||||
MvOW          7687 AATTCTCATTCTAAAAAAATAATTAAATTATTTATAATAAAATAAATTAT   7736

MvNW          7987 CATAAAAAAATTTATTATACTTAAAAAAAATAAAAAAAATCTTCTAATAA   8036
                   .||||||||||||||||||||||||||||||||||||||||||||||..|
MvOW          7737 TATAAAAAAATTTATTATACTTAAAAAAAATAAAAAAAATCTTCTAACGA   7786
```

```
MvNW          8037 AACAAATTGAATTTTTAAAAATAACTTAAAATGATATTTATCATATTTTC   8086
                   |||||||||||||||||||||||||.||||||||||||||||||.|||
MvOW          7787 AACAAATTGAATTTTTAAAAATAACCTAAAATGATATTTATCATATATTC   7836

MvNW          8087 GACACCACAAATCAATATTTTAATTTAAATTATTTAAGTTAAAATCAAAT   8136
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          7837 GACACCACAAATCAATATTTTAATTTAAATTATTTAAGTTAAAATCAAAT   7886

MvNW          8137 TATTACATAATCAATCTTTATAACTAAAATATTTAAAGGTAATCAATGTA   8186
                   |||||||||||||||||||||||||.|||||||||||||||||||||||||
MvOW          7887 TATTACATAATCAATCTTTATAATTAAAATATTTAAAGGTAATCAATGTA   7936

MvNW          8187 ATATTATTAATAAATATTCACGAGATAAGCCTGTATAAAATCTATAAATT   8236
                   |||||||||||||||||||||||||||||.|||.||||||||||||||||
MvOW          7937 ATATTATTAATAAATATTCACGAGATAAACCTATATAAAATCTATAAATT   7986

MvNW          8237 CCTGAATAATATTTACCATGTTGAATATAAGAATATAAATATAGTCTATA   8286
                   |||||||||||||||||||||||||||||||||||||||||||.|||||||
MvOW          7987 CCTGAATAATATTTACCATGTTGAATATAAGAATATAAATATAATCTATA   8036

MvNW          8287 ACCAGCTCTAAAAAAAGAAATAAATATTAATATTAATATTGAAATTCAAG   8336
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          8037 ACCAGCTCTAAAAAAAGAAATAAATATTAATATTAATATTGAAATTCAAG   8086

MvNW          8337 ATCATCTTATTAATCTATTAATTAAACTAATTTCTCCCATTAAATTTAAA   8386
                   |||||||||.|||||||||||||||.|||||||||.|||||||||||||||
MvOW          8087 ATCATCTTAATAATCTATTAATTAGACTAATTTCACCCATTAAATTTAAA   8136

MvNW          8387 GAAGGTGGTGCTGCTATATTTGAAGATATAATTAAAAATCATCATAGTCT   8436
                   ||||||||||||||||||||||||||||||||||||||||||.|.|||||
MvOW          8137 GAAGGTGGTGCTGCTATATTTGAAGATATAATTAAAAATCATCACAATCT   8186

MvNW          8437 TATTGAAGGTATAAAATTTATTATACCCTTATTAATATATAATCTTCGTC   8486
                   |||||||||||||||||||||.||||||||||.||||||||.|||||||||
MvOW          8187 TATTGAAGGTATAAAATTTATCATACCCTTATTGATATATAAACTTCGTC   8236
```

```
MvNW            8487 TATGTAATCGTTCATAATTAATATTAGCTAAACAAAATATACCAGAAGAA   8536
                     ||||||||||||||||||||||||.|||||||||||||||.|||||||||
MvOW            8237 TATGTAATCGTTCATAATTAATATTTGCTAAACAAAATATCCCAGAAGAA   8286

MvNW            8537 CATAATCCATGACCAATTATTAAAATATAAGAACCAATAAATCCTCAATA   8586
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            8287 CATAATCCATGACCAATTATTAAAATATAAGAACCAATAAATCCTCAATA   8336

MvNW            8587 ATTTATTGTTATAATTCCACCAATAACTATTCTTATATGAGCAACCGAAG   8636
                     ||||||.||||||||.||.||||||||||||||||||||||||.||||
MvOW            8337 ATTTATAGTTATAATCCCCCCAATAACTATTCTTATATGAGCAACTGAAG   8386

MvNW            8637 AATAAGCAATTAATGATTTAATATCAACTTGACATAAACACTTTAATCTA   8686
                     ||||||||||||||.||||||||||||||||||||.||||||.|||||||||
MvOW            8387 AATAAGCAATTAAAGATTTAATATCAACTTGACAAAAACATTTTAATCTA   8436

MvNW            8687 ATATAAAATCCACCTACTAAACTAATAGTAATAAAAATGATATTATATTT   8736
                     ||||||||.||.||.||||||||||||||||||||||.||||||||||||
MvOW            8437 ATATAAAAACCTCCAACTAAACTAATAGTAATAAAAATAATATTATATTT   8486

MvNW            8737 TAAATTTATATTTTGTAATATAATTATTAAACGAATTAAACCATAACCTC   8786
                     ||||||||||||||||||||||||||||||||||||||.|||||||.||||
MvOW            8487 TAAATTTATATTTTGTAATATAATTATTAAACGAATTAGACCATACCCTC   8536

MvNW            8787 CTAATTTTAATATAATACCAGCTAAGATTATAGAACCTGAAACTGGTGCT   8836
                     ||||.||||||||||.||.||||||.||||||||||||||||||.|||
MvOW            8537 CTAACTTTAATATAATTCCTGCTAAAATTATAGAACCTGAAACTGGAGCT   8586

MvNW            8837 TCAACATGAGCTTTAGGTAATCATAAATGAACAAAATATATTGGTATTTT   8886
                     ||||||||.||||||||||||||||||||||||||||||||||||||||||
MvOW            8587 TCAACATGGGCTTTAGGTAATCATAAATGAACAAAATATATTGGTATTTT   8636

MvNW            8887 AACTAAAAAAGCTATAATTATAGAAAAATATAATAAATATAAATCAAAAT   8936
                     .||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            8637 TACTAAAAAAGCTATAATTATAGAAAAATATAATAAATATAAATCAAAAT   8686
```

```
MvNW            8937 TAAAAAATTTCATAAAATAAATTATAATATGATTTACTTCATTAAAAATA   8986
                     ||||||||||.||||||||||||||||||.||||||||||||||||||
MvOW            8687 TAAAAAATTTTATAAAATAAATTATAATATAATTTACTTCATTAAAAATA   8736

MvNW            8987 TAAAAAATACCTAATAATAAAGGTAAAGAAACAAATAAAGTATAAAATAA   9036
                     ||||||||||||||||||.||||||||||||||||||||||||||||||||
MvOW            8737 TAAAAAATACCTAATAACAAAGGTAAAGAAACAAATAAAGTATAAAATAA   8786

MvNW            9037 TAAATATATTCCAGCTTGAATTCGTTCAGGTTGATAACCTCAACCAATAA   9086
                     ||||||||||||||||||||||.||||||||||||||||||||||||||||
MvOW            8787 TAAATATATTCCAGCTTGAATCCGTTCAGGTTGATAACCTCAACCAATAA   8836

MvNW            9087 TTAATAATAATGTAGGAATTAATCTTCCTTCAAAAAATAAATAAAATATA   9136
                     ||||||.|||.|||||||||||||||.||.||||||||||||||||||||
MvOW            8837 TTAATATTAAAGTAGGAATTAATCTCCCCTCAAAAAATAAATAAAATATA   8886

MvNW            9137 AATATATTTATAACTCTAAAAGTTAAATATAATATTATTAATAGAAAAAT   9186
                     ||||||.|||||||||||||||||||||||||||||||||||||.|||||||
MvOW            8887 AATATGTTTATAACTCTAAAAGTTAAATATAATATTATTAATAAAAAAAT   8936

MvNW            9187 TAAATTAAATAAAAAAAAATTTAAATAATAATCTTCTTTATATAAATTTT   9236
                     ||||||.|||||||||||||||||||||||.|.|..||||||||||||||
MvOW            8937 TAAATTGAATAAAAAAAAATTTAAATAATAACCCTGCTTATATAAATTTT   8986

MvNW            9237 CACTAGCCATAATTATTAAAATACAAATTCAAACTCTTAATATAATTAAA   9286
                     |.||||||.|||||||||||||||||||||||||||||||||.|||.|||
MvOW            8987 CTCTAGCTATAATTATTAAAATACAAATTCAAACTCTTAATAAAATCAAA   9036

MvNW            9287 CCATAAGATAAAATATCACATGAATATATATAACTAAAATTACTATAAGT   9336
                     ||||||||||||||||||||||||||||||||.||||||||||.|||||||
MvOW            9037 CCATAAGATAAAATATCACATGAATATATATATCTAAAATTACAATAAGT   9086

MvNW            9337 TTCAATACTTAATGTTAAATTTATTAATAAAAATATTATAAAAAATAAAA   9386
                     ||||||.||||||||||||||||||||||||||||||||||||||||||||
MvOW            9087 TTCAATTCTTAATGTTAAATTTATTAATAAAAATATTATAAAAAATAAAA   9136
```

```
MvNW            9387 TTATTTGAACCATTCAATATATATTAAAATTAAAACATAAAGGAATTATA   9436
                     ||||||||||||||||||||||||||||||||||||||||||||||||.|
MvOW            9137 TTATTTGAACCATTCAATATATATTAAAATTAAAACATAAAGGAATTAAA   9186

MvNW            9437 AAAATTATTATAAATAAAAATTTTATCATTATAATAAATTAAAACTTTGA   9486
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            9187 AAAATTATTATAAATAAAAATTTTATCATTATAATAAATTAAAACTTTGA   9236

MvNW            9487 AAATAATCATTTCCATGAGTACGAATTATTGAAACTAAAATTGATAAACC   9536
                     ||||||||||.||||||||||||||||||||||||||||||||||||||||
MvOW            9237 AAATAATCATTACCATGAGTACGAATTATTGAAACTAAAATTGATAAACC   9286

MvNW            9537 TAAAGCTCCCTCACAAACTGAGAAAACTAAAAAAACTATTAATATATATA   9586
                     ||||||.||.||||||.||.||.||||||||||||||||||||||||||||
MvOW            9287 TAAAGCCCCTTCACATACGGAAAAAACTAAAAAAACTATTAATATATATA   9336

MvNW            9587 TATCATATTCAATATAATTAAATAATAAAATTATAAAAAAAAAAAATTCTT   9636
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            9337 TATCATATTCAATATAATTAAATAATAAAATTATAAAAAAAAAAAATTCTT   9386

MvNW            9637 AAAACAATAAATTCTAATCTTAATAAAACAATCAATAAATGCTTATGTTT   9686
                     ||||||||||.||||||||||||||||||.||||||||||||||||||||
MvOW            9387 AAAACAATAAACTCTAATCTTAATAAAACAATTAATAAATGCTTATGTTT   9436

MvNW            9687 AGAAACAAAAATTATATTACCTAAAATAAATATAATAATAACTAAAATTC   9736
                     ||||||||||||||||||.||||||||||.||||||||||||||||||||
MvOW            9437 AGAAACAAAAATTATATTTCCTAAAATAAACATAATAATAACTAAAATTC   9486

MvNW            9737 ATATATTTATAACTATCATTAGTTTTTATAGTTTAAAATAAAACATTGGT   9786
                     ||||                                        ||||||
MvOW            9487 ATAT----------------------------------------ATTGGT   9496

MvNW            9787 CTTGTAAATCAAAAATAAGATAAATTTTTTAAAAACATCAAAGAAAAAGA   9836
                     |.||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW            9497 CCTGTAAATCAAAAATAAGATAAATTTTTTAAAAACATCAAAGAAAAAGA   9546
```

```
MvNW          9837 TTTTTCTTTATCAATAATCTCCAAAATTATTATTTTTATTAAACTATTCT   9886
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          9547 TTTTTCTTTATCAATAATCTCCAAAATTATTATTTTTATTAAACTATTCT   9596

MvNW          9887 TTGAAATTATTAAAATATTCTTATCTCTATTAATTATTATATTTTCTTTT   9936
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          9597 TTGAAATTATTAAAATATTCTTATCTCTATTAATTATTATATTTTCTTTT   9646

MvNW          9937 TTTATAATTTTTTTAAATCATCCTTTATCAATAGGATTAATAATTTTAAT   9986
                   .|.|||||||||.||||||||||||||||||||||||||||||||||||||
MvOW          9647 CTAATAATTTTTCTAAATCATCCTTTATCAATAGGATTAATAATTTTAAT   9696

MvNW          9987 TCAAACTATATTAACTTGTTTAATTTCAAGAATTATAATATCAACATATT  10036
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          9697 TCAAACTATATTAACTTGTTTAATTTCAAGAATTATAATATCAACATATT   9746

MvNW         10037 GATTCTCTTATATTTTATTTTTAACCTTTTTAGGAGGATTATTAGTATTA  10086
                   |||||||||||||.|||||||.||||||||||||||||||||||||||||||
MvOW          9747 GATTCTCTTATATCTTATTTCTAACCTTTTTAGGAGGATTATTAGTATTA   9796

MvNW         10087 TTTATTTATGTATCTAGAATTGCATCAAATGAAATATTTACAATTTCATT  10136
                   |||||||||||||||.||||||||||||||||||||||||||||||||||||
MvOW          9797 TTTATTTATGTATCAAGAATTGCATCAAATGAAATATTTACAATTTCATT   9846

MvNW         10137 TACTATAAAAATCATAATAATAATTTGTTTTATTATTATTATTATTGTAA  10186
                   ||||||||||||.||||||||||..|..|||.|||.|||||||||.|||
MvOW          9847 TACTATAAAAATAATAATAATAATAAGAATTACTATCATTATTATTATAA   9896

MvNW         10187 GAATTATTAATATAAATAATTTAAAATGAATAAATTTTAATACAAATTTA  10236
                   ||||||||||.||||||||||||||||||||||||||||||||||||||||
MvOW          9897 GAATTATTAAAATAAATAATTTAAAATGAATAAATTTTAATACAAATTTA   9946

MvNW         10237 GAAATAAATAATTTTTTTAATAAATTCATATTTTTTAATAATGAAAATAA  10286
                   ||||||||||||.||||||||||||||||||||||||||||||||||||||
MvOW          9947 GAAATAAATAAATTTTTTAATAAATTCATATTTTTTAATAATGAAAATAA   9996
```

```
MvNW          10287 AATTAATTTATCTAAATTATATAATAACCAAACATTTTTATTAATAATAA  10336
                    |||||||||||||||||||||||||||||||||||||||.|||||||||
MvOW           9997 AATTAATTTATCTAAATTATATAATAACCAAACATTTTTAATAATAATAA  10046

MvNW          10337 TAATAATTATTTACTTATTTATTACATTAATTGCAGTAGTAAAAATCACA  10386
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10047 TAATAATTATTTACTTATTTATTACATTAATTGCAGTAGTAAAAATCACA  10096

MvNW          10387 AATATTTTTTATGGTCCTTTACGATCTTCATTTTAACAAATGATAAATAT  10436
                    |||||||||||||||||||||||||||||||||| ||||||||||||||
MvOW          10097 AATATTTTTTATGGTCCTTTACGATCTTCATTTTAA-AAATGATAAATAT  10145

MvNW          10437 ATTTAAACCAATTCGAAAAACACACCCTATTTTAAAAATTATTAATGGAT  10486
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10146 ATTTAAACCAATTCGAAAAACACACCCTATTTTAAAAATTATTAATGGAT  10195

MvNW          10487 CTTTTGTAGATCTACCATCTCCATCAAATATTTCATCATTATGAAATTTT  10536
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10196 CTTTTGTAGATCTACCATCTCCATCAAATATTTCATCATTATGAAATTTT  10245

MvNW          10537 GGATCACTTTTATTTATATGCTTAATAATTCAAATTATTACTGGATTATT  10586
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10246 GGATCACTTTTATTTATATGCTTAATAATTCAAATTATTACTGGATTATT  10295

MvNW          10587 TTTAACTATATATTATACAGCAAATATTGAATTAGCTTTTTATAGAGTAA  10636
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10296 TTTAACTATATATTATACAGCAAATATTGAATTAGCTTTTTATAGAGTAA  10345

MvNW          10637 ATTATATTTGCCGAAATGTTAATTATGGATGATTAATTCGAACTTTACAT  10686
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10346 ATTATATTTGCCGAAATGTTAATTATGGATGATTAATTCGAACTTTACAT  10395

MvNW          10687 GCAAATGGAGCATCTTTTTTTTTTATTTGTATTTATATTCATATTGGACG  10736
                    |||||||||||||||||||||||.|||||||||||||||||||||||||
MvOW          10396 GCAAATGGAGCATCTTTTTTTTTTTTTTTGTATTTATATTCATATTGGACG  10445
```

```
MvNW          10737 AGGAATTTATTATGAATCTTTTAATTACAAATACACATGAATAGTAGGTG  10786
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10446 AGGAATTTATTATGAATCTTTTAATTACAAATACACATGAATAGTAGGTG  10495

MvNW          10787 TAATTATTTTATTTTTATTAATAGCAACAGCTTTTATAGGATATGTTCTC  10836
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10496 TAATTATTTTATTTTTATTAATAGCAACAGCTTTTATAGGATATGTTCTC  10545

MvNW          10837 CCTTGAGGACAAATATCATTTTGAGGTGCAACTGTTATTACTAATTTATT  10886
                    ||||||||||||||||||||||||||.||||||||||||||||||||||||
MvOW          10546 CCTTGAGGACAAATATCATTTTGAGGAGCAACTGTTATTACTAATTTATT  10595

MvNW          10887 ATCTGCCATCCCTTATTTAGGTACAACATTAGTAAATTGAATTTGAGGTG  10936
                    ||||||||||||||||||||||.||||||||||||||||||||||||||.|
MvOW          10596 ATCTGCCATCCCTTATTTAGGAACAACATTAGTAAATTGAATTTGAGGAG  10645

MvNW          10937 GATTTGCTATTGATAATGCCACTTTAACTCGATTTTATACTTTTCATTTT  10986
                    ||||||||||||||||..||||||||||.||||||||||||||||||||||
MvOW          10646 GATTTGCTATTGATAACACCACTTTAACGCGATTTTATACTTTTCATTTT  10695

MvNW          10987 ATTTTACCTTTTATTATTTTAATAATAAGAATAATTCATTTATTATTCCT  11036
                    ||||||||||||||||||.||||||||||||||||||||||.|||||||
MvOW          10696 ATTTTACCTTTTATTATATTAATAATAAGAATAATTCATTTACTATTCCT  10745

MvNW          11037 TCATCAAACAGGATCTAATAATCCTTTAGGAATCAATAGAAATTTAGATA  11086
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10746 TCATCAAACAGGATCTAATAATCCTTTAGGAATCAATAGAAATTTAGATA  10795

MvNW          11087 AAATTCCTTTTCATCCATTTTTTATATTTAAGGATTTAATTGGATTTATT  11136
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10796 AAATTCCTTTTCATCCATTTTTTATATTTAAGGATTTAATTGGATTTATT  10845

MvNW          11137 TTAGTTATATTTATATTAATTTTATTAACACTTACAAACCCTTATTTATT  11186
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          10846 TTAGTTATATTTATATTAATTTTATTAACACTTACAAACCCTTATTTATT  10895
```

```
MvNW          11187 AGGAGATCCAGATAATTTTATCCCTGCCAATCCATTAGTAACTCCAATTC  11236
                    ||||||.||||||||||.|||||.|||||.||||||||||||||||||
MvOW          10896 AGGAGACCCAGATAATTTCATCCCAGCCAACCCATTAGTAACTCCAATTC  10945

MvNW          11237 ATATTCAACCAGAATGATATTTTTTATTTGCTTATGCTATTTTACGATCA  11286
                    ||||||||||||||||||||||||||||||.||||||||||||||||||
MvOW          10946 ATATTCAACCAGAATGATATTTTTTATTTGCCTATGCTATTTTACGATCA  10995

MvNW          11287 ATTCCTAATAAATTAGGGGGAGTTATTGCTTTAGTAATATCAATTCTTAT  11336
                    ||||||||||||||||.||||||||||||||||||||||||||||||||
MvOW          10996 ATTCCTAATAAATTAGGAGGAGTTATTGCTTTAGTAATATCAATTCTTAT  11045

MvNW          11337 TTTAATTATTTTACCAATAACTTTTATAAAAAAAAATACAAGGAATTCAAT  11386
                    ||||||||||||||||||||||||||.||||||||.|||||||||||||||
MvOW          11046 TTTAATTATTTTACCAATAACTTTTAAAAAAAAAAAACAAGGAATTCAAT  11095

MvNW          11387 TTTATCCATTAAATCAAATTATATTTTGAATAATAGTAACAACAATTATT  11436
                    |||||.||||||||||||||||||||||||||||||||||||||||||||
MvOW          11096 TTTACCCATTAAATCAAATTATATTTTGAATAATAGTAACAACAATTATT  11145

MvNW          11437 TTATTAACATGAATTGGAGCACGACCTGTAGAAGATCCTTATATTATTGT  11486
                    .||.||||||||||||||||||||||||||||.||||||.||||||.||
MvOW          11146 CTACTAACATGAATTGGAGCACGACCTGTAGAAGACCCTTACATTATCGT  11195

MvNW          11487 GGGACAAATTTTAACAATTTTATATTTTTCATATTATATCTTTAATCCTT  11536
                    .|||||||||||||||||||||||.||.||||||||||||||||||||.|
MvOW          11196 AGGACAAATTTTAACAATTTTATACTTCTCATATTATATCTTTAATCCCT  11245

MvNW          11537 TAGTTAGTATATACTGAGATAAATTAATTTTTAATTAATTAATGAGCTTG  11586
                    ||.||||.||||||||||||||||||||||||||||||||||||||||||
MvOW          11246 TAATTAGAATATACTGAGATAAATTAATTTTTAATTAATTAATGAGCTTG  11295

MvNW          11587 TAAAAGCATTTGTTTTGAAAACTTAAGAAAGAATATATTATTCTATTAAT  11636
                    ||||||||||||||||||||||||||||||||||.|||||||||||||||
MvOW          11296 TAAAAGCATTTGTTTTGAAAACTTAAGAAAGAATAAATTATTCTATTAAT  11345
```

```
MvNW          11637 TTATACTAAAAATAATATAATAATTAAAAGAAAAAAATTTTTAATCCTAA  11686
                    ||||||||||||||||||     .||||||||||||||||||||||||||
MvOW          11346 TTATACTAAAAATAATAT-----CTAAAAGAAAAAAATTTTTAATCCTAA  11390

MvNW          11687 ATAAAATATCATAAAATTTAATGATAAAGGTAAATAAATTTTTCAAGCTA  11736
                    |||||||||.|||||||||||||||||||||||||||||.||||||||
MvOW          11391 ATAAAATATTATAAAATTTAATGATAAAGGTAAATAAATTTCTCAAGCTA  11440

MvNW          11737 AATATATTAATTTATCATATCGATAACGAGGTAAAGTCCCTCGAACTCAA  11786
                    ||||||||.|.||||||||||||||||||||||||||||||||||||||
MvOW          11441 AATATATTGACTTATCATATCGATAACGAGGTAAAGTCCCTCGAACTCAA  11490

MvNW          11787 ATAAATAAAAAAGAAATTAATCTTAATTTTAAATAAAAAAAGAAATCTAA  11836
                    |||||||||||||||||||||||.||.||.||||||||||.||||||||
MvOW          11491 ATAAATAAAAAAGAAATTAATCTCAACTTCAAATAAAAAAAAAAATCTAA  11540

MvNW          11837 TGAAAATCCACCTATATATAATAAAATAAATAATAATCTTATAAATAAAA  11886
                    |||||||||.|||||||||||||||.||||||||||||||||||||||||
MvOW          11541 TGAAAATCCTCCTATATATAATAAAACAAATAATAATCTTATAAATAAAA  11590

MvNW          11887 TTCTAGAATATTCAGCTAAAAAAATTAATGCAAATCCACCTCTTCTATAT  11936
                    |||||||||||||||||||||||||||||||||||.||.||||||||||||
MvOW          11591 TTCTAGAATATTCAGCTAAAAAAATTAATGCAAAACCCCCTCTTCTATAT  11640

MvNW          11937 TCAATATTAAATCCTGAAACTAATTCTCTTTCCCCTTCAGCAAAATCAAA  11986
                    |||||||||||||||||||||||||||||||||||.|||||||||||||||
MvOW          11641 TCAATATTAAATCCTGAAACTAATTCTCTTTCCCCCTCAGCAAAATCAAA  11690

MvNW          11987 AGGAGTACGATTAGTCTCAGCTAATCTAGAAGAAAATCAACATATTCTTA  12036
                    ||||||.||||||||.||.||||||||||||||||||||||||||||||||
MvOW          11691 AGGAGTCCGATTAGTTTCTGCTAATCTAGAAGAAAATCAACATATTCTTA  11740

MvNW          12037 ACGGAATTATTAAAAAAAAAAATCAAATTAAATTTTGATAATATGAAAAA  12086
                    |.|||||||||||||||||||||||||||||||||||||||.||||||||
MvOW          11741 ATGGAATTATTAAAAAAAAAAATCAAATTAAATTTTGATAATAAGAAAAA  11790
```

```
MvNW          12087 CTAATTATATTAAAATCTATAACTATAATAATTCTAGATATTAAAATTAA  12136
                    ||||||||||||||||||||||.||||||||||||||||||||||||||
MvOW          11791 CTAATTATATTAAAATCTATAATTATAATAATTCTAGATATTAAAATTAA  11840

MvNW          12137 AGCTAAACTAACTTCATAAGAAATAGTTTGAGCAACAGCTCGTAAACCTC  12186
                    ||||||||||||||||||||||||||||||||||.||||||||||||||||
MvOW          11841 AGCTAAACTAACTTCATAAGAAATAGTTTGAGCTACAGCTCGTAAACCTC  11890

MvNW          12187 CTAATAAAGAATAATTAGAATTAGAAGATCAACCAGCAATTATAACTGTA  12236
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          11891 CTAATAAAGAATAATTAGAATTAGAAGATCAACCAGCAATTATAACTGTA  11940

MvNW          12237 TATACACCCATTCTCGTACAACATAAAAAAAATAAAATACCTAAATTAAA  12286
                    ||.||||||.||||||.||||||||||||||||||||||||||||||||||
MvOW          11941 TAAACACCTATTCTTGTACAACATAAAAAAAATAAAATACCTAAATTAAA  11990

MvNW          12287 TCTAATAAAATTAAAATAATAAGGAATTATTATTCAAATTATTAAAGACA  12336
                    ||||||||||||||||||||||||||||||||||||||||||||||||||.|
MvOW          11991 TCTAATAAAATTAAAATAATAAGGAATTATTATTCAAATTATTAAAGATA  12040

MvNW          12337 AAATAAATCTAATGACAGGAGAAAAATAATAAATTAAATAATTAGAAAAT  12386
                    ||||||||||||||.||||||||||||||||.||||||||||||||||||
MvOW          12041 AAATAAATCTAATTACAGGAGAAAAATAATAAGTTAAATAATTAGAAAAT  12090

MvNW          12387 TTAGGATAAGTTTGTTCTTTAGTAAATAACTTAATAGCATCTGAAAAAGG  12436
                    ||||||||||||||||||||||||||||||.||||||||||||||||||||
MvOW          12091 TTAGGATAAGTTTGTTCTTTAGTAAATAATTTAATAGCATCTGAAAAAGG  12140

MvNW          12437 TTGTAAAATTCCTATTAAACCAACCTTATTAGGCCCTTTACGAATTTGAA  12486
                    ||||||||||||||.||||||||.||||||||||||||||||||||||||
MvOW          12141 TTGTAAAATTCCTATAAAACCAACTTTATTAGGCCCTTTACGAATTTGAA  12190

MvNW          12487 TATAACCTAAAACTTTTCGCTCTAATAAAGTTAAGAAAGCAACACCAATT  12536
                    ||||.||||||||.||.||.|||
MvOW          12191 TATATCCTAAAACCTTACGTTCT--------------------------  12213
```

```
MvNW          12537 AAAACCCCTACAATTAAAATTAATAARCCTAAATTAGAAAATTTAGGATA  12586

MvOW          12214 --------------------------------------------------  12213

MvNW          12587 AGTTTGTTCTTTAGTAAATAACTTAATAGCATCTGAAAAAGGTTGTAAAA  12636

MvOW          12214 --------------------------------------------------  12213

MvNW          12637 TTCCTATTAAACCAACCTTATTAGGCCCTTTACGAATTTGAATATAACCT  12686

MvOW          12214 --------------------------------------------------  12213

MvNW          12687 AAAACTTTTCGCTCCAATAAAGTTAAGAAAGCAACACCAATTAAAACCCC  12736
                           ||||||||||.||||||||||||||||.||.||
MvOW          12214 ---------------AATAAAGTTAAAAAAGCAACACCAATTAATACTCC  12248

MvNW          12737 TACAATTAAAATTAATAAACCTAAAAAAATTAAAATTATATCTAATATTA  12786
                     ||.|||||||||||||||||||||||||||||||||||.|||||||
MvOW          12249 TAAAATTAAAATTAATAAACCTAAAAAAATTAAAATTATATCCAATATTA  12298

MvNW          12787 TCATTACTACCTATATAAATAAAATTATACATTTATGATTTCTAAAACCA  12836
                     ||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          12299 TCATTACTACCTATATAAATAAAATTATACATTTATGATTTCTAAAACCA  12348

MvNW          12837 TTACATTTTTCTGCCAAAATAGCTTAATAAATTATTATAATATTTATTTT  12886
                     |||||||||||||||||||||||||||..||..||||.||||.|||
MvOW          12349 TTACATTTTTCTGCCAAAATAGCTTAATAAAACATATTAATTTTTAATTT  12398

MvNW          12887 ATAAATTCATTAATAAAATTCTT-TAATTAAATTTAATTTAAATATTTAT  12935
                     |||.|||..|||||||||||||| .||.|||||||||||||||||||||
MvOW          12399 ATATATTTGTTAATAAAATTCTTCAAACTAAATTTAATTTAAATATTTAT  12448

MvNW          12936 TCCTTTCGTACTAAAATATTTTTTTTATTTAAAGATAGAAACCAACCTGG  12985
                     ||||||||||||||||||.|||||||..||||||||||||||||||||
MvOW          12449 TCCTTTCGTACTAAAATATCTTTTTTGCTTAAAGATAGAAACCAACCTGG  12498
```

```
MvNW          12986 CTCACACCGGTTTGAACTCAGATCATGTAAGATTTTAATGATCGAACAGA  13035
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          12499 CTCACACCGGTTTGAACTCAGATCATGTAAGATTTTAATGATCGAACAGA  12548

MvNW          13036 TCAAAATTTTAGACTTTTGCA-TATAAATTTTATCTTAATCCAACATCGA  13084
                    || ||||||||.|||||||||  ||||||  ||||||||||||||||||||
MvOW          12549 TC-AAATTTTAAACTTTTGCATTATAAA--TTATCTTAATCCAACATCGA  12595

MvNW          13085 GGTCGCAAACTTTTTTTTTTATTTGAACTAAAAAAAAAAAATTACGCTGTT  13134
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          12596 GGTCGCAAACTTTTTTTTTTATTTGAACTAAAAAAAAAAAATTACGCTGTT  12645

MvNW          13135 ATCCCTAAGGTAATTTTTTCTTTTAATCAAAATAATTTTGGATCATATAC  13184
                    |||||||||||||||||||||||.|||||||||||||||||||||||.|||.
MvOW          12646 ATCCCTAAGGTAATTTTTTCTTCTAATCAAAATAATTTTGGATCAAATAT  12695

MvNW          13185 TCACTTATTAATGAATCTTATTAAAAAAAAGTTAATTATATTTTTCTATC  13234
                    |||||||||||||||.||...|||||||||||||.|||||||||.||||
MvOW          12696 TCACTTATTAATGAATTTTTAAAAAAAAAAGTTAAATATATTTTTTTATC  12745

MvNW          13235 ACCCCAACAAAATAAATATTATTATTTTAATTATTAACTATATAAATAAT  13284
                    |||||||||||||||||||||||||||||||||||||.|||||||||||||
MvOW          12746 ACCCCAACAAAATAAATATTATTATTTTAATTATTAAATATATAAATAAT  12795

MvNW          13285 TTTAATAAAAAAATTTATCAAACTCTATAGGGTCTTCTCGTCTTTTAAGT  13334
                    ||.||||||||..||||||.|||||||||||||||||||||||||||||.|
MvOW          12796 TTCAATAAAAACGTTTATTAAACTCTATAGGGTCTTCTCGTCTTTTAAAT  12845

MvNW          13335 TTATTTTAACTTTTTAATTAAAAAATTAAATTTTTTTAATAAAATTGAGA  13384
                    ||||||||||||||||||||||||||||||||||||.|||||||||||||||
MvOW          12846 TTATTTTAACTTTTTAATTAAAAAATTAAATTTTTCTAATAAAATTGAGA  12895

MvNW          13385 CAGTTTATATTTCATCCAATCTTTCATACAAGTCACCAATTAAGAGACTA  13434
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          12896 CAGTTTATATTTCATCCAATCTTTCATACAAGTCACCAATTAAGAGACTA  12945
```

```
MvNW          13435 ATGATTATGCTACCTTTGTACAGTCAATATACTGCAGCCCTTTAATTAAA  13484
                    |||||||||||||||
MvOW          12946 ATGATTATGCTACCT-----------------------------------  12960

MvNW          13485 AATCAGTGGGCAGATTAGACTTTAAATTATTATTCAAAAAGACATGTTTT  13534

MvOW          12961 -------------------------------------------------  12960

MvNW          13535 TGATAAACAAGTG-----------------------------------  13547
                              ||
MvOW          12961 -----------TGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  12999

MvNW          13548 -------------------------------------------------  13547

MvOW          13000 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  13049

MvNW          13548 -----------AATATAAAATTTTGCCGAATTACTTAATTTTATTTTTCA  13586
                              |||||||||||||||||||||||||||.||||||||||||
MvOW          13050 NNNNNNNNNNNAATATAAAATTTTGCCGAATTACTTAACTTTATTTTTCA  13099

MvNW          13587 ATAAATTATTAATACTATTTAATTATTAAATACTAATTTTATCATTATAA  13636
                    ||||   |||||||.|||||||.||||||||||||||||||||||||||.
MvOW          13100 ATAA---ATTAATATTATTTATTTATTAAATACTAATTTTATCATTATAT  13146

MvNW          13637 TTAATTTTTAATTATTAAAAATTATTTTTTTATAAAAAATTAAA---CTT  13683
                    |||||.||||.||||||||||.|||||||||||||||||||||   .||
MvOW          13147 TTAATCTTTATTTATTAAAAAATATTTTTTTATAAAAAATTAAATTTTTT  13196

MvNW          13684 TTTTAAATTTTAATTTAATTAAAATTAATTTATAATAAATTAAAAATTAT  13733
                    ||||||||||||||.||||||||||||||||||||||||||||||||||
MvOW          13197 TTTTAAATTTTAATTAAATTAAAATTAATTTATAATAAATTAAAAATTAT  13246

MvNW          13734 AAACAATATAAATTTTAAAAATTTTAATT-AATACATTATTATTATTATA  13782
                    |||||||||||||||||||||||||||| ||.||   |.||||||||||
MvOW          13247 AAACAATATAAATTTTAAAAATTTTAATTAAAAAC---AATATTATTATA  13293
```

```
MvNW          13783 AATATTTAATTTAAAGCTTATCCCTTAAAATATAATTTTTTACTTATAAT  13832
                    ||||||||||||| ||||||| ||||||||||||||||||..||||||||
MvOW          13294 AATATTTAATTT-AAGCTTA-CCCTTAAAATATAATTTTTTCTTTATAAT  13341

MvNW          13833 AAATTAATTAATTAATTTATTATAAAGAAAAAATAAAATTAAATTTTTTT  13882
                    |||||||||||||||||||||||||| ||||||||||||||||||||||
MvOW          13342 AAATTAATTAATTAATTTATTATAAAG-AAAAATAAAATTAAATTTTTTT  13390

MvNW          13883 CTAAAAAAACTAGATATCTTAAAAAACGATTAACATTTCATTTCAAATTA  13932
                    |||||||||||| ||||||||||||||||||||||| |.||||||||||||
MvOW          13391 CTAAAAAAACT-GATATCTTAAAAAACGATTAACA-TCCATTTCAAATTA  13438

MvNW          13933 ATTATTAAAAATATTTATGCTACAATAACTTTTATAATTAATTATCTCTT  13982
                    ||||||.||||||||||.|| ||||||||||||||||||||||||||||||
MvOW          13439 ATTATTTAAAATATTTAGGC-ACAATAACTTTTATAATTAATTATCTCTT  13487

MvNW          13983 TTTAATTCGAGAAATATTTTAACTAAAATTTTAATTAATAAACTCTGATA  14032
                    |||||||||||||.|||||||||||||||||||||||||||||| ||||
MvOW          13488 TTTAATTCGAGAATTATTTTAACTAAAATTTTAATTAATAAACTC-GATA  13536

MvNW          14033 CACAAGATACAATAAATAAAATTTACTTTTAAATAAATTCATTTTCAAAT  14082
                    ||| ||||||||||||||||||||||||||||||||||||||||||||||
MvOW          13537 CAC-AGATACAATAAATAAAATTTACTTTTAAATAAATTCATTTTCAAAT  13585

MvNW          14083 TTATTTAAAATTTCTTATACAATACTAATTGACTATAAAATTTATAATTT  14132
                    ||||||||||.|||||||||||||||.|..|||||||||.|||||||||
MvOW          13586 TTATTTAAAACTTCTTATACAATACTATTACACTATAAAACTTATAATTT  13635

MvNW          14133 TTTTTTTATTAATACTAAAACCCCCATTTTAATATTAAAATTATTTTTAT  14182
                    ||||||||||||||| ||||||||||||||||||||||||||||||||||
MvOW          13636 TTTTTTTATTAATACT-AAACCCCCATTTTAATATTAAAATTATTTTTAT  13684

MvNW          14183 TATTTATAAATTATTAATTATTCATCTTCAAATTAATTGAATAACATCAA  14232
                    ||||||||||||||||||||||.||.|||||||| |||||||||.|||||
MvOW          13685 TATTTATAAATTATTAATTATTAATTTTCAAATT-ATTGAATAATATCAA  13733
```

```
MvNW          14233 TATCATTTCAATGTAAATGAAATACTTAATCAAGCTCTAATTTGTTATTT  14282
                    ||||||||||||||||||||||||||.||||||||||||||||.||||||
MvOW          13734 TATCATTTCAATGTAAATGAAATACTTTATCAAGCTCTAATTTGGTATTT  13783

MvNW          14283 CTAGAAACACTTTCCAGTACCTCTACTTTGTTACGACTTATTCCAATTTA  14332
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          13784 CTAGAAACACTTTCCAGTACCTCTACTTTGTTACGACTTATTCCAATTTA  13833

MvNW          14333 TTAATGAAAGCGACGGGCAATATGTACATATTTTAATTTTTAATCATTTT  14382
                    |||||||||||||||||||||||||||||.||||||||||||||||||||
MvOW          13834 TTAATGAAAGCGACGGGCAATATGTACATATCTTAATTTTTAATCATTTT  13883

MvNW          14383 AATAAATTAAATAAAATTACATTTAAATCCACTTTCAATTAATTTTTACA  14432
                    ||||||||||||||||||||||||||||||||||||||||||||||.||||||
MvOW          13884 AATAAATTAAATAAAATTACATTTAAATCCACTTTCAATTAATTATTACA  13933

MvNW          14433 AATTAATATTCATATAAATAAATTCATTGTAATCCATTATATTCTTAATT  14482
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          13934 AATTAATATTCATATAAATAAATTCATTGTAATCCATTATATTCTTAATT  13983

MvNW          14483 ATAATCTGCATCTTGATCTGATTTAATTTTTTATTTAAATTTTTAAATAT  14532
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
MvOW          13984 ATAATCTGCATCTTGATCTGATTTAATTTTTTATTTAAATTTTTAAATAT  14033

MvNW          14533 TATTTTTATTTTAAAATATTTTTATAACAACGATATACAAAATAATAAAT  14582
                    ||||.|||||||| |||||||||
MvOW          14034 TATTCTTATTTT-AAATATTTT----------------------------  14054

MvNW          14583 TAAGTAAATTTATTCGTGGATTATCAATTATTAAACAGATTCCTCTAAAT  14632

MvOW          14055 --------------------------------------------------  14054

MvNW          14633 GAACTAAAATACCGCCAAATTATTTAAGTTTCAATAAATGATTATATACT  14682

MvOW          14055 --------------------------------------------------  14054

MvNW          14683 ATTTTAGTATTATAAATTTAAATTTTTAATAATAGGGTATCTAATCCTAG  14732
```

```
MvOW          14055 ----------------------------------------------------  14054

MvNW          14733 TTTATAAATAAAATTTATTAAATCATAAATAAAAATTTAATTTTAATTAA  14782

MvOW          14055 ----------------------------------------------------  14054

MvNW          14783 ATTAAAATTTCACCTTATAATTTAATATTTAATTAAAAAAATATTAATTA  14832

MvOW          14055 ----------------------------------------------------  14054

MvNW          14833 TTAATTACTAATAAATTTTAATTTAATTTTTGTTTAACCGCAACTGCTGG  14882

MvOW          14055 ----------------------------------------------------  14054

MvNW          14883 CACAAAATTTGTTATTAATTTAAATATTACTAAATCTTAATTTCTTAAAT  14932

MvOW          14055 ----------------------------------------------------  14054

MvNW          14933 TTTTAATATTAATTACTACTTGTATTTATTAAATATTATTAAAATAGTTA  14982

MvOW          14055 ----------------------------------------------------  14054

MvNW          14983 ATAATTAACACTAAAATTTATATGCAAAATAAATTTATAATAAAATCTTT  15032

MvOW          14055 ----------------------------------------------------  14054

MvNW          15033 AAACTATAAAAAATTTATTTATTGTAGGATTTTAGACATAGTTTTTTTTT  15082

MvOW          14055 ----------------------------------------------------  14054

MvNW          15083 TTTTTTTTTATATATATAAAATTTAATATAAATTATTAAATATTAAATAT  15132

MvOW          14055 ----------------------------------------------------  14054

MvNW          15133 TTTCTTTCTTTTTCTTCTTTATAACATTAATATTAAAAATTAATACGTAG  15182

MvOW          14055 ----------------------------------------------------  14054
```

```
MvNW          15183 ATTCATCGATTAATAATCATTTAAATAAATAATTAATTAATATATTTTAA  15232
MvOW          14055 --------------------------------------------------  14054
MvNW          15233 AATTAATTAAATTGAAATTTAAAATATTAATTTTACTAAATTAATTAATT  15282
MvOW          14055 --------------------------------------------------  14054
MvNW          15283 AATTTAATTAATATTAAAAATATTAATAAATTAAATATTTAATATATATA  15332
MvOW          14055 --------------------------------------------------  14054
MvNW          15333 TATATATATAAATATAAACCGTTTTTAATATTTTTTCTATAAATAAAAAA  15382
MvOW          14055 --------------------------------------------------  14054
MvNW          15383 AAA  15385
MvOW          14055 ---  14054
```

APPENDIX F
Counts of all the genes identified in gene ontology analysis of all contigs

APPENDIX F1. *A. curvipes.*

GO Level	Term (Name)	#Sequence	Parents (Name)	Category
2	metabolic process	2166	biological_process	Biological Process
Null	single-organism process	1313	biological_process	Biological Process
Null	single-organism cellular process	1088	single-organism process, cellular process	Biological Process
3	primary metabolic process	1616	metabolic process	Biological Process
2	cellular process	1880	biological_process	Biological Process
3	regulation of biological process	870	biological regulation, biological_process	Biological Process
3	multicellular organismal development	684	single-multicellular organism process, developmental process	Biological Process
2	developmental process	748	biological_process	Biological Process
3	cellular component organization	755	cellular component organization or biogenesis	Biological Process
3	organic substance metabolic process	1588	metabolic process	Biological Process
4	protein metabolic process	962	primary metabolic process, macromolecule metabolic process	Biological Process
2	response to stimulus	684	biological_process	Biological Process
4	transport	638	establishment of localization	Biological Process
3	catabolic process	519	metabolic process	Biological Process
4	nucleobase-containing compound metabolic process	564	heterocycle metabolic process, primary metabolic process, organic cyclic compound metabolic process, cellular aromatic compound metabolic process, cellular nitrogen compound metabolic process	Biological Process
2	biological regulation	914	biological_process	Biological Process

APPENDIX F1. (cont.)

Null	single-multicellular organism process	685	multicellular organismal process, single-organism process	Biological Process
3	cellular metabolic process	1270	metabolic process, cellular process	Biological Process
3	biosynthetic process	713	metabolic process	Biological Process
4	cell differentiation	427	cellular developmental process	Biological Process
5	cellular protein metabolic process	693	protein metabolic process, cellular macromolecule metabolic process	Biological Process
2	cellular component organization or biogenesis	755	biological_process	Biological Process
5	organelle organization	473	single-organism cellular process, cellular component organization	Biological Process
3	macromolecule metabolic process	1050	organic substance metabolic process	Biological Process
6	translation	376	cellular macromolecule biosynthetic process, cellular protein metabolic process, gene expression	Biological Process
4	anatomical structure morphogenesis	359	anatomical structure development, developmental process	Biological Process
4	signal transduction	336	single organism signaling, cell communication, cellular response to stimulus, regulation of cellular process	Biological Process
3	response to stress	334	response to stimulus	Biological Process
6	cellular protein modification process	331	cellular protein metabolic process, protein modification process	Biological Process
3	establishment of localization	638	biological_process, localization	Biological Process
4	cellular macromolecule metabolic process	770	cellular metabolic process, macromolecule metabolic process	Biological Process
4	heterocycle metabolic process	564	cellular metabolic process	Biological Process
4	cellular nitrogen compound metabolic process	564	cellular metabolic process, nitrogen compound metabolic process	Biological Process

APPENDIX F1. (cont.)

Null	organic cyclic compound metabolic process	564	organic substance metabolic process	Biological Process
4	cellular aromatic compound metabolic process	564	cellular metabolic process	Biological Process
3	anatomical structure development	401	developmental process	Biological Process
3	cell cycle	299	single-organism cellular process	Biological Process
3	cell communication	422	single-organism cellular process	Biological Process
2	multicellular organismal process	685	biological_process	Biological Process
Null	single organism signaling	409	single-organism process, signaling	Biological Process
3	cellular developmental process	427	developmental process, single-organism cellular process	Biological Process
6	cytoskeleton organization	256	organelle organization	Biological Process
2	reproduction	243	biological_process	Biological Process
4	gene expression	399	macromolecule metabolic process	Biological Process
5	cellular macromolecule biosynthetic process	376	cellular macromolecule metabolic process, cellular biosynthetic process, macromolecule biosynthetic process	Biological Process
4	regulation of cellular process	336	cellular process, regulation of biological process	Biological Process
3	cellular response to stimulus	336	response to stimulus, single-organism cellular process	Biological Process
5	protein modification process	331	macromolecule modification, protein metabolic process	Biological Process
2	localization	639	biological_process	Biological Process
3	nitrogen compound metabolic process	564	metabolic process	Biological Process
4	carbohydrate metabolic process	185	primary metabolic process, organic substance metabolic process	Biological Process

APPENDIX F1. (cont.)

4	generation of precursor metabolites and energy	184	cellular metabolic process	Biological Process
6	protein transport	184	organic substance transport, establishment of protein localization	Biological Process
2	signaling	409	biological_process	Biological Process
5	ion transport	145	single-organism transport	Biological Process
4	embryo development	140	single-organism developmental process, multicellular organismal development, anatomical structure development	Biological Process
3	cell death	136	death, single-organism cellular process	Biological Process
4	cellular biosynthetic process	376	cellular metabolic process, biosynthetic process	Biological Process
4	macromolecule biosynthetic process	376	organic substance biosynthetic process, macromolecule metabolic process	Biological Process
4	macromolecule modification	331	macromolecule metabolic process	Biological Process
4	lipid metabolic process	115	primary metabolic process, organic substance metabolic process, single-organism metabolic process	Biological Process
5	organic substance transport	184	Transport	Biological Process
5	establishment of protein localization	184	establishment of localization, protein localization	Biological Process
6	DNA metabolic process	108	nucleic acid metabolic process, cellular macromolecule metabolic process	Biological Process
3	cell-cell signaling	105	single organism signaling, cell communication	Biological Process
3	behavior	105	response to stimulus	Biological Process
3	response to external stimulus	103	response to stimulus	Biological Process
3	cellular homeostasis	95	homeostatic process, single-organism cellular process	Biological Process
2	growth	97	biological_process	Biological Process
Null	single-organism metabolic process	148	metabolic process	Biological Process

APPENDIX F1. (cont.)

2	cell proliferation	89	single-organism process	Biological Process
Null	single-organism transport	145	single-organism process, transport	Biological Process
Null	single-organism developmental process	140	single-organism process, developmental process	Biological Process
2	death	136	single-organism process	Biological Process
Null	organic substance biosynthetic process	376	organic substance metabolic process, biosynthetic process	Biological Process
3	response to abiotic stimulus	79	response to stimulus	Biological Process
3	response to biotic stimulus	77	response to stimulus	Biological Process
4	protein localization	185	macromolecule localization	Biological Process
5	nucleic acid metabolic process	108	macromolecule metabolic process, nucleobase-containing compound metabolic process	Biological Process
4	homeostatic process	95	regulation of biological quality	Biological Process
6	mitochondrion organization	53	organelle organization	Biological Process
3	response to endogenous stimulus	43	response to stimulus	Biological Process
3	macromolecule localization	185	Localization	Biological Process
3	secondary metabolic process	38	single-organism metabolic process	Biological Process
3	regulation of biological quality	95	biological regulation	Biological Process
2	viral reproduction	29	multi-organism cellular process	Biological Process
3	cell recognition	25	single-multicellular organism process, single-organism cellular process	Biological Process
7	regulation of gene expression, epigenetic	25	regulation of gene expression	Biological Process
5	cytoplasm organization	22	single-organism cellular process, cellular component organization	Biological Process

APPENDIX F1. (cont.)

3	cell growth	19	growth, single-organism cellular process	Biological Process
Null	multi-organism cellular process	29	multi-organism process, cellular process	Biological Process
2	multi-organism process	32	biological_process	Biological Process
4	symbiosis, encompassing mutualism through parasitism	16	interspecies interaction between organisms	Biological Process
6	regulation of gene expression	25	gene expression, regulation of macromolecule metabolic process	Biological Process
3	interspecies interaction between organisms	16	multi-organism process	Biological Process
5	regulation of macromolecule metabolic process	25	macromolecule metabolic process, regulation of metabolic process	Biological Process
4	regulation of metabolic process	25	regulation of biological process, metabolic process	Biological Process
7	glycine receptor clustering	1	postsynaptic membrane organization, neurotransmitter-gated ion channel clustering	Biological Process
8	neurotransmitter-gated ion channel clustering	1	receptor clustering, synapse assembly	Biological Process
4	postsynaptic membrane organization	1	cellular membrane organization	Biological Process
6	synapse assembly	1	nervous system development, synapse organization, cellular component assembly	Biological Process
3	cellular membrane organization	1	single-organism cellular process, membrane organization	Biological Process
7	receptor clustering	1	protein localization to membrane	Biological Process
4	protein localization to membrane	1	cellular membrane organization, cellular protein localization	Biological Process
4	cellular component assembly	1	cellular component biogenesis, cellular component organization	Biological Process

APPENDIX F1. (cont.)

4	membrane organization	1	cellular component organization	Biological Process
5	nervous system development	1	system development	Biological Process
4	synapse organization	1	single-organism cellular process, cellular component organization	Biological Process
4	system development	1	multicellular organismal development, anatomical structure development	Biological Process
5	cellular protein localization	1	cellular macromolecule localization, protein localization	Biological Process
3	cellular component biogenesis	1	cellular component organization or biogenesis	Biological Process
4	cellular macromolecule localization	1	macromolecule localization, cellular localization	Biological Process
3	cellular localization	1	localization, single-organism cellular process	Biological Process
2	cell	2494	cellular_component	Cellular Component
4	intracellular part	1949	cell part, intracellular	Cellular Component
5	cytoplasm	1502	intracellular part	Cellular Component
4	intracellular	2151	cell part	Cellular Component
3	cell part	2208	cell, cellular_component	Cellular Component
5	cytoplasmic part	1071	intracellular part, cytoplasm	Cellular Component
6	intracellular membrane-bounded organelle	1287	intracellular organelle, membrane-bounded organelle	Cellular Component
5	intracellular organelle	1639	organelle, intracellular part	Cellular Component
3	protein complex	733	macromolecular complex	Cellular Component

APPENDIX F1. (cont.)

7	nucleus	656	intracellular membrane-bounded organelle	Cellular Component
2	organelle	1692	cellular_component	Cellular Component
2	macromolecular complex	966	cellular_component	Cellular Component
3	membrane-bounded organelle	1287	Organelle	Cellular Component
6	intracellular non-membrane-bounded organelle	687	intracellular organelle, non-membrane-bounded organelle	Cellular Component
6	mitochondrion	424	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
7	cytoskeleton	312	intracellular non-membrane-bounded organelle	Cellular Component
6	cytosol	273	cytoplasmic part	Cellular Component
3	non-membrane-bounded organelle	687	Organelle	Cellular Component
6	ribosome	245	ribonucleoprotein complex, cytoplasmic part, intracellular non-membrane-bounded organelle	Cellular Component
6	lipid particle	199	cytoplasmic part	Cellular Component
6	nuclear part	278	intracellular organelle part, nucleus	Cellular Component
4	plasma membrane	187	cell part, cell periphery, membrane	Cellular Component
7	nuclear lumen	236	nuclear part, intracellular organelle lumen	Cellular Component
6	nucleoplasm	147	nuclear part, nuclear lumen	Cellular Component

APPENDIX F1. (cont.)

5	ribonucleoprotein complex	245	macromolecular complex, intracellular part	Cellular Component
4	intracellular organelle part	321	organelle part, intracellular organelle, intracellular part	Cellular Component
6	endoplasmic reticulum	132	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
2	extracellular region	186	cellular_component	Cellular Component
4	cell periphery	188	cell part	Cellular Component
2	membrane	187	cellular_component	Cellular Component
6	nucleolus	106	nuclear part, nuclear lumen, intracellular non-membrane-bounded organelle	Cellular Component
6	Golgi apparatus	104	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	intracellular organelle lumen	236	intracellular organelle part, organelle lumen	Cellular Component
7	chromosome	95	intracellular non-membrane-bounded organelle	Cellular Component
3	organelle part	321	cellular_component, organelle	Cellular Component
4	extracellular space	78	extracellular region part	Cellular Component
7	cytoplasmic membrane-bounded vesicle	67	cytoplasmic vesicle, intracellular membrane-bounded organelle, membrane-bounded vesicle	Cellular Component
4	organelle lumen	236	organelle part, membrane-enclosed lumen	Cellular Component
7	microtubule organizing center	57	microtubule cytoskeleton, cytoskeletal part	Cellular Component
3	extracellular region part	90	cellular_component, extracellular region	Cellular Component

APPENDIX F1. (cont.)

5	nuclear envelope	48	nuclear part, organelle envelope, endomembrane system	Cellular Component
4	membrane-bounded vesicle	67	Vesicle	Cellular Component
6	cytoplasmic vesicle	67	cytoplasmic part, vesicle, intracellular organelle	Cellular Component
2	membrane-enclosed lumen	236	cellular_component	Cellular Component
6	cytoskeletal part	57	intracellular organelle part, cytoskeleton	Cellular Component
8	microtubule cytoskeleton	57	Cytoskeleton	Cellular Component
6	nuclear chromosome	30	nuclear part, nuclear lumen, chromosome	Cellular Component
4	organelle envelope	48	intracellular organelle part, membrane-bounded organelle, envelope	Cellular Component
4	endomembrane system	48	cell part	Cellular Component
6	vacuole	36	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
3	vesicle	67	Organelle	Cellular Component
6	endosome	24	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
7	peroxisome	21	Microbody	Cellular Component
4	envelope	49	cell part	Cellular Component
4	proteinaceous extracellular matrix	15	extracellular region part, extracellular matrix	Cellular Component
8	lysosome	13	lytic vacuole	Cellular Component

APPENDIX F1. (cont.)

6	microbody	21	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
2	extracellular matrix	15	cellular_component	Cellular Component
5	cilium	8	cell projection, intracellular membrane-bounded organelle	Cellular Component
7	lytic vacuole	13	Vacuole	Cellular Component
4	cell projection	8	cell part	Cellular Component
6	plastid	4	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
4	external encapsulating structure	3	cell part, cell periphery	Cellular Component
5	cell envelope	1	Envelope	Cellular Component
5	cell wall	1	external encapsulating structure	Cellular Component
5	thylakoid	1	intracellular part	Cellular Component
2	binding	2108	molecular_function	Molecular Function
2	catalytic activity	2218	molecular_function	Molecular Function
4	nucleotide binding	806	small molecule binding, nucleoside phosphate binding	Molecular Function
3	hydrolase activity	925	catalytic activity	Molecular Function
3	protein binding	668	Binding	Molecular Function
3	organic cyclic compound binding	1169	Binding	Molecular Function

APPENDIX F1. (cont.)

null	heterocyclic compound binding	1164	Binding	Molecular Function
3	small molecule binding	806	Binding	Molecular Function
null	nucleoside phosphate binding	806	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
3	transferase activity	571	catalytic activity	Molecular Function
3	nucleic acid binding	548	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
2	transporter activity	343	molecular_function	Molecular Function
2	structural molecule activity	304	molecular_function	Molecular Function
4	peptidase activity	226	hydrolase activity	Molecular Function
4	RNA binding	251	nucleic acid binding	Molecular Function
4	DNA binding	179	nucleic acid binding	Molecular Function
5	kinase activity	186	transferase activity, transferring phosphorus-containing groups	Molecular Function
6	protein kinase activity	101	kinase activity, phosphotransferase activity, alcohol group as acceptor	Molecular Function
2	enzyme regulator activity	95	molecular_function	Molecular Function
4	transferase activity, transferring phosphorus-containing groups	186	transferase activity	Molecular Function
4	cytoskeletal protein binding	94	protein binding	Molecular Function

APPENDIX F1. (cont.)

5	translation factor activity, nucleic acid binding	82	RNA binding	Molecular Function
6	calcium ion binding	72	metal ion binding	Molecular Function
2	electron carrier activity	63	molecular_function	Molecular Function
5	actin binding	62	cytoskeletal protein binding	Molecular Function
5	phosphotransferase activity, alcohol group as acceptor	101	transferase activity, transferring phosphorus-containing groups	Molecular Function
3	lipid binding	53	Binding	Molecular Function
4	receptor binding	47	protein binding	Molecular Function
2	receptor activity	45	molecular_function	Molecular Function
4	hydrolase activity, acting on ester bonds	99	hydrolase activity	Molecular Function
5	metal ion binding	72	cation binding	Molecular Function
5	nuclease activity	41	hydrolase activity, acting on ester bonds	Molecular Function
3	sequence-specific DNA binding transcription factor activity	40	nucleic acid binding transcription factor activity	Molecular Function
7	phosphoprotein phosphatase activity	40	phosphatase activity	Molecular Function
3	carbohydrate binding	39	Binding	Molecular Function
Null	transcription regulator activity	37	obsolete_molecular_function	Molecular Function

APPENDIX F1. (cont.)

6	ion channel activity	28	substrate-specific channel activity, ion transmembrane transporter activity	Molecular Function
3	signal transducer activity	27	molecular transducer activity	Molecular Function
4	cation binding	72	ion binding	Molecular Function
8	motor activity	25	nucleoside-triphosphatase activity	Molecular Function
6	phosphatase activity	41	phosphoric ester hydrolase activity	Molecular Function
2	nucleic acid binding transcription factor activity	40	molecular_function	Molecular Function
Null	obsolete_molecular_function	37		Molecular Function
2	nutrient reservoir activity	21	molecular_function	Molecular Function
3	chromatin binding	21	Binding	Molecular Function
5	carboxylic ester hydrolase activity	17	hydrolase activity, acting on ester bonds	Molecular Function
5	ion transmembrane transporter activity	28	substrate-specific transmembrane transporter activity	Molecular Function
5	substrate-specific channel activity	28	substrate-specific transmembrane transporter activity, channel activity	Molecular Function
2	molecular transducer activity	27	molecular_function	Molecular Function
3	ion binding	72	Binding	Molecular Function
7	nucleoside-triphosphatase activity	25	pyrophosphatase activity	Molecular Function
5	phosphoric ester hydrolase activity	41	hydrolase activity, acting on ester bonds	Molecular Function

APPENDIX F1. (cont.)

2	antioxidant activity	13	molecular_function	Molecular Function
4	substrate-specific transmembrane transporter activity	28	substrate-specific transporter activity, transmembrane transporter activity	Molecular Function
5	channel activity	28	passive transmembrane transporter activity	Molecular Function
6	pyrophosphatase activity	25	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Molecular Function
3	transmembrane transporter activity	28	transporter activity	Molecular Function
3	substrate-specific transporter activity	28	transporter activity	Molecular Function
4	passive transmembrane transporter activity	28	transmembrane transporter activity	Molecular Function
2	translation regulator activity	6	molecular_function	Molecular Function
5	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	25	hydrolase activity, acting on acid anhydrides	Molecular Function
4	hydrolase activity, acting on acid anhydrides	25	hydrolase activity	Molecular Function
8	inositol monophosphate phosphatase activity	1	inositol phosphate phosphatase activity	Molecular Function
3	oxygen binding	1	Binding	Molecular Function
7	inositol phosphate phosphatase activity	1	phosphatase activity	Molecular Function

APPENDIX F2. ***A. craccivora.***

GO Level	Term (Name)	#Sequence	Parents (Name)	Category
3	regulation of biological process	946	biological regulation, biological_process	Biological Process
2	metabolic process	2170	biological_process	Biological Process
3	catabolic process	565	metabolic process	Biological Process
3	multicellular organismal development	566	single-multicellular organism process, developmental process	Biological Process
4	nucleobase-containing compound metabolic process	614	heterocycle metabolic process, primary metabolic process, organic cyclic compound metabolic process, cellular aromatic compound metabolic process, cellular nitrogen compound metabolic process	Biological Process
3	biosynthetic process	806	metabolic process	Biological Process
4	transport	651	establishment of localization	Biological Process
4	signal transduction	408	single organism signaling, cell communication, cellular response to stimulus, regulation of cellular process	Biological Process
4	cell differentiation	401	cellular developmental process	Biological Process
3	cellular component organization	654	cellular component organization or biogenesis	Biological Process
6	cellular protein modification process	358	cellular protein metabolic process, protein modification process	Biological Process
6	translation	328	cellular macromolecule biosynthetic process, cellular protein metabolic process, gene expression	Biological Process
4	protein metabolic process	906	primary metabolic process, macromolecule metabolic process	Biological Process
4	anatomical structure morphogenesis	316	anatomical structure development, developmental process	Biological Process

APPENDIX F2. (cont.)

3	response to stress	287	response to stimulus	Biological Process
5	organelle organization	413	single-organism cellular process, cellular component organization	Biological Process
3	cell cycle	245	single-organism cellular process	Biological Process
2	reproduction	219	biological_process	Biological Process
4	carbohydrate metabolic process	203	primary metabolic process, organic substance metabolic process	Biological Process
4	generation of precursor metabolites and energy	199	cellular metabolic process	Biological Process
6	cytoskeleton organization	186	organelle organization	Biological Process
5	ion transport	185	single-organism transport	Biological Process
6	protein transport	185	organic substance transport, establishment of protein localization	Biological Process
4	lipid metabolic process	173	primary metabolic process, organic substance metabolic process, single-organism metabolic process	Biological Process
4	embryo development	136	single-organism developmental process, multicellular organismal development, anatomical structure development	Biological Process
3	cell-cell signaling	128	single organism signaling, cell communication	Biological Process
6	DNA metabolic process	117	nucleic acid metabolic process, cellular macromolecule metabolic process	Biological Process
3	cell death	112	death, single-organism cellular process	Biological Process
3	behavior	111	response to stimulus	Biological Process
3	response to external stimulus	96	response to stimulus	Biological Process
2	cell proliferation	93	single-organism process	Biological Process
2	growth	100	biological_process	Biological Process
3	cellular homeostasis	71	homeostatic process, single-organism cellular process	Biological Process
3	response to abiotic stimulus	68	response to stimulus	Biological Process

APPENDIX F2. (cont.)

3	secondary metabolic process	50	single-organism metabolic process	Biological Process
3	response to endogenous stimulus	45	response to stimulus	Biological Process
2	viral reproduction	41	multi-organism cellular process	Biological Process
6	mitochondrion organization	39	organelle organization	Biological Process
3	response to biotic stimulus	35	response to stimulus	Biological Process
3	cell communication	495	single-organism cellular process	Biological Process
3	cell recognition	24	single-multicellular organism process, single-organism cellular process	Biological Process
7	regulation of gene expression, epigenetic	24	regulation of gene expression	Biological Process
3	primary metabolic process	1652	metabolic process	Biological Process
4	symbiosis, encompassing mutualism through parasitism	21	interspecies interaction between organisms	Biological Process
3	cell growth	19	growth, single-organism cellular process	Biological Process
5	cytoplasm organization	17	single-organism cellular process, cellular component organization	Biological Process
5	intracellular signal transduction	1	signal transduction	Biological Process
5	cellular response to drug	1	cellular response to chemical stimulus, response to drug	Biological Process
2	cellular process	1881	biological_process	Biological Process
4	cellular response to chemical stimulus	1	response to chemical stimulus, cellular response to stimulus	Biological Process
3	regulation of biological quality	71	biological regulation	Biological Process
2	localization	651	biological_process	Biological Process

APPENDIX F2. (cont.)

2	response to stimulus	662	biological_process	Biological Process
3	organic substance metabolic process	1638	metabolic process	Biological Process
3	cellular developmental process	401	developmental process, single-organism cellular process	Biological Process
4	cellular biosynthetic process	328	cellular metabolic process, biosynthetic process	Biological Process
Null	organic substance biosynthetic process	328	organic substance metabolic process, biosynthetic process	Biological Process
Null	single-organism cellular process	1023	single-organism process, cellular process	Biological Process
5	cellular protein metabolic process	671	protein metabolic process, cellular macromolecule metabolic process	Biological Process
Null	single-organism metabolic process	210	metabolic process	Biological Process
3	anatomical structure development	352	developmental process	Biological Process
Null	multi-organism cellular process	41	multi-organism process, cellular process	Biological Process
5	cellular macromolecule biosynthetic process	328	cellular macromolecule metabolic process, cellular biosynthetic process, macromolecule biosynthetic process	Biological Process
4	cellular macromolecule metabolic process	760	cellular metabolic process, macromolecule metabolic process	Biological Process
4	regulation of metabolic process	24	regulation of biological process, metabolic process	Biological Process
2	cellular component organization or biogenesis	654	biological_process	Biological Process
2	multi-organism process	46	biological_process	Biological Process
6	regulation of gene expression	24	gene expression, regulation of macromolecule metabolic process	Biological Process
4	heterocycle metabolic process	614	cellular metabolic process	Biological Process
5	organic substance transport	185	Transport	Biological Process
5	establishment of protein localization	185	establishment of localization, protein localization	Biological Process

APPENDIX F2. (cont.)

2	death	112	single-organism process	Biological Process
Null	single organism signaling	480	single-organism process, signaling	Biological Process
5	nucleic acid metabolic process	117	macromolecule metabolic process, nucleobase-containing compound metabolic process	Biological Process
5	regulation of macromolecule metabolic process	24	macromolecule metabolic process, regulation of metabolic process	Biological Process
Null	single-organism process	1256	biological_process	Biological Process
3	macromolecule metabolic process	1002	organic substance metabolic process	Biological Process
4	homeostatic process	71	regulation of biological quality	Biological Process
4	cellular nitrogen compound metabolic process	614	cellular metabolic process, nitrogen compound metabolic process	Biological Process
2	signaling	480	biological_process	Biological Process
Null	single-organism transport	185	single-organism process, transport	Biological Process
Null	single-organism developmental process	136	single-organism process, developmental process	Biological Process
3	response to chemical stimulus	1	response to stimulus	Biological Process
Null	organic cyclic compound metabolic process	614	organic substance metabolic process	Biological Process
2	multicellular organismal process	566	biological_process	Biological Process
4	macromolecule biosynthetic process	328	organic substance biosynthetic process, macromolecule metabolic process	Biological Process
2	biological regulation	970	biological_process	Biological Process
3	macromolecule localization	185	Localization	Biological Process
3	nitrogen compound metabolic process	614	metabolic process	Biological Process
3	establishment of localization	651	biological_process, localization	Biological Process
4	regulation of cellular process	408	cellular process, regulation of biological process	Biological Process

APPENDIX F2. (cont.)

4	cellular aromatic compound metabolic process	614	cellular metabolic process	Biological Process
4	response to drug	1	response to chemical stimulus	Biological Process
3	interspecies interaction between organisms	21	multi-organism process	Biological Process
4	gene expression	351	macromolecule metabolic process	Biological Process
Null	all	0		Biological Process
4	macromolecule modification	358	macromolecule metabolic process	Biological Process
3	cellular metabolic process	1300	metabolic process, cellular process	Biological Process
3	cellular response to stimulus	409	response to stimulus, single-organism cellular process	Biological Process
4	protein localization	185	macromolecule localization	Biological Process
2	developmental process	636	biological_process	Biological Process
Null	single-multicellular organism process	566	multicellular organismal process, single-organism process	Biological Process
5	protein modification process	358	macromolecule modification, protein metabolic process	Biological Process
3	protein complex	790	macromolecular complex	Cellular Component
2	cell	2405	cellular_component	Cellular Component
5	cytoplasm	1340	intracellular part	Cellular Component
7	nucleus	690	intracellular membrane-bounded organelle	Cellular Component
4	intracellular	2046	cell part	Cellular Component
6	mitochondrion	314	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	ribosome	275	ribonucleoprotein complex, cytoplasmic part, intracellular non-membrane-bounded organelle	Cellular Component
6	nucleoplasm	244	nuclear part, nuclear lumen	Cellular Component
7	cytoskeleton	244	intracellular non-membrane-bounded organelle	Cellular Component
4	plasma membrane	207	cell part, cell periphery, membrane	Cellular Component

APPENDIX F2. (cont.)

6	cytosol	190	cytoplasmic part	Cellular Component
6	nucleolus	129	nuclear part, nuclear lumen, intracellular non-membrane-bounded organelle	Cellular Component
7	chromosome	153	intracellular non-membrane-bounded organelle	Cellular Component
6	endoplasmic reticulum	122	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
2	organelle	1602	cellular_component	Cellular Component
6	lipid particle	87	cytoplasmic part	Cellular Component
6	Golgi apparatus	86	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
2	extracellular region	112	cellular_component	Cellular Component
7	cytoplasmic membrane-bounded vesicle	53	cytoplasmic vesicle, intracellular membrane-bounded organelle, membrane-bounded vesicle	Cellular Component
6	nuclear chromosome	50	nuclear part, nuclear lumen, chromosome	Cellular Component
5	nuclear envelope	48	nuclear part, organelle envelope, endomembrane system	Cellular Component
6	endosome	38	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
7	microtubule organizing center	38	microtubule cytoskeleton, cytoskeletal part	Cellular Component
4	extracellular space	30	extracellular region part	Cellular Component
6	vacuole	48	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
8	lysosome	21	lytic vacuole	Cellular Component
7	peroxisome	19	Microbody	Cellular Component
4	proteinaceous extracellular matrix	19	extracellular region part, extracellular matrix	Cellular Component
6	plastid	13	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
5	cell wall	7	external encapsulating structure	Cellular Component

APPENDIX F2. (cont.)

5	cilium	6	cell projection, intracellular membrane-bounded organelle	Cellular Component
4	external encapsulating structure	11	cell part, cell periphery	Cellular Component
5	cell envelope	1	Envelope	Cellular Component
4	organelle envelope	48	intracellular organelle part, membrane-bounded organelle, envelope	Cellular Component
3	extracellular region part	47	cellular_component, extracellular region	Cellular Component
7	lytic vacuole	21	Vacuole	Cellular Component
4	endomembrane system	48	cell part	Cellular Component
3	membrane-bounded organelle	1184	Organelle	Cellular Component
3	vesicle	53	Organelle	Cellular Component
4	envelope	49	cell part	Cellular Component
6	intracellular organelle lumen	334	intracellular organelle part, organelle lumen	Cellular Component
Null	all	0		Cellular Component
2	extracellular matrix	19	cellular_component	Cellular Component
6	intracellular non-membrane-bounded organelle	713	intracellular organelle, non-membrane-bounded organelle	Cellular Component
4	organelle lumen	334	organelle part, membrane-enclosed lumen	Cellular Component
6	nuclear part	369	intracellular organelle part, nucleus	Cellular Component
3	non-membrane-bounded organelle	713	Organelle	Cellular Component
6	intracellular membrane-bounded organelle	1184	intracellular organelle, membrane-bounded organelle	Cellular Component
4	membrane-bounded vesicle	53	Vesicle	Cellular Component
5	cytoplasmic part	910	intracellular part, cytoplasm	Cellular Component
3	cell part	2114	cell, cellular_component	Cellular Component
6	cytoskeletal part	38	intracellular organelle part, cytoskeleton	Cellular Component
5	intracellular organelle	1531	organelle, intracellular part	Cellular Component

APPENDIX F2. (cont.)

2	membrane-enclosed lumen	334	cellular_component	Cellular Component
5	ribonucleoprotein complex	275	macromolecular complex, intracellular part	Cellular Component
2	membrane	207	cellular_component	Cellular Component
6	microbody	19	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	cytoplasmic vesicle	53	cytoplasmic part, vesicle, intracellular organelle	Cellular Component
2	macromolecular complex	1024	cellular_component	Cellular Component
4	intracellular organelle part	393	organelle part, intracellular organelle, intracellular part	Cellular Component
7	nuclear lumen	334	nuclear part, intracellular organelle lumen	Cellular Component
8	microtubule cytoskeleton	38	Cytoskeleton	Cellular Component
4	cell projection	6	cell part	Cellular Component
4	cell periphery	215	cell part	Cellular Component
3	organelle part	393	cellular_component, organelle	Cellular Component
4	intracellular part	1827	cell part, intracellular	Cellular Component
4	nucleotide binding	828	small molecule binding, nucleoside phosphate binding	Molecular Function
2	catalytic activity	2012	molecular_function	Molecular Function
2	binding	2097	molecular_function	Molecular Function
3	hydrolase activity	867	catalytic activity	Molecular Function
3	protein binding	657	Binding	Molecular Function
3	transferase activity	592	catalytic activity	Molecular Function
2	structural molecule activity	259	molecular_function	Molecular Function
2	transporter activity	272	molecular_function	Molecular Function
4	DNA binding	192	nucleic acid binding	Molecular Function
4	peptidase activity	189	hydrolase activity	Molecular Function
4	RNA binding	276	nucleic acid binding	Molecular Function

APPENDIX F2. (cont.)

3	nucleic acid binding	592	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
6	protein kinase activity	122	kinase activity, phosphotransferase activity, alcohol group as acceptor	Molecular Function
2	enzyme regulator activity	101	molecular_function	Molecular Function
5	translation factor activity, nucleic acid binding	89	RNA binding	Molecular Function
6	calcium ion binding	87	metal ion binding	Molecular Function
5	kinase activity	208	transferase activity, transferring phosphorus-containing groups	Molecular Function
3	lipid binding	65	Binding	Molecular Function
5	actin binding	61	cytoskeletal protein binding	Molecular Function
2	receptor activity	53	molecular_function	Molecular Function
3	sequence-specific DNA binding transcription factor activity	53	nucleic acid binding transcription factor activity	Molecular Function
7	phosphoprotein phosphatase activity	51	phosphatase activity	Molecular Function
3	carbohydrate binding	50	Binding	Molecular Function
2	electron carrier activity	48	molecular_function	Molecular Function
4	receptor binding	42	protein binding	Molecular Function
4	cytoskeletal protein binding	95	protein binding	Molecular Function
5	nuclease activity	38	hydrolase activity, acting on ester bonds	Molecular Function
3	chromatin binding	37	Binding	Molecular Function
3	signal transducer activity	30	molecular transducer activity	Molecular Function
8	motor activity	29	nucleoside-triphosphatase activity	Molecular Function
Null	transcription regulator activity	28	obsolete_molecular_function	Molecular Function

APPENDIX F2. (cont.)

2	antioxidant activity	19	molecular_function	Molecular Function
6	ion channel activity	16	substrate-specific channel activity, ion transmembrane transporter activity	Molecular Function
3	organic cyclic compound binding	1213	Binding	Molecular Function
5	carboxylic ester hydrolase activity	9	hydrolase activity, acting on ester bonds	Molecular Function
2	translation regulator activity	2	molecular_function	Molecular Function
4	neurotransmitter transporter activity	2	transporter activity	Molecular Function
6	lead ion binding	1	metal ion binding	Molecular Function
3	small molecule binding	829	Binding	Molecular Function
3	oxygen binding	1	Binding	Molecular Function
4	substrate-specific transmembrane transporter activity	16	substrate-specific transporter activity, transmembrane transporter activity	Molecular Function
5	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	29	hydrolase activity, acting on acid anhydrides	Molecular Function
2	molecular transducer activity	30	molecular_function	Molecular Function
5	ion transmembrane transporter activity	16	substrate-specific transmembrane transporter activity	Molecular Function
5	metal ion binding	88	cation binding	Molecular Function
5	channel activity	16	passive transmembrane transporter activity	Molecular Function
5	phosphoric ester hydrolase activity	51	hydrolase activity, acting on ester bonds	Molecular Function
4	cation binding	88	ion binding	Molecular Function

APPENDIX F2. (cont.)

4	hydrolase activity, acting on acid anhydrides	29	hydrolase activity	Molecular Function
4	hydrolase activity, acting on ester bonds	98	hydrolase activity	Molecular Function
3	transmembrane transporter activity	16	transporter activity	Molecular Function
6	phosphatase activity	51	phosphoric ester hydrolase activity	Molecular Function
5	phosphotransferase activity, alcohol group as acceptor	122	transferase activity, transferring phosphorus-containing groups	Molecular Function
2	nucleic acid binding transcription factor activity	53	molecular_function	Molecular Function
Null	all	0		Molecular Function
3	substrate-specific transporter activity	16	transporter activity	Molecular Function
3	ion binding	88	Binding	Molecular Function
4	passive transmembrane transporter activity	16	transmembrane transporter activity	Molecular Function
Null	obsolete_molecular_function	28		Molecular Function
Null	nucleoside phosphate binding	828	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
6	pyrophosphatase activity	29	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Molecular Function
Null	heterocyclic compound binding	1203	Binding	Molecular Function
7	nucleoside-triphosphatase activity	29	pyrophosphatase activity	Molecular Function
5	substrate-specific channel activity	16	substrate-specific transmembrane transporter activity, channel activity	Molecular Function

APPENDIX F2. (cont.)

4	transferase activity, transferring phosphorus-containing groups	208	transferase activity	Molecular Function

APPENDIX F3. ***C. tomentosicollis.***

GO Level	Term (Name)	#Sequence	Parents (Name)	Category
2	metabolic process	2379	biological_process	Biological Process
3	regulation of biological process	850	biological regulation, biological_process	Biological Process
3	multicellular organismal development	680	single-multicellular organism process, developmental process	Biological Process
3	catabolic process	619	metabolic process	Biological Process
4	nucleobase-containing compound metabolic process	572	heterocycle metabolic process, primary metabolic process, organic cyclic compound metabolic process, cellular aromatic compound metabolic process, cellular nitrogen compound metabolic process	Biological Process
3	cellular component organization	708	cellular component organization or biogenesis	Biological Process
4	cell differentiation	413	cellular developmental process	Biological Process
4	transport	620	establishment of localization	Biological Process
3	biosynthetic process	759	metabolic process	Biological Process
4	protein metabolic process	998	primary metabolic process, macromolecule metabolic process	Biological Process
6	translation	371	cellular macromolecule biosynthetic process, cellular protein metabolic process, gene expression	Biological Process
4	signal transduction	360	single organism signaling, cell communication, cellular response to stimulus, regulation of cellular process	Biological Process
4	anatomical structure morphogenesis	347	anatomical structure development, developmental process	Biological Process

APPENDIX F3. (cont.)

6	cellular protein modification process	344	cellular protein metabolic process, protein modification process	Biological Process
3	response to stress	316	response to stimulus	Biological Process
3	cell cycle	249	single-organism cellular process	Biological Process
4	generation of precursor metabolites and energy	235	cellular metabolic process	Biological Process
5	organelle organization	420	single-organism cellular process, cellular component organization	Biological Process
2	reproduction	217	biological_process	Biological Process
4	carbohydrate metabolic process	216	primary metabolic process, organic substance metabolic process	Biological Process
6	cytoskeleton organization	206	organelle organization	Biological Process
6	protein transport	199	organic substance transport, establishment of protein localization	Biological Process
4	embryo development	163	single-organism developmental process, multicellular organismal development, anatomical structure development	Biological Process
3	cell death	159	death, single-organism cellular process	Biological Process
4	lipid metabolic process	157	primary metabolic process, organic substance metabolic process, single-organism metabolic process	Biological Process
5	ion transport	133	single-organism transport	Biological Process
3	response to external stimulus	111	response to stimulus	Biological Process
3	cell-cell signaling	110	single organism signaling, cell communication	Biological Process
3	cellular homeostasis	110	homeostatic process, single-organism cellular process	Biological Process

APPENDIX F3. (cont.)

3	behavior	100	response to stimulus	Biological Process
2	growth	104	biological_process	Biological Process
2	cell proliferation	89	single-organism process	Biological Process
3	response to biotic stimulus	74	response to stimulus	Biological Process
6	DNA metabolic process	72	nucleic acid metabolic process, cellular macromolecule metabolic process	Biological Process
3	response to abiotic stimulus	68	response to stimulus	Biological Process
6	mitochondrion organization	55	organelle organization	Biological Process
3	response to endogenous stimulus	43	response to stimulus	Biological Process
3	secondary metabolic process	41	single-organism metabolic process	Biological Process
2	viral reproduction	39	multi-organism cellular process	Biological Process
3	cell recognition	30	single-multicellular organism process, single-organism cellular process	Biological Process
3	cell communication	450	single-organism cellular process	Biological Process
3	primary metabolic process	1712	metabolic process	Biological Process
4	symbiosis, encompassing mutualism through parasitism	22	interspecies interaction between organisms	Biological Process
3	cell growth	20	growth, single-organism cellular process	Biological Process
7	regulation of gene expression, epigenetic	19	regulation of gene expression	Biological Process
5	cytoplasm organization	18	single-organism cellular process, cellular component organization	Biological Process
5	intracellular signal transduction	3	signal transduction	Biological Process
2	death	159	single-organism process	Biological Process
2	cellular process	1936	biological_process	Biological Process

APPENDIX F3. (cont.)

3	regulation of biological quality	110	biological regulation	Biological Process
2	localization	620	biological_process	Biological Process
2	response to stimulus	648	biological_process	Biological Process
3	organic substance metabolic process	1697	metabolic process	Biological Process
3	cellular developmental process	413	developmental process, single-organism cellular process	Biological Process
4	cellular biosynthetic process	371	cellular metabolic process, biosynthetic process	Biological Process
Null	organic substance biosynthetic process	371	organic substance metabolic process, biosynthetic process	Biological Process
Null	single-organism cellular process	1038	single-organism process, cellular process	Biological Process
5	cellular protein metabolic process	698	protein metabolic process, cellular macromolecule metabolic process	Biological Process
Null	single-organism metabolic process	186	metabolic process	Biological Process
3	anatomical structure development	398	developmental process	Biological Process
Null	multi-organism cellular process	39	multi-organism process, cellular process	Biological Process
5	cellular macromolecule biosynthetic process	371	cellular macromolecule metabolic process, cellular biosynthetic process, macromolecule biosynthetic process	Biological Process
4	cellular macromolecule metabolic process	744	cellular metabolic process, macromolecule metabolic process	Biological Process
4	regulation of metabolic process	19	regulation of biological process, metabolic process	Biological Process

APPENDIX F3. (cont.)

2	cellular component organization or biogenesis	708	biological_process	Biological Process
2	multi-organism process	44	biological_process	Biological Process
6	regulation of gene expression	19	gene expression, regulation of macromolecule metabolic process	Biological Process
4	heterocycle metabolic process	572	cellular metabolic process	Biological Process
5	organic substance transport	199	Transport	Biological Process
5	establishment of protein localization	199	establishment of localization, protein localization	Biological Process
Null	single organism signaling	431	single-organism process, signaling	Biological Process
5	nucleic acid metabolic process	72	macromolecule metabolic process, nucleobase-containing compound metabolic process	Biological Process
5	regulation of macromolecule metabolic process	19	macromolecule metabolic process, regulation of metabolic process	Biological Process
Null	single-organism process	1285	biological_process	Biological Process
3	macromolecule metabolic process	1054	organic substance metabolic process	Biological Process
4	homeostatic process	110	regulation of biological quality	Biological Process
4	cellular nitrogen compound metabolic process	572	cellular metabolic process, nitrogen compound metabolic process	Biological Process
2	signaling	431	biological_process	Biological Process
Null	single-organism transport	133	single-organism process, transport	Biological Process
Null	single-organism developmental process	163	single-organism process, developmental process	Biological Process
Null	organic cyclic compound metabolic process	572	organic substance metabolic process	Biological Process

APPENDIX F3. (cont.)

2	multicellular organismal process	680	biological_process	Biological Process
4	macromolecule biosynthetic process	371	organic substance biosynthetic process, macromolecule metabolic process	Biological Process
2	biological regulation	879	biological_process	Biological Process
3	macromolecule localization	199	Localization	Biological Process
3	nitrogen compound metabolic process	572	metabolic process	Biological Process
3	establishment of localization	620	biological_process, localization	Biological Process
4	regulation of cellular process	360	cellular process, regulation of biological process	Biological Process
4	cellular aromatic compound metabolic process	572	cellular metabolic process	Biological Process
3	interspecies interaction between organisms	22	multi-organism process	Biological Process
4	gene expression	390	macromolecule metabolic process	Biological Process
Null	all	0		Biological Process
4	macromolecule modification	344	macromolecule metabolic process	Biological Process
3	cellular metabolic process	1322	metabolic process, cellular process	Biological Process
3	cellular response to stimulus	360	response to stimulus, single-organism cellular process	Biological Process
4	protein localization	199	macromolecule localization	Biological Process
2	developmental process	741	biological_process	Biological Process
Null	single-multicellular organism process	680	multicellular organismal process, single-organism process	Biological Process
5	protein modification process	344	macromolecule modification, protein metabolic process	Biological Process

APPENDIX F3. (cont.)

3	protein complex	765	macromolecular complex	Cellular Component
2	cell	2588	cellular_component	Cellular Component
5	cytoplasm	1564	intracellular part	Cellular Component
6	mitochondrion	465	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
7	nucleus	642	intracellular membrane-bounded organelle	Cellular Component
4	intracellular	2204	cell part	Cellular Component
7	cytoskeleton	317	intracellular non-membrane-bounded organelle	Cellular Component
6	ribosome	242	ribonucleoprotein complex, cytoplasmic part, intracellular non-membrane-bounded organelle	Cellular Component
6	cytosol	232	cytoplasmic part	Cellular Component
4	plasma membrane	219	cell part, cell periphery, membrane	Cellular Component
6	nucleoplasm	159	nuclear part, nuclear lumen	Cellular Component
6	lipid particle	144	cytoplasmic part	Cellular Component
6	endoplasmic reticulum	137	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	nucleolus	103	nuclear part, nuclear lumen, intracellular non-membrane-bounded organelle	Cellular Component
2	organelle	1728	cellular_component	Cellular Component
6	Golgi apparatus	97	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
2	extracellular region	153	cellular_component	Cellular Component
7	chromosome	107	intracellular non-membrane-bounded organelle	Cellular Component

APPENDIX F3. (cont.)

7	cytoplasmic membrane-bounded vesicle	66	cytoplasmic vesicle, intracellular membrane-bounded organelle, membrane-bounded vesicle	Cellular Component
7	microtubule organizing center	56	microtubule cytoskeleton, cytoskeletal part	Cellular Component
4	extracellular space	49	extracellular region part	Cellular Component
6	endosome	35	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
5	nuclear envelope	34	nuclear part, organelle envelope, endomembrane system	Cellular Component
6	nuclear chromosome	33	nuclear part, nuclear lumen, chromosome	Cellular Component
6	vacuole	53	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
8	lysosome	25	lytic vacuole	Cellular Component
7	peroxisome	23	Microbody	Cellular Component
4	proteinaceous extracellular matrix	14	extracellular region part, extracellular matrix	Cellular Component
5	cilium	10	cell projection, intracellular membrane-bounded organelle	Cellular Component
6	plastid	3	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	cytoplasmic chromosome	1	cytoplasmic part, chromosome	Cellular Component
4	organelle envelope	34	intracellular organelle part, membrane-bounded organelle, envelope	Cellular Component
3	extracellular region part	61	cellular_component, extracellular region	Cellular Component
7	lytic vacuole	25	Vacuole	Cellular Component

APPENDIX F3. (cont.)

4	endomembrane system	34	cell part	Cellular Component
3	membrane-bounded organelle	1307	Organelle	Cellular Component
3	vesicle	66	Organelle	Cellular Component
4	envelope	34	cell part	Cellular Component
6	intracellular organelle lumen	241	intracellular organelle part, organelle lumen	Cellular Component
Null	all	0		Cellular Component
2	extracellular matrix	14	cellular_component	Cellular Component
6	intracellular non-membrane-bounded organelle	695	intracellular organelle, non-membrane-bounded organelle	Cellular Component
4	organelle lumen	241	organelle part, membrane-enclosed lumen	Cellular Component
6	nuclear part	272	intracellular organelle part, nucleus	Cellular Component
3	non-membrane-bounded organelle	695	Organelle	Cellular Component
6	intracellular membrane-bounded organelle	1307	intracellular organelle, membrane-bounded organelle	Cellular Component
4	membrane-bounded vesicle	66	Vesicle	Cellular Component
5	cytoplasmic part	1098	intracellular part, cytoplasm	Cellular Component
3	cell part	2276	cell, cellular_component	Cellular Component
6	cytoskeletal part	56	intracellular organelle part, cytoskeleton	Cellular Component
5	intracellular organelle	1663	organelle, intracellular part	Cellular Component
2	membrane-enclosed lumen	241	cellular_component	Cellular Component
5	ribonucleoprotein complex	242	macromolecular complex, intracellular part	Cellular Component

APPENDIX F3. (cont.)

2	membrane	219	cellular_component	Cellular Component
6	microbody	23	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	cytoplasmic vesicle	66	cytoplasmic part, vesicle, intracellular organelle	Cellular Component
2	macromolecular complex	1003	cellular_component	Cellular Component
4	intracellular organelle part	312	organelle part, intracellular organelle, intracellular part	Cellular Component
7	nuclear lumen	241	nuclear part, intracellular organelle lumen	Cellular Component
8	microtubule cytoskeleton	56	Cytoskeleton	Cellular Component
4	cell projection	10	cell part	Cellular Component
4	cell periphery	219	cell part	Cellular Component
4	intracellular part	1985	cell part, intracellular	Cellular Component
3	organelle part	312	cellular_component, organelle	Cellular Component
2	catalytic activity	2538	molecular_function	Molecular Function
4	nucleotide binding	865	small molecule binding, nucleoside phosphate binding	Molecular Function
2	binding	2272	molecular_function	Molecular Function
3	hydrolase activity	1112	catalytic activity	Molecular Function
3	protein binding	708	Binding	Molecular Function
3	transferase activity	630	catalytic activity	Molecular Function
2	structural molecule activity	315	molecular_function	Molecular Function
2	transporter activity	331	molecular_function	Molecular Function
4	peptidase activity	283	hydrolase activity	Molecular Function

APPENDIX F3. (cont.)

4	RNA binding	284	nucleic acid binding	Molecular Function
4	DNA binding	177	nucleic acid binding	Molecular Function
3	nucleic acid binding	590	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
6	protein kinase activity	124	kinase activity, phosphotransferase activity, alcohol group as acceptor	Molecular Function
2	enzyme regulator activity	104	molecular_function	Molecular Function
5	translation factor activity, nucleic acid binding	100	RNA binding	Molecular Function
2	electron carrier activity	90	molecular_function	Molecular Function
5	kinase activity	212	transferase activity, transferring phosphorus-containing groups	Molecular Function
6	calcium ion binding	83	metal ion binding	Molecular Function
3	lipid binding	82	Binding	Molecular Function
5	actin binding	70	cytoskeletal protein binding	Molecular Function
2	receptor activity	61	molecular_function	Molecular Function
3	carbohydrate binding	56	Binding	Molecular Function
3	sequence-specific DNA binding transcription factor activity	53	nucleic acid binding transcription factor activity	Molecular Function
4	receptor binding	52	protein binding	Molecular Function
7	phosphoprotein phosphatase activity	46	phosphatase activity	Molecular Function
4	cytoskeletal protein binding	104	protein binding	Molecular Function
Null	transcription regulator activity	33	obsolete_molecular_function	Molecular Function

APPENDIX F3. (cont.)

3	signal transducer activity	31	molecular transducer activity	Molecular Function
3	chromatin binding	30	Binding	Molecular Function
6	ion channel activity	30	substrate-specific channel activity, ion transmembrane transporter activity	Molecular Function
2	antioxidant activity	29	molecular_function	Molecular Function
5	nuclease activity	28	hydrolase activity, acting on ester bonds	Molecular Function
8	motor activity	26	nucleoside-triphosphatase activity	Molecular Function
5	carboxylic ester hydrolase activity	17	hydrolase activity, acting on ester bonds	Molecular Function
3	organic cyclic compound binding	1247	Binding	Molecular Function
4	neurotransmitter transporter activity	3	transporter activity	Molecular Function
2	translation regulator activity	1	molecular_function	Molecular Function
2	protein tag	1	molecular_function	Molecular Function
6	7SK snRNA binding	1	snRNA binding	Molecular Function
7	inositol phosphate phosphatase activity	1	phosphatase activity	Molecular Function
3	oxygen binding	1	Binding	Molecular Function
4	substrate-specific transmembrane transporter activity	30	substrate-specific transporter activity, transmembrane transporter activity	Molecular Function
5	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	26	hydrolase activity, acting on acid anhydrides	Molecular Function
2	molecular transducer activity	31	molecular_function	Molecular Function

APPENDIX F3. (cont.)

5	ion transmembrane transporter activity	30	substrate-specific transmembrane transporter activity	Molecular Function
5	metal ion binding	83	cation binding	Molecular Function
5	channel activity	30	passive transmembrane transporter activity	Molecular Function
5	phosphoric ester hydrolase activity	47	hydrolase activity, acting on ester bonds	Molecular Function
4	cation binding	83	ion binding	Molecular Function
4	hydrolase activity, acting on acid anhydrides	26	hydrolase activity	Molecular Function
4	hydrolase activity, acting on ester bonds	92	hydrolase activity	Molecular Function
3	transmembrane transporter activity	30	transporter activity	Molecular Function
5	snRNA binding	1	RNA binding	Molecular Function
6	phosphatase activity	47	phosphoric ester hydrolase activity	Molecular Function
5	phosphotransferase activity, alcohol group as acceptor	124	transferase activity, transferring phosphorus-containing groups	Molecular Function
3	small molecule binding	865	Binding	Molecular Function
2	nucleic acid binding transcription factor activity	53	molecular_function	Molecular Function
Null	all	0		Molecular Function
3	substrate-specific transporter activity	30	transporter activity	Molecular Function
3	ion binding	83	Binding	Molecular Function
4	passive transmembrane transporter activity	30	transmembrane transporter activity	Molecular Function
Null	obsolete_molecular_function	33		Molecular Function

APPENDIX F3. (cont.)

Null	nucleoside phosphate binding	865	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
6	pyrophosphatase activity	26	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Molecular Function
Null	heterocyclic compound binding	1243	Binding	Molecular Function
7	nucleoside-triphosphatase activity	26	pyrophosphatase activity	Molecular Function
5	substrate-specific channel activity	30	substrate-specific transmembrane transporter activity, channel activity	Molecular Function
4	transferase activity, transferring phosphorus-containing groups	212	transferase activity	Molecular Function

APPENDIX F4. ***M. sjostedti.***

GO Level	Term (Name)	#Sequence	Parents (Name)	Category
2	metabolic process	1862	biological_process	Biological Process
Null	single-organism process	1001	biological_process	Biological Process
Null	single-organism cellular process	810	single-organism process, cellular process	Biological Process
3	primary metabolic process	1402	metabolic process	Biological Process
2	cellular process	1567	biological_process	Biological Process
3	regulation of biological process	644	biological regulation, biological_process	Biological Process
2	developmental process	544	biological_process	Biological Process
3	multicellular organismal development	495	single-multicellular organism process, developmental process	Biological Process
3	cellular component organization	557	cellular component organization or biogenesis	Biological Process
3	organic substance metabolic process	1384	metabolic process	Biological Process
4	protein metabolic process	817	primary metabolic process, macromolecule metabolic process	Biological Process
4	nucleobase-containing compound metabolic process	512	heterocycle metabolic process, primary metabolic process, organic cyclic compound metabolic process, cellular aromatic compound metabolic process, cellular nitrogen compound metabolic process	Biological Process
4	transport	506	establishment of localization	Biological Process

APPENDIX F4. (cont.)

3	catabolic process	428	metabolic process	Biological Process
3	cellular metabolic process	1123	metabolic process, cellular process	Biological Process
2	response to stimulus	477	biological_process	Biological Process
2	biological regulation	655	biological_process	Biological Process
5	cellular protein metabolic process	608	protein metabolic process, cellular macromolecule metabolic process	Biological Process
3	biosynthetic process	641	metabolic process	Biological Process
5	organelle organization	396	single-organism cellular process, cellular component organization	Biological Process
4	cell differentiation	355	cellular developmental process	Biological Process
6	translation	352	cellular macromolecule biosynthetic process, cellular protein metabolic process, gene expression	Biological Process
3	macromolecule metabolic process	889	organic substance metabolic process	Biological Process
null	single-multicellular organism process	495	multicellular organismal process, single-organism process	Biological Process
2	cellular component organization or biogenesis	557	biological_process	Biological Process
4	heterocycle metabolic process	512	cellular metabolic process	Biological Process
4	cellular nitrogen compound metabolic process	512	cellular metabolic process, nitrogen compound metabolic process	Biological Process
null	organic cyclic compound metabolic process	512	organic substance metabolic process	Biological Process

APPENDIX F4. (cont.)

4	cellular aromatic compound metabolic process	512	cellular metabolic process	Biological Process
4	cellular macromolecule metabolic process	677	cellular metabolic process, macromolecule metabolic process	Biological Process
6	cellular protein modification process	265	cellular protein metabolic process, protein modification process	Biological Process
3	establishment of localization	506	biological_process, localization	Biological Process
4	anatomical structure morphogenesis	260	anatomical structure development, developmental process	Biological Process
3	cell cycle	257	single-organism cellular process	Biological Process
3	response to stress	240	response to stimulus	Biological Process
4	signal transduction	232	single organism signaling, cell communication, cellular response to stimulus, regulation of cellular process	Biological Process
3	anatomical structure development	303	developmental process	Biological Process
4	gene expression	367	macromolecule metabolic process	Biological Process
6	cytoskeleton organization	216	organelle organization	Biological Process
3	cellular developmental process	355	developmental process, single-organism cellular process	Biological Process
5	cellular macromolecule biosynthetic process	352	cellular macromolecule metabolic process, cellular biosynthetic process, macromolecule biosynthetic process	Biological Process

APPENDIX F4. (cont.)

3	cell communication	308	single-organism cellular process	Biological Process
2	multicellular organismal process	495	biological_process	Biological Process
null	single organism signaling	292	single-organism process, signaling	Biological Process
2	reproduction	184	biological_process	Biological Process
3	nitrogen compound metabolic process	512	metabolic process	Biological Process
4	generation of precursor metabolites and energy	164	cellular metabolic process	Biological Process
5	protein modification process	265	macromolecule modification, protein metabolic process	Biological Process
2	localization	506	biological_process	Biological Process
6	protein transport	153	organic substance transport, establishment of protein localization	Biological Process
4	regulation of cellular process	232	cellular process, regulation of biological process	Biological Process
3	cellular response to stimulus	232	response to stimulus, single-organism cellular process	Biological Process
4	carbohydrate metabolic process	136	primary metabolic process, organic substance metabolic process	Biological Process
5	ion transport	128	single-organism transport	Biological Process
4	cellular biosynthetic process	352	cellular metabolic process, biosynthetic process	Biological Process
4	macromolecule biosynthetic process	352	organic substance biosynthetic process, macromolecule metabolic process	Biological Process

APPENDIX F4. (cont.)

4	embryo development	122	single-organism developmental process, multicellular organismal development, anatomical structure development	Biological Process
2	signaling	292	biological_process	Biological Process
4	lipid metabolic process	111	primary metabolic process, organic substance metabolic process, single-organism metabolic process	Biological Process
3	cell death	109	death, single-organism cellular process	Biological Process
4	macromolecule modification	265	macromolecule metabolic process	Biological Process
6	DNA metabolic process	95	nucleic acid metabolic process, cellular macromolecule metabolic process	Biological Process
2	cell proliferation	95	single-organism process	Biological Process
5	organic substance transport	153	transport	Biological Process
5	establishment of protein localization	153	establishment of localization, protein localization	Biological Process
3	cell-cell signaling	85	single organism signaling, cell communication	Biological Process
null	single-organism metabolic process	129	metabolic process	Biological Process
null	single-organism transport	128	single-organism process, transport	Biological Process
null	organic substance biosynthetic process	352	organic substance metabolic process, biosynthetic process	Biological Process

APPENDIX F4. (cont.)

3	behavior	74	response to stimulus	Biological Process
3	response to external stimulus	74	response to stimulus	Biological Process
null	single-organism developmental process	122	single-organism process, developmental process	Biological Process
3	cellular homeostasis	70	homeostatic process, single-organism cellular process	Biological Process
2	death	110	single-organism process	Biological Process
2	growth	67	biological_process	Biological Process
5	nucleic acid metabolic process	95	macromolecule metabolic process, nucleobase-containing compound metabolic process	Biological Process
4	protein localization	153	macromolecule localization	Biological Process
3	response to abiotic stimulus	55	response to stimulus	Biological Process
4	homeostatic process	70	regulation of biological quality	Biological Process
2	viral reproduction	39	multi-organism cellular process	Biological Process
6	mitochondrion organization	38	organelle organization	Biological Process
3	response to biotic stimulus	36	response to stimulus	Biological Process
3	macromolecule localization	153	localization	Biological Process
3	response to endogenous stimulus	30	response to stimulus	Biological Process
3	secondary metabolic process	27	single-organism metabolic process	Biological Process
3	regulation of biological quality	70	biological regulation	Biological Process
null	multi-organism cellular process	39	multi-organism process, cellular process	Biological Process
2	multi-organism process	43	biological_process	Biological Process
4	symbiosis, encompassing mutualism through parasitism	21	interspecies interaction between organisms	Biological Process

APPENDIX F4. (cont.)

3	cell growth	20	growth, single-organism cellular process	Biological Process
7	regulation of gene expression, epigenetic	19	regulation of gene expression	Biological Process
5	cytoplasm organization	17	single-organism cellular process, cellular component organization	Biological Process
3	cell recognition	15	single-multicellular organism process, single-organism cellular process	Biological Process
3	interspecies interaction between organisms	21	multi-organism process	Biological Process
6	regulation of gene expression	19	gene expression, regulation of macromolecule metabolic process	Biological Process
5	regulation of macromolecule metabolic process	19	macromolecule metabolic process, regulation of metabolic process	Biological Process
4	regulation of metabolic process	19	regulation of biological process, metabolic process	Biological Process
5	intracellular signal transduction	2	signal transduction	Biological Process
2	cell	2076	cellular_component	Cellular Component
4	intracellular part	1679	cell part, intracellular	Cellular Component
5	cytoplasm	1292	intracellular part	Cellular Component
4	intracellular	1834	cell part	Cellular Component
3	cell part	1876	cell, cellular_component	Cellular Component
5	cytoplasmic part	898	intracellular part, cytoplasm	Cellular Component
6	intracellular membrane-bounded organelle	1057	intracellular organelle, membrane-bounded organelle	Cellular Component
3	protein complex	642	macromolecular complex	Cellular Component

APPENDIX F4. (cont.)

5	intracellular organelle	1402	organelle, intracellular part	Cellular Component
7	nucleus	579	intracellular membrane-bounded organelle	Cellular Component
2	macromolecular complex	859	cellular_component	Cellular Component
2	organelle	1434	cellular_component	Cellular Component
6	intracellular non-membrane-bounded organelle	636	intracellular organelle, non-membrane-bounded organelle	Cellular Component
3	membrane-bounded organelle	1057	organelle	Cellular Component
6	mitochondrion	315	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
7	cytoskeleton	276	intracellular non-membrane-bounded organelle	Cellular Component
3	non-membrane-bounded organelle	636	organelle	Cellular Component
6	ribosome	219	ribonucleoprotein complex, cytoplasmic part, intracellular non-membrane-bounded organelle	Cellular Component
6	cytosol	204	cytoplasmic part	Cellular Component
6	nuclear part	270	intracellular organelle part, nucleus	Cellular Component
7	nuclear lumen	239	nuclear part, intracellular organelle lumen	Cellular Component
4	plasma membrane	161	cell part, cell periphery, membrane	Cellular Component
6	nucleoplasm	141	nuclear part, nuclear lumen	Cellular Component
5	ribonucleoprotein complex	219	macromolecular complex, intracellular part	Cellular Component
4	intracellular organelle part	297	organelle part, intracellular organelle, intracellular part	Cellular Component

APPENDIX F4. (cont.)

6	lipid particle	126	cytoplasmic part	Cellular Component
6	endoplasmic reticulum	110	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	intracellular organelle lumen	239	intracellular organelle part, organelle lumen	Cellular Component
6	nucleolus	102	nuclear part, nuclear lumen, intracellular non-membrane-bounded organelle	Cellular Component
4	cell periphery	161	cell part	Cellular Component
2	membrane	161	cellular_component	Cellular Component
7	chromosome	101	intracellular non-membrane-bounded organelle	Cellular Component
6	Golgi apparatus	82	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
3	organelle part	297	cellular_component, organelle	Cellular Component
2	extracellular region	86	cellular_component	Cellular Component
4	organelle lumen	239	organelle part, membrane-enclosed lumen	Cellular Component
7	cytoplasmic membrane-bounded vesicle	61	cytoplasmic vesicle, intracellular membrane-bounded organelle, membrane-bounded vesicle	Cellular Component
6	nuclear chromosome	43	nuclear part, nuclear lumen, chromosome	Cellular Component
5	nuclear envelope	43	nuclear part, organelle envelope, endomembrane system	Cellular Component
2	membrane-enclosed lumen	239	cellular_component	Cellular Component
4	membrane-bounded vesicle	61	vesicle	Cellular Component

APPENDIX F4. (cont.)

6	cytoplasmic vesicle	61	cytoplasmic part, vesicle, intracellular organelle	Cellular Component
7	microtubule organizing center	36	microtubule cytoskeleton, cytoskeletal part	Cellular Component
6	vacuole	43	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
6	endosome	32	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
4	organelle envelope	43	intracellular organelle part, membrane-bounded organelle, envelope	Cellular Component
4	endomembrane system	43	cell part	Cellular Component
3	vesicle	61	organelle	Cellular Component
6	cytoskeletal part	36	intracellular organelle part, cytoskeleton	Cellular Component
8	microtubule cytoskeleton	36	cytoskeleton	Cellular Component
8	lysosome	19	lytic vacuole	Cellular Component
7	peroxisome	17	microbody	Cellular Component
4	extracellular space	16	extracellular region part	Cellular Component
4	envelope	43	cell part	Cellular Component
3	extracellular region part	20	cellular_component, extracellular region	Cellular Component
7	lytic vacuole	19	vacuole	Cellular Component
6	microbody	17	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component
4	proteinaceous extracellular matrix	6	extracellular region part, extracellular matrix	Cellular Component
6	plastid	5	cytoplasmic part, intracellular membrane-bounded organelle	Cellular Component

APPENDIX F4. (cont.)

5	cilium	4	cell projection, intracellular membrane-bounded organelle	Cellular Component
2	extracellular matrix	6	cellular_component	Cellular Component
4	cell projection	4	cell part	Cellular Component
5	cell wall	1	external encapsulating structure	Cellular Component
4	external encapsulating structure	1	cell part, cell periphery	Cellular Component
2	binding	1760	molecular_function	Molecular Function
2	catalytic activity	1766	molecular_function	Molecular Function
4	nucleotide binding	683	small molecule binding, nucleoside phosphate binding	Molecular Function
3	hydrolase activity	744	catalytic activity	Molecular Function
3	protein binding	567	binding	Molecular Function
3	organic cyclic compound binding	1016	binding	Molecular Function
Null	heterocyclic compound binding	1010	binding	Molecular Function
3	small molecule binding	683	binding	Molecular Function
Null	nucleoside phosphate binding	683	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
3	transferase activity	470	catalytic activity	Molecular Function
3	nucleic acid binding	506	heterocyclic compound binding, organic cyclic compound binding	Molecular Function
2	structural molecule activity	280	molecular_function	Molecular Function
2	transporter activity	255	molecular_function	Molecular Function
4	RNA binding	256	nucleic acid binding	Molecular Function
4	peptidase activity	174	hydrolase activity	Molecular Function
4	DNA binding	159	nucleic acid binding	Molecular Function

APPENDIX F4. (cont.)

5	kinase activity	159	transferase activity, transferring phosphorus-containing groups	Molecular Function
5	translation factor activity, nucleic acid binding	104	RNA binding	Molecular Function
4	cytoskeletal protein binding	112	protein binding	Molecular Function
6	protein kinase activity	94	kinase activity, phosphotransferase activity, alcohol group as acceptor	Molecular Function
2	enzyme regulator activity	84	molecular_function	Molecular Function
6	calcium ion binding	78	metal ion binding	Molecular Function
5	actin binding	76	cytoskeletal protein binding	Molecular Function
4	transferase activity, transferring phosphorus-containing groups	159	transferase activity	Molecular Function
3	lipid binding	66	binding	Molecular Function
2	electron carrier activity	58	molecular_function	Molecular Function
5	phosphotransferase activity, alcohol group as acceptor	94	transferase activity, transferring phosphorus-containing groups	Molecular Function
5	metal ion binding	78	cation binding	Molecular Function
7	phosphoprotein phosphatase activity	42	phosphatase activity	Molecular Function
3	sequence-specific DNA binding transcription factor activity	35	nucleic acid binding transcription factor activity	Molecular Function
4	receptor binding	34	protein binding	Molecular Function
8	motor activity	30	nucleoside-triphosphatase activity	Molecular Function

APPENDIX F4. (cont.)

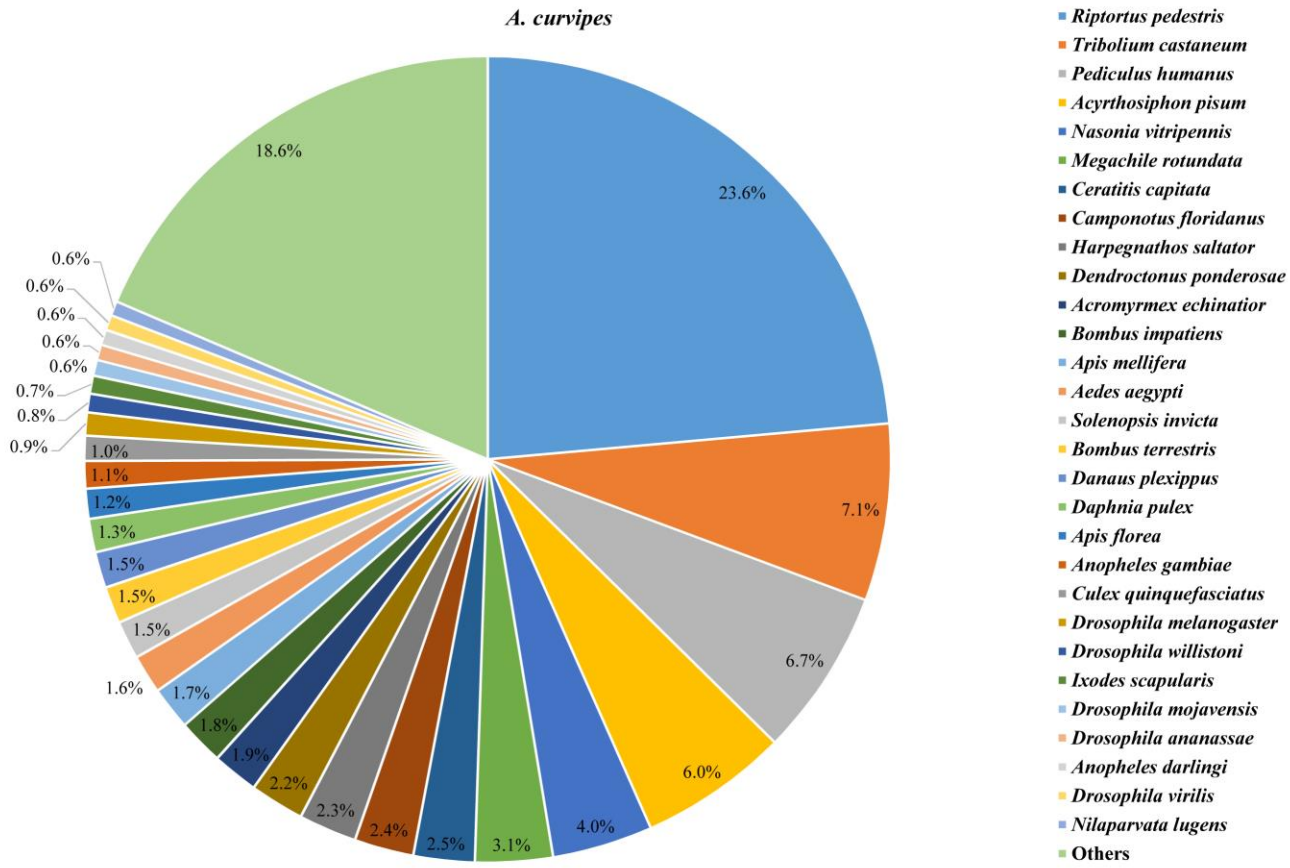
4	cation binding	78	ion binding	Molecular Function
4	hydrolase activity, acting on ester bonds	73	hydrolase activity	Molecular Function
6	phosphatase activity	42	phosphoric ester hydrolase activity	Molecular Function
3	chromatin binding	25	binding	Molecular Function
5	nuclease activity	24	hydrolase activity, acting on ester bonds	Molecular Function
null	transcription regulator activity	22	obsolete_molecular_function	Molecular Function
3	carbohydrate binding	22	binding	Molecular Function
2	receptor activity	21	molecular_function	Molecular Function
2	nucleic acid binding transcription factor activity	35	molecular_function	Molecular Function
7	nucleoside-triphosphatase activity	30	pyrophosphatase activity	Molecular Function
6	ion channel activity	18	substrate-specific channel activity, ion transmembrane transporter activity	Molecular Function
3	ion binding	78	binding	Molecular Function
5	phosphoric ester hydrolase activity	42	hydrolase activity, acting on ester bonds	Molecular Function
null	obsolete_molecular_function	22		Molecular Function
2	antioxidant activity	12	molecular_function	Molecular Function
3	signal transducer activity	12	molecular transducer activity	Molecular Function
5	ion transmembrane transporter activity	18	substrate-specific transmembrane transporter activity	Molecular Function

APPENDIX F4. (cont.)

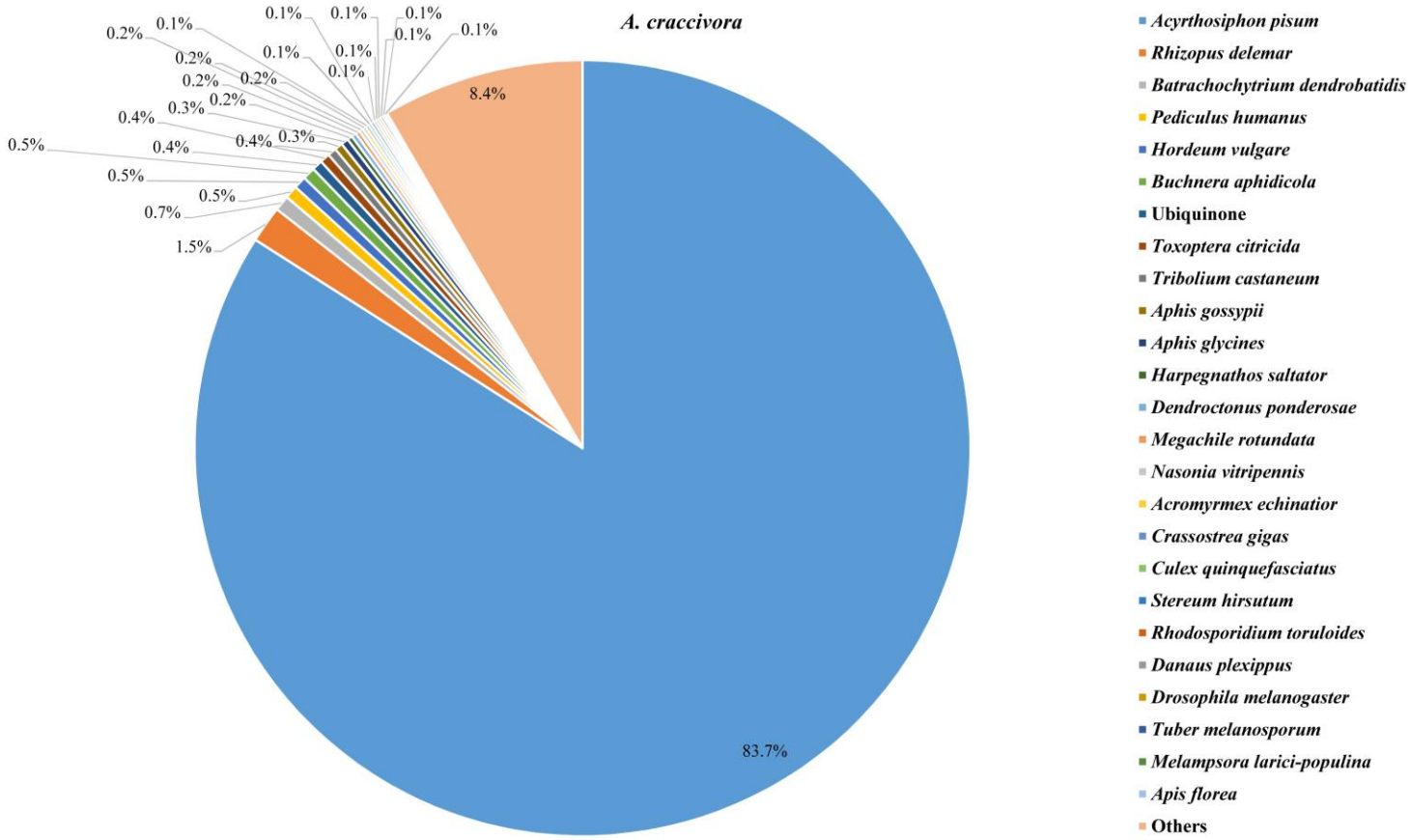
6	pyrophosphatase activity	30	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Molecular Function
5	substrate-specific channel activity	18	substrate-specific transmembrane transporter activity, channel activity	Molecular Function
2	molecular transducer activity	12	molecular_function	Molecular Function
5	carboxylic ester hydrolase activity	7	hydrolase activity, acting on ester bonds	Molecular Function
4	substrate-specific transmembrane transporter activity	18	substrate-specific transporter activity, transmembrane transporter activity	Molecular Function
5	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	30	hydrolase activity, acting on acid anhydrides	Molecular Function
5	channel activity	18	passive transmembrane transporter activity	Molecular Function
2	nutrient reservoir activity	6	molecular_function	Molecular Function
4	hydrolase activity, acting on acid anhydrides	30	hydrolase activity	Molecular Function
3	transmembrane transporter activity	18	transporter activity	Molecular Function
3	substrate-specific transporter activity	18	transporter activity	Molecular Function
4	passive transmembrane transporter activity	18	transmembrane transporter activity	Molecular Function
2	translation regulator activity	1	molecular_function	Molecular Function

APPENDIX G
Species distribution of the top Blastx hits

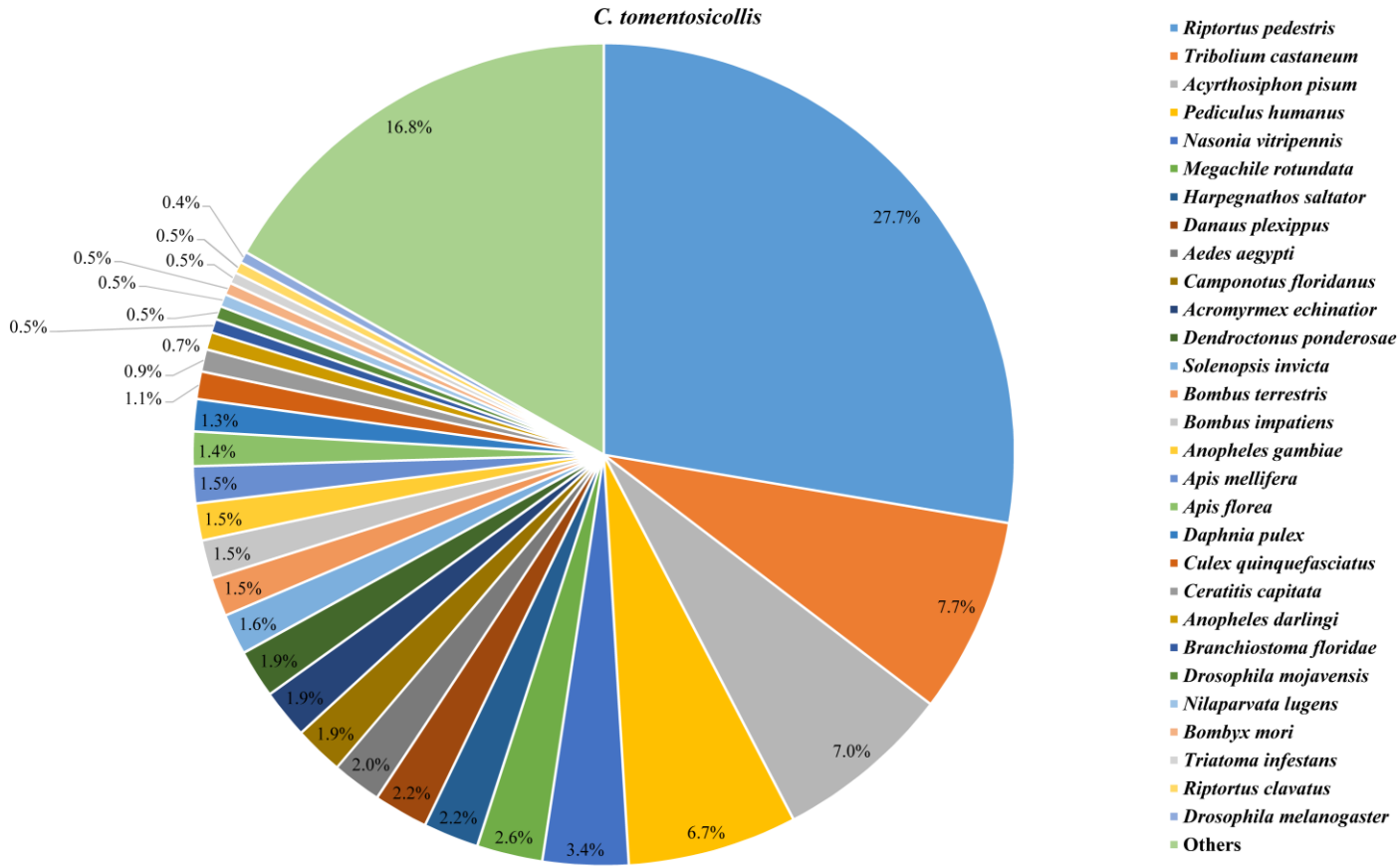
APPENDIX G1. *A. curvipes.*



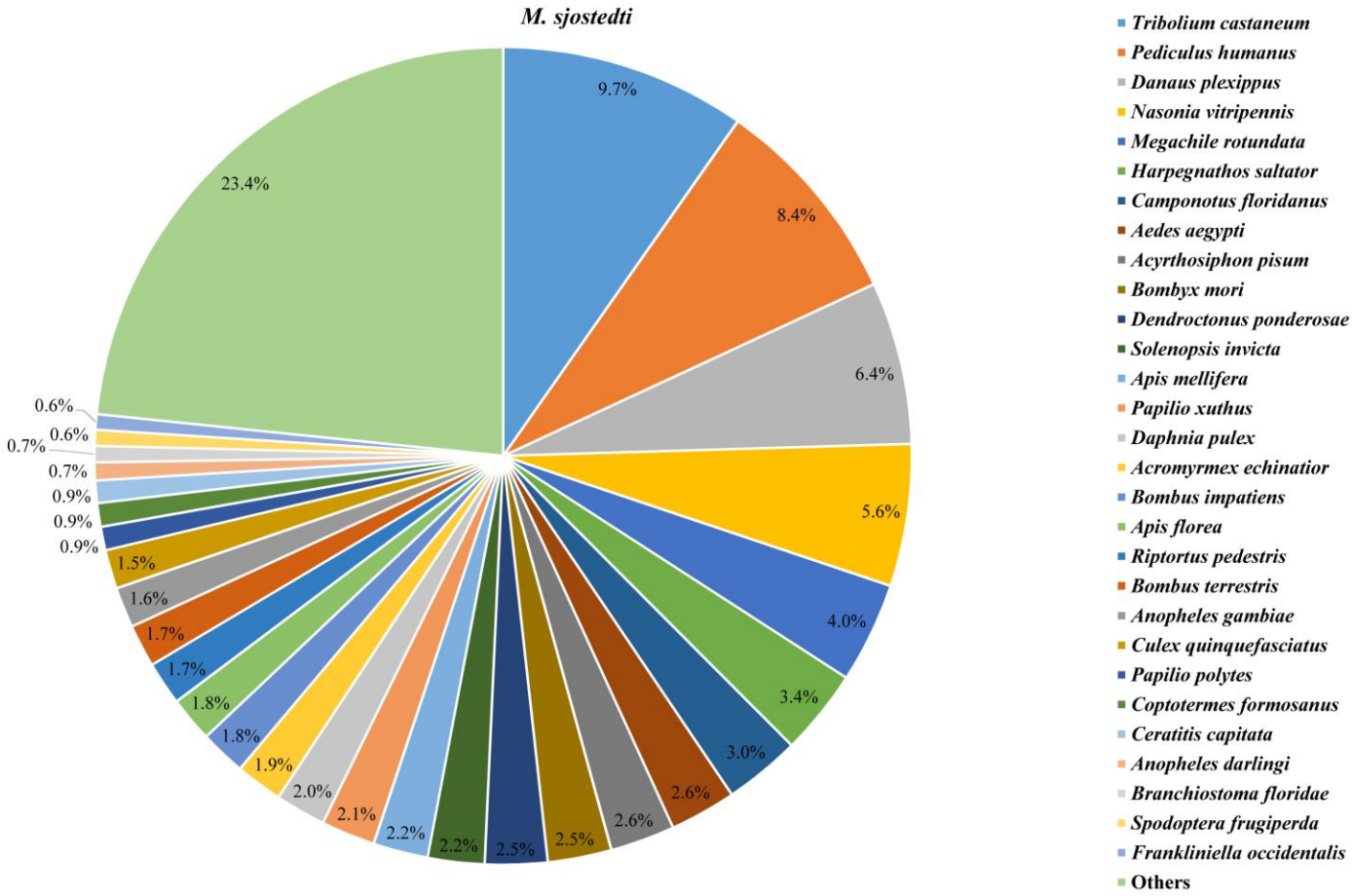
APPENDIX G2. *A. craccivora.*



APPENDIX G3. *C. tomentosicollis.*



APPENDIX G4. *M. sjostedti.*



APPENDIX H
InterProScan results from *A. craccivora* contigs that showed Blastn hits to *B. aphidicola*

Contig ID	InterProScan			
Aphis 47	noIPR	Unintegrated	unintegrated	SignalP-NN(euk) (SIGNALP)
Aphis 47	noIPR	Unintegrated	unintegrated	tmhmm (TMHMM)
Aphis 47	noIPR	Unintegrated	unintegrated	
Aphis 65	IPR025472	Protein of unknown function DUF4323	Family	PF14211 (PFAM)
Aphis 65	IPR025472	Protein of unknown function DUF4323	Family	
Aphis 221	IPR000850	Adenylate kinase	Family	PR00094 (PRINTS)
Aphis 221	IPR000850	Adenylate kinase	Family	PTHR23359 (PANTHER)
Aphis 221	IPR000850	Adenylate kinase	Family	PF00406 (PFAM)
Aphis 221	IPR000850	Adenylate kinase	Family	GO:0005524
Aphis 221	IPR000850	Adenylate kinase	Family	GO:0006139
Aphis 221	IPR000850	Adenylate kinase	Family	GO:0019205
Aphis 221	IPR027417	P-loop containing nucleoside triphosphate hydrolase	Domain	SSF52540 (SUPERFAMILY)
Aphis 221	IPR027417	P-loop containing nucleoside triphosphate hydrolase	Domain	
Aphis 221	noIPR	Unintegrated	unintegrated	G3DSA:3.40.50.300 (GENE3D)
Aphis 221	noIPR	Unintegrated	unintegrated	
Aphis 367	noIPR	Unintegrated	unintegrated	SignalP-NN(euk) (SIGNALP)
Aphis 367	noIPR	Unintegrated	unintegrated	
Aphis 1561	IPR001347	Sugar isomerase (SIS)	Domain	PF01380 (PFAM)
Aphis 1561	IPR001347	Sugar isomerase (SIS)	Domain	PS51464 (PROFILE)
Aphis 1561	IPR001347	Sugar isomerase (SIS)	Domain	GO:0005975

APPENDIX H. (cont.)

Aphis 1561	IPR001347	Sugar isomerase (SIS)	Domain	GO:0030246
Aphis 1561	IPR005855	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	Family	PTHR10937:SF0 (PANTHER)
Aphis 1561	IPR005855	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	Family	GO:0004360
Aphis 1561	IPR005855	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	Family	GO:0005737
Aphis 1561	IPR005855	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	Family	GO:0016051
Aphis 1561	noIPR	unintegrated	unintegrated	G3DSA:3.40.50.10490 (GENE3D)
Aphis 1561	noIPR	unintegrated	unintegrated	PTHR10937 (PANTHER)
Aphis 1561	noIPR	unintegrated	unintegrated	SSF53697 (SUPERFAMILY)
Aphis 1561	noIPR	unintegrated	unintegrated	
Aphis 2076	IPR002508	Cell wall hydrolase/autolysin, catalytic	Domain	G3DSA:3.40.630.40 (GENE3D)
Aphis 2076	IPR002508	Cell wall hydrolase/autolysin, catalytic	Domain	PF01520 (PFAM)
Aphis 2076	IPR002508	Cell wall hydrolase/autolysin, catalytic	Domain	GO:0008745
Aphis 2076	IPR002508	Cell wall hydrolase/autolysin, catalytic	Domain	GO:0009253
Aphis 2076	noIPR	unintegrated	unintegrated	PTHR30404 (PANTHER)
Aphis 2076	noIPR	unintegrated	unintegrated	PTHR30404:SF0 (PANTHER)
Aphis 2076	noIPR	unintegrated	unintegrated	
Aphis 2391	IPR001353	Proteasome, subunit alpha/beta	Family	PF00227 (PFAM)
Aphis 2391	IPR001353	Proteasome, subunit alpha/beta	Family	GO:0004298
Aphis 2391	IPR001353	Proteasome, subunit alpha/beta	Family	GO:0005839
Aphis 2391	IPR001353	Proteasome, subunit alpha/beta	Family	GO:0051603

APPENDIX H. (cont.)

Aphis 2391	noIPR	Unintegrated	unintegrated	G3DSA:3.60.20.10 (GENE3D)
Aphis 2391	noIPR	Unintegrated	unintegrated	PTHR32194 (PANTHER)
Aphis 2391	noIPR	Unintegrated	unintegrated	PTHR32194:SF0 (PANTHER)
Aphis 2391	noIPR	Unintegrated	unintegrated	SSF56235 (SUPERFAMILY)
Aphis 2391	noIPR	Unintegrated	unintegrated	
Aphis 2472	IPR004160	Translation elongation factor EFTu/EF1A, C-terminal	Domain	PF03143 (PFAM)
Aphis 2472	IPR004160	Translation elongation factor EFTu/EF1A, C-terminal	Domain	GO:0005525
Aphis 2472	IPR004161	Translation elongation factor EFTu/EF1A, domain 2	Domain	PF03144 (PFAM)
Aphis 2472	IPR004161	Translation elongation factor EFTu/EF1A, domain 2	Domain	GO:0005525
Aphis 2472	IPR004541	Translation elongation factor EFTu/EF1A, bacterial/organelle	Family	PTHR23115:SF31 (PANTHER)
Aphis 2472	IPR004541	Translation elongation factor EFTu/EF1A, bacterial/organelle	Family	GO:0003746
Aphis 2472	IPR004541	Translation elongation factor EFTu/EF1A, bacterial/organelle	Family	GO:0005525
Aphis 2472	IPR004541	Translation elongation factor EFTu/EF1A, bacterial/organelle	Family	GO:0005622
Aphis 2472	IPR004541	Translation elongation factor EFTu/EF1A, bacterial/organelle	Family	GO:0006414
Aphis 2472	IPR009000	Translation elongation/initiation factor/Ribosomal, beta-barrel	Domain	SSF50447 (SUPERFAMILY)
Aphis 2472	IPR009000	Translation elongation/initiation factor/Ribosomal, beta-barrel	Domain	
Aphis 2472	IPR009001	Translation elongation factor EF1A/initiation factor IF2gamma, C-terminal	Domain	SSF50465 (SUPERFAMILY)
Aphis 2472	IPR009001	Translation elongation factor EF1A/initiation factor IF2gamma, C-terminal	Domain	
Aphis 2472	noIPR	Unintegrated	unintegrated	G3DSA:2.40.30.10 (GENE3D)
Aphis 2472	noIPR	Unintegrated	unintegrated	PTHR23115 (PANTHER)

APPENDIX H. (cont.)

Aphis 2472	noIPR	unintegrated	unintegrated	
Aphis 2530	IPR000529	Ribosomal protein S6	Family	PF01250 (PFAM)
Aphis 2530	IPR000529	Ribosomal protein S6	Family	TIGR00166 (TIGRFAMs)
Aphis 2530	IPR000529	Ribosomal protein S6	Family	SSF54995 (SUPERFAMILY)
Aphis 2530	IPR000529	Ribosomal protein S6	Family	GO:0003735
Aphis 2530	IPR000529	Ribosomal protein S6	Family	GO:0005840
Aphis 2530	IPR000529	Ribosomal protein S6	Family	GO:0006412
Aphis 2530	IPR000529	Ribosomal protein S6	Family	GO:0019843
Aphis 2530	IPR014717	Translation elongation factor EF1B/ribosomal protein S6	Domain	G3DSA:3.30.70.60 (GENE3D)
Aphis 2530	IPR014717	Translation elongation factor EF1B/ribosomal protein S6	Domain	
Aphis 2530	IPR020815	Ribosomal protein S6, conserved site	Conserved_site	PS01048 (PROSITE)
Aphis 2530	IPR020815	Ribosomal protein S6, conserved site	Conserved_site	GO:0003735
Aphis 2530	IPR020815	Ribosomal protein S6, conserved site	Conserved_site	GO:0005840
Aphis 2530	IPR020815	Ribosomal protein S6, conserved site	Conserved_site	GO:0006412
Aphis 2530	IPR020815	Ribosomal protein S6, conserved site	Conserved_site	GO:0019843
Aphis 2736	noIPR	unintegrated	unintegrated	SignalP-NN(euk) (SIGNALP)
Aphis 2736	noIPR	unintegrated	unintegrated	
Aphis 3831	IPR006847	Translation initiation factor IF-2, N-terminal	Domain	PF04760 (PFAM)
Aphis 3831	IPR006847	Translation initiation factor IF-2, N-terminal	Domain	GO:0003743
Aphis 3831	IPR006847	Translation initiation factor IF-2, N-terminal	Domain	GO:0006413
Aphis 3831	IPR009061	DNA binding domain, putative	Domain	SSF46955 (SUPERFAMILY)
Aphis 3831	IPR009061	DNA binding domain, putative	Domain	GO:0000166

APPENDIX H. (cont.)

Aphis 3831	IPR013575	Initiation factor 2 associated domain, bacterial	Domain	PF08364 (PFAM)
Aphis 3831	IPR013575	Initiation factor 2 associated domain, bacterial	Domain	
Aphis 3831	noIPR	Unintegrated	unintegrated	G3DSA:3.30.56.50 (GENE3D)
Aphis 3831	noIPR	Unintegrated	unintegrated	
Aphis 3870	IPR016484	GTP-binding protein EngA	Family	PTHR11649:SF5 (PANTHER)
Aphis 3870	IPR016484	GTP-binding protein EngA	Family	GO:0005525
Aphis 3870	noIPR	Unintegrated	unintegrated	G3DSA:3.40.50.300 (GENE3D)
Aphis 3870	noIPR	Unintegrated	unintegrated	PTHR11649 (PANTHER)
Aphis 3870	noIPR	Unintegrated	unintegrated	
Aphis 4564	IPR000795	Elongation factor, GTP-binding domain	Domain	PF00009 (PFAM)
Aphis 4564	IPR000795	Elongation factor, GTP-binding domain	Domain	GO:0003924
Aphis 4564	IPR000795	Elongation factor, GTP-binding domain	Domain	GO:0005525
Aphis 4564	IPR005225	Small GTP-binding protein domain	Domain	TIGR00231 (TIGRFAMs)
Aphis 4564	IPR005225	Small GTP-binding protein domain	Domain	GO:0005525
Aphis 4564	IPR006847	Translation initiation factor IF-2, N-terminal	Domain	PF04760 (PFAM)
Aphis 4564	IPR006847	Translation initiation factor IF-2, N-terminal	Domain	GO:0003743
Aphis 4564	IPR006847	Translation initiation factor IF-2, N-terminal	Domain	GO:0006413
Aphis 4564	IPR015760	Translation initiation factor IF- 2	Family	PTHR23115:SF41 (PANTHER)
Aphis 4564	IPR015760	Translation initiation factor IF- 2	Family	
Aphis 4564	IPR027417	P-loop containing nucleoside triphosphate hydrolase	Domain	SSF52540 (SUPERFAMILY)
Aphis 4564	IPR027417	P-loop containing nucleoside triphosphate hydrolase	Domain	

APPENDIX H. (cont.)

Aphis 4564	noIPR	unintegrated	unintegrated	G3DSA:3.40.50.300 (GENE3D)
Aphis 4564	noIPR	unintegrated	unintegrated	PTHR23115 (PANTHER)
Aphis 4564	noIPR	unintegrated	unintegrated	
Aphis 4568	IPR000819	Peptidase M17, leucyl aminopeptidase, C-terminal	Domain	PF00883 (PFAM)
Aphis 4568	IPR000819	Peptidase M17, leucyl aminopeptidase, C-terminal	Domain	GO:0004177
Aphis 4568	IPR000819	Peptidase M17, leucyl aminopeptidase, C-terminal	Domain	GO:0005622
Aphis 4568	IPR000819	Peptidase M17, leucyl aminopeptidase, C-terminal	Domain	GO:0006508
Aphis 4568	noIPR	unintegrated	unintegrated	G3DSA:3.40.630.10 (GENE3D)
Aphis 4568	noIPR	unintegrated	unintegrated	PTHR11963 (PANTHER)
Aphis 4568	noIPR	unintegrated	unintegrated	PTHR11963:SF4 (PANTHER)
Aphis 4568	noIPR	unintegrated	unintegrated	SSF53187 (SUPERFAMILY)
Aphis 4568	noIPR	unintegrated	unintegrated	
Aphis 5021	IPR027417	P-loop containing nucleoside triphosphate hydrolase	Domain	SSF52540 (SUPERFAMILY)
Aphis 5021	IPR027417	P-loop containing nucleoside triphosphate hydrolase	Domain	
Aphis 5021	noIPR	unintegrated	unintegrated	G3DSA:3.40.50.300 (GENE3D)
Aphis 5021	noIPR	unintegrated	unintegrated	PTHR19211 (PANTHER)
Aphis 5021	noIPR	unintegrated	unintegrated	PTHR19211:SF7 (PANTHER)
Aphis 5021	noIPR	unintegrated	unintegrated	
Aphis 5225	IPR002133	S-adenosylmethionine synthetase	Family	PTHR11964 (PANTHER)
Aphis 5225	IPR002133	S-adenosylmethionine synthetase	Family	GO:0004478
Aphis 5225	IPR002133	S-adenosylmethionine synthetase	Family	GO:0005524

APPENDIX H. (cont.)

Aphis 5225	IPR002133	S-adenosylmethionine synthetase	Family	GO:0006556
Aphis 5225	IPR022628	S-adenosylmethionine synthetase, N-terminal	Domain	PF00438 (PFAM)
Aphis 5225	IPR022628	S-adenosylmethionine synthetase, N-terminal	Domain	GO:0004478
Aphis 5225	IPR022628	S-adenosylmethionine synthetase, N-terminal	Domain	GO:0006556
Aphis 5225	IPR022636	S-adenosylmethionine synthetase superfamily	Domain	SSF55973 (SUPERFAMILY)
Aphis 5225	IPR022636	S-adenosylmethionine synthetase superfamily	Domain	GO:0004478
Aphis 5225	IPR022636	S-adenosylmethionine synthetase superfamily	Domain	GO:0006556
Aphis 5225	noIPR	Unintegrated	unintegrated	G3DSA:3.30.300.10 (GENE3D)
Aphis 5225	noIPR	Unintegrated	unintegrated	PTHR11964:SF0 (PANTHER)
Aphis 5225	noIPR	Unintegrated	unintegrated	
Aphis 5524	IPR008580	Domain of unknown function DUF862, eukaryotic	Domain	PF05903 (PFAM)
Aphis 5524	IPR008580	Domain of unknown function DUF862, eukaryotic	Domain	
Aphis 5524	noIPR	unintegrated	unintegrated	PTHR12378 (PANTHER)
Aphis 5524	noIPR	unintegrated	unintegrated	
Aphis 5691	IPR001451	Bacterial transferase hexapeptide repeat	Repeat	PF00132 (PFAM)
Aphis 5691	IPR001451	Bacterial transferase hexapeptide repeat	Repeat	
Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	PTHR22572:SF17 (PANTHER)
Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	GO:0000287
Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	GO:0000902

APPENDIX H. (cont.)

Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	GO:0003977
Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	GO:0005737
Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	GO:0009103
Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	GO:0009252
Aphis 5691	IPR005882	Bifunctional UDP-N-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate N-acetyltransferase	Family	GO:0019134
Aphis 5691	IPR011004	Trimeric LpxA-like	Domain	SSF51161 (SUPERFAMILY)
Aphis 5691	IPR011004	Trimeric LpxA-like	Domain	GO:0016740
Aphis 5691	noIPR	unintegrated	unintegrated	G3DSA:2.160.10.10 (GENE3D)
Aphis 5691	noIPR	unintegrated	unintegrated	PTHR22572 (PANTHER)
Aphis 5691	noIPR	unintegrated	unintegrated	
Aphis 5984	IPR001844	Chaperonin Cpn60	Family	PR00298 (PRINTS)
Aphis 5984	IPR001844	Chaperonin Cpn60	Family	GO:0005737
Aphis 5984	IPR001844	Chaperonin Cpn60	Family	GO:0042026
Aphis 5984	IPR002423	Chaperonin Cpn60/TCP-1	Family	PTHR11353 (PANTHER)
Aphis 5984	IPR002423	Chaperonin Cpn60/TCP-1	Family	PF00118 (PFAM)

APPENDIX H. (cont.)

Aphis 5984	IPR002423	Chaperonin Cpn60/TCP-1	Family	SSF48592 (SUPERFAMILY)
Aphis 5984	IPR002423	Chaperonin Cpn60/TCP-1	Family	GO:0005524
Aphis 5984	IPR002423	Chaperonin Cpn60/TCP-1	Family	GO:0044267
Aphis 5984	IPR018370	Chaperonin Cpn60, conserved site	Conserved_site	PS00296 (PROSITE)
Aphis 5984	IPR018370	Chaperonin Cpn60, conserved site	Conserved_site	GO:0005524
Aphis 5984	IPR018370	Chaperonin Cpn60, conserved site	Conserved_site	GO:0005737
Aphis 5984	IPR018370	Chaperonin Cpn60, conserved site	Conserved_site	GO:0006457
Aphis 5984	IPR027409	GroEL-like apical domain	Domain	G3DSA:3.50.7.10 (GENE3D)
Aphis 5984	IPR027409	GroEL-like apical domain	Domain	SSF52029 (SUPERFAMILY)
Aphis 5984	IPR027409	GroEL-like apical domain	Domain	
Aphis 5984	IPR027413	GroEL-like equatorial domain	Domain	G3DSA:1.10.560.10 (GENE3D)
Aphis 5984	IPR027413	GroEL-like equatorial domain	Domain	
Aphis 5984	noIPR	unintegrated	unintegrated	PTHR11353:SF10 (PANTHER)
Aphis 5984	noIPR	unintegrated	unintegrated	
Aphis 6159	IPR011603	2-oxoglutarate dehydrogenase, E1 component	Family	PTHR23152 (PANTHER)
Aphis 6159	IPR011603	2-oxoglutarate dehydrogenase, E1 component	Family	GO:0004591
Aphis 6159	IPR011603	2-oxoglutarate dehydrogenase, E1 component	Family	GO:0006099
Aphis 6159	IPR011603	2-oxoglutarate dehydrogenase, E1 component	Family	GO:0030976
Aphis 6159	IPR011603	2-oxoglutarate dehydrogenase, E1 component	Family	GO:0055114
Aphis 6159	noIPR	unintegrated	unintegrated	PTHR23152:SF0 (PANTHER)
Aphis 6159	noIPR	unintegrated	unintegrated	

APPENDIX H. (cont.)

Aphis 6899	noIPR	unintegrated	unintegrated	SignalP-NN(euk) (SIGNALP)
Aphis 6899	noIPR	unintegrated	unintegrated	
Aphis 7020	IPR002903	Ribosomal RNA small subunit methyltransferase H	Family	PF01795 (PFAM)
Aphis 7020	IPR002903	Ribosomal RNA small subunit methyltransferase H	Family	GO:0008168
Aphis 7209	noIPR	unintegrated	unintegrated	G3DSA:3.40.1280.10 (GENE3D)
Aphis 7209	noIPR	unintegrated	unintegrated	
Aphis 7344	IPR000454	ATPase, F0 complex, subunit C	Family	PR00124 (PRINTS)
Aphis 7344	IPR000454	ATPase, F0 complex, subunit C	Family	G3DSA:1.20.20.10 (GENE3D)
Aphis 7344	IPR000454	ATPase, F0 complex, subunit C	Family	GO:0015078
Aphis 7344	IPR000454	ATPase, F0 complex, subunit C	Family	GO:0015986
Aphis 7344	IPR002379	V-ATPase proteolipid subunit C-like domain	Domain	PF00137 (PFAM)
Aphis 7344	IPR002379	V-ATPase proteolipid subunit C-like domain	Domain	SSF81333 (SUPERFAMILY)
Aphis 7344	IPR002379	V-ATPase proteolipid subunit C-like domain	Domain	GO:0015078
Aphis 7344	IPR002379	V-ATPase proteolipid subunit C-like domain	Domain	GO:0015991
Aphis 7344	IPR005953	ATPase, F0 complex, subunit C, bacterial/chloroplast	Family	TIGR01260 (TIGRFAMs)
Aphis 7344	IPR005953	ATPase, F0 complex, subunit C, bacterial/chloroplast	Family	GO:0015078
Aphis 7344	IPR005953	ATPase, F0 complex, subunit C, bacterial/chloroplast	Family	GO:0015986
Aphis 7344	IPR020537	ATPase, F0 complex, subunit C, DCCD-binding site	Binding_site	PS00605 (PROSITE)
Aphis 7344	IPR020537	ATPase, F0 complex, subunit C, DCCD-binding site	Binding_site	

APPENDIX H. (cont.)

Aphis 7344	noIPR	Unintegrated	unintegrated	SignalP-NN(euk) (SIGNALP)
Aphis 7344	noIPR	Unintegrated	unintegrated	tmhmm (TMHMM)
Aphis 7344	noIPR	Unintegrated	unintegrated	

APPENDIX I

Summary of all SNPs detected in contigs including sequence description, length, organism, minimum e-value, number of GOs, and number of SNPs associated with each contig.

APPENDIX I1. *A. curvipes.*

Contig ID	Sequence Description	Length (bp)	Organism	Minimum E-value	#GOs	#SNPs
Anop 7	elongation factor 1-alpha	2152	*Riptortus pedestris*	0	6	1
Anop 13	atp synthase subunit mitochondrial-like	2165	*Toxoptera citricida*	0	6	1
Anop 14	ankyrin repeat protein	1054	*Candidatus Amoebophilus asiaticus*	5.64E-58	6	4
Anop 15	ankyrin repeat protein	874	*Candidatus Amoebophilus asiaticus*	4.05E-47	6	10
Anop 17	cg7630 cg7630-pa	772	*Triatoma brasiliensis*	3.58E-18	0	2
Anop 23	rrna intron-encoded homing endonuclease	2738	*Oxytricha trifallax*	1.17E-55	1	1
Anop 25	cg41536 cg41536- partial	3200	*Daphnia pulex*	1.90E-57	0	7
Anop 26	trifunctional purine biosynthetic protein adenosine-3	1487	*Tribolium castaneum*	0	6	2
Anop 27	transferrin	305	*Riptortus clavatus*	9.41E-08	4	4
Anop 31	transferrin	1344	*Riptortus clavatus*	0	4	5
Anop 33	achain crystal structure of engineered northeast structural genomics consortium target	639	synthetic construct	1.17E-31	4	3

APPENDIX I1. (cont.)

Anop 34	orf16-lacz fusion protein	1781	*Heliobacterium modesticaldum Ice1*	2.91E-53	3
Anop 35	N/A	851			0
Anop 50	atp-citrate synthase	3596	*Acyrthosiphon pisum*	0	6
Anop 51	atp synthase f0 subunit 6	321	*Stictopleurus subviridis*	1.47E-38	4
Anop 52	atp synthase f0 subunit 6	315	*Stictopleurus subviridis*	5.60E-37	4
Anop 53	atpase subunit 6	310	*Riptortus pedestris*	2.66E-22	4
Anop 54	atp synthase f0 subunit 6	393	*Stictopleurus subviridis*	5.99E-47	4
Anop 55	pleiotrophin-like protein	1400	*Tribolium castaneum*	4.47E-40	2
Anop 58	basic juvenile hormone sensitive hemolymph protein	714	*Riptortus clavatus*	2.30E-119	6
Anop 64	midline fasciclin	3028	*Tribolium castaneum*	5.98E-81	3
Anop 71	arylphorin receptor	4026	*Calliphora vicina*	0	3
Anop 72	N/A	355			0
Anop 84	apolipophorins	4036	*Apis mellifera*	3.14E-102	1
Anop 87	adp atp translocase	1573	*Triatoma infestans*	0	4
Anop 92	15-hydroxyprostaglandin dehydrogenase	1077	*Acromyrmex echinatior*	3.00E-43	4
Anop 97	ribosomal protein s18	571	*Cicindela campestris*	2.94E-86	4
Anop 99	hemi_pyrap ame: full=hemiptericin	603	*Hemiptericin*	5.33E-23	3
Anop 106	cytochrome b	463	*Aeschyntelus notatus*	3.73E-61	7

APPENDIX I1. (cont.)

Anop 107	cytochrome b	849	*Hydaropsis longirostris*	3.70E-129	7
Anop 112	enolase	3401	*Dendroctonus ponderosae*	0	4
Anop 113	atp synthase subunit mitochondrial	1231	*Apis mellifera*	0	6
Anop 120	glyceraldehyde-3-phosphate dehydrogenase	400	*Maconellicoccus hirsutus*	4.12E-31	7
Anop 121	glyceraldhyde-3-phosphate partial	389	*Apolygus lucorum*	1.51E-64	3
Anop 123	glyceraldehyde 3 phosphate dehydrogenase 1	535	*Apolygus lucorum*	1.06E-97	8
Anop 125	cytochrome c oxidase subunit partial	519	*Aeschyntelus notatus*	1.65E-86	8
Anop 133	apolipophorin-iii precursor	580	*Riptortus clavatus*	1.13E-62	3
Anop 152	tropomyosin 1	1814	*Megachile rotundata*	2.51E-158	3
Anop 153	myosin heavy muscle isoform 1	6145	*Acyrthosiphon pisum*	0	42
Anop 159	N/A	358			0
Anop 161	trifunctional enzyme beta subunit (tp-beta)	832	*Tribolium castaneum*	3.07E-135	7
Anop 162	vitellogenin	4518	*Riptortus clavatus*	0	2
Anop 163	cytochrome c oxidase subunit iii	1109	*Stictopleurus subviridis*	5.33E-55	4
Anop 164	cytochrome c oxidase subunit iii	367	*Stictopleurus subviridis*	1.94E-62	4
Anop 165	proactivator polypeptide	2113	*Tribolium castaneum*	1.32E-137	4

APPENDIX I1. (cont.)

Anop 169	lipoyltransferase mitochondrial-like	3307	*Bombus terrestris*	4.31E-122	2
Anop 175	fatty acid synthase	7421	*Nasonia vitripennis*	0	6
Anop 184	nadh dehydrogenase subunit 5	604	*Stictopleurus subviridis*	6.57E-35	4
Anop 187	hexamerin 1	784	*Riptortus clavatus*	8.69E-113	6
Anop 188	vitellogenin	2422	*Riptortus clavatus*	7.45E-168	4
Anop 198	ribosomal protein l9	709	*Tribolium castaneum*	5.21E-119	4
Anop 205	actin	1506	*Drosophila melanogaster*	0	13
Anop 213	ribosomal protein s25	413	*Diaphorina citri*	1.57E-40	3
Anop 214	probable bax inhibitor 1-like	1533	*Triatoma infestans*	2.98E-111	1
Anop 215	N/A	295			0
Anop 236	heat shock protein 70	776	*Pyrrhocoris apterus*	6.14E-154	2
Anop 238	N/A	382			0
Anop 241	mitochondrial porin	1644	*Homalodisca vitripennis*	4.96E-133	11
Anop 249	cathepsin l-like	438	*Triatoma brasiliensis*	5.67E-63	1
Anop 258	ribosomal protein s9	690	*Meladema coriacea*	7.39E-119	5
Anop 261	heat shock protein 70	571	*Pyrrhocoris apterus*	9.12E-96	4
Anop 266	heat shock protein 90	1073	*Camponotus floridanus*	0	4
Anop 274	serine threonine-protein phosphatase 6 regulatory ankyrin repeat subunit a-like	290	*Strongylocentrotus purpuratus*	1.32E-15	2
Anop 275	imaginal disc growth factor	901	*Oncometopia nigricans*	3.47E-105	2
Anop 277	imaginal disc growth factor	419	*Pieris rapae*	3.05E-34	4

APPENDIX I1. (cont.)

Anop 282	cytochrome oxidase subunit 1	610	*Sethenira ferruginea*	1.40E-74	10
Anop 285	N/A	434			0
Anop 290	N/A	424			0
Anop 291	nadh dehydrogenase subunit 2	455	*Chauliops fallax*	3.37E-24	5
Anop 293	nadh dehydrogenase subunit 2	553	*Aeschyntelus notatus*	9.67E-33	5
Anop 296	N/A	673			0
Anop 300	cytochrome oxidase subunit 1	356	*Homoeocerus sp. ST-2009*	5.22E-28	10
Anop 304	probable atp-dependent rna helicase ddx17-like	2214	*Tribolium castaneum*	0	6
Anop 316	ribosomal protein p1	526	*Triatoma infestans*	7.86E-33	3
Anop 318	N/A	508			0
Anop 324	N/A	563			0
Anop 325	vitellogenin	2012	*Lethocerus deyrollei*	1.28E-124	2
Anop 332	cytochrome p450	901	*Tribolium castaneum*	3.01E-34	2
Anop 334	cytochrome oxidase subunit i	219	*Cletus punctiger*	4.87E-35	10
Anop 337	ribosomal protein l23a	1033	*Tribolium castaneum*	8.98E-79	7
Anop 344	nadh dehydrogenase subunit 2	360	*Aeschyntelus notatus*	3.83E-13	4
Anop 348	apolipophorin-iii precursor	477	*Riptortus clavatus*	3.45E-59	3
Anop 349	apolipophorin-iii precursor	472	*Riptortus clavatus*	5.47E-63	3
Anop 351	ankyrin repeat domain protein	951	*Strongylocentrotus purpuratus*	8.78E-35	3
Anop 354	hypothetical protein EAI_16042	681	*Harpegnathos saltator*	9.95E-07	0
Anop 358	nadh dehydrogenase subunit i	1020	*Aeschyntelus notatus*	1.09E-108	4
Anop 364	cytochrome b	848	*Riptortus pedestris*	8.85E-42	7

APPENDIX I1. (cont.)

Anop 365	disulfide isomerase	1998	*Litopenaeus vannamei*	0	12
Anop 376	abp2_ripcl ame: full=probable antibacterial peptide flags: precursor	945	*Riptortus clavatus*	2.01E-44	1
Anop 377	superoxide dismutase	778	*Triatoma infestans*	2.65E-82	4
Anop 380	ubiquitin	228	*Cherax quadricarinatus*	1.16E-46	61
Anop 389	N/A	518			4
Anop 392	atp synthase delta mitochondrial	566	*Culex quinquefasciatus*	1.11E-53	4
Anop 395	40s ribosomal protein s2	1113	*Tribolium castaneum*	4.75E-139	4
Anop 422	N/A	342			0
Anop 433	ribosomal protein l7a	943	*Solenopsis invicta*	2.72E-97	2
Anop 447	ornithine decarboxylase	1221	*Pediculus humanus corporis*	1.12E-42	2
Anop 467	mitochondrial phosphate carrier protein	1338	*Aedes aegypti*	7.12E-164	4
Anop 471	cytochrome oxidase subunit partial	361	*Anoplocnemis phasianus*	1.43E-51	10
Anop 473	cytochrome oxidase subunit 1	330	*Leptocorisa vericornis*	7.21E-55	10
Anop 480	ribonuclease uk114-like isoform 1	816	*Drosophila virilis*	3.86E-58	2
Anop 485	ankyrin repeat protein	524	*synthetic construct*	9.81E-25	1
Anop 494	N/A	662			0
Anop 503	AGAP010360-PA	817	*Anopheles gambiae str. PEST*	4.83E-11	3
Anop 517	N/A	244			0
Anop 520	ribosomal protein l24	544	*Bombyx mori*	9.31E-61	2
Anop 523	ribosomal protein l14	544	*Lygus lineolaris*	1.26E-68	3

APPENDIX I1. (cont.)

Anop 527	luciferin-regenerating enzyme	1169	*Nasonia vitripennis*	4.87E-70	3
Anop 534	N/A	602			0
Anop 541	isoform cra_b	403	*Oncopeltus fasciatus*	8.92E-20	7
Anop 552	acyl- -binding protein	702	*Rhodnius prolixus*	3.61E-39	2
Anop 561	i-type lysozyme	512	*Nilaparvata lugens*	3.77E-41	3
Anop 568	peroxiredoxin 1	1112	*Coptotermes formosanus*	8.53E-79	10
Anop 569	elongation factor 1 delta	753	*Graphocephala atropunctata*	6.56E-65	9
Anop 576	alpha-glucosidase	1276	*Aedes aegypti*	3.27E-90	4
Anop 582	cg31997 cg31997-pa	616	*Megachile rotundata*	2.14E-43	3
Anop 597	polyadenylate-binding protein 1-like isoform 1	2456	*Bombus terrestris*	0	4
Anop 602	luciferin-regenerating enzyme	843	*Nasonia vitripennis*	2.46E-45	3
Anop 609	cathepsin l	523	*Drosophila mojavensis*	1.20E-61	5
Anop 615	cg12324 protein	466	*Triatoma infestans*	8.81E-86	5
Anop 622	pheromone-degrading enzyme	734	*Pyrrhocoris apterus*	4.12E-50	1
Anop 643	phosphatidylethanolamine-binding protein	325	*Apis mellifera*	9.09E-11	0
Anop 648	N/A	733			0
Anop 658	N/A	374			0
Anop 662	ferritin heavy chain	1009	*Nasonia vitripennis*	2.92E-56	6
Anop 664	salivary secreted cystatin 3 precursor	574	*Oncopeltus fasciatus*	6.72E-16	5
Anop 674	odorant-binding protein	732	*Apolygus lucorum*	3.77E-17	1
Anop 678	superoxide dismutase	1406	Cu-Zn	8.83E-50	18

APPENDIX I1. (cont.)

Anop 683	N/A	350			0
Anop 708	N/A	1002			0
Anop 714	N/A	686			0
Anop 729	ubiquinol-cytochrome c reductase complex 14 kda protein	709	*Papilio xuthus*	5.40E-43	6
Anop 744	mitochondrial cytochrome c oxidase subunit 5b isoform 1	562	*Triatoma infestans*	8.47E-51	4
Anop 758	N/A	451			0
Anop 759	cytochrome c	651	*Graphocephala atropunctata*	1.71E-62	9
Anop 771	N/A	830			0
Anop 772	fructose -bisphosphate aldolase	1414	*Daphnia pulex*	0	21
Anop 782	PREDICTED: hypothetical protein LOC100648520	386	*Bombus terrestris*	7.11E-07	0
Anop 799	cytochrome c oxidase polypeptide iv	673	*Locusta migratoria*	4.36E-71	7
Anop 800	ribosomal protein l10	771	*Acyrthosiphon pisum*	1.12E-147	3
Anop 805	N/A	402			0
Anop 807	40s ribosomal protein s8-like	709	*Triatoma infestans*	1.59E-123	3
Anop 809	ribosomal protein l31	696	*Harpegnathos saltator*	1.71E-62	3
Anop 818	glutamine synthetase 2	867	*Nilaparvata lugens*	1.07E-173	7
Anop 825	transferrin	1095	*Riptortus clavatus*	4.20E-136	4
Anop 828	60s ribosomal protein l5	972	*Laodelphax striatella*	4.66E-168	7
Anop 829	N/A	644			0

APPENDIX I1. (cont.)

Anop 843	fk506-binding protein	638	*Daphnia pulex*	1.31E-63	6
Anop 849	40s ribosomal protein s11	667	*Maconellicoccus hirsutus*	3.68E-74	3
Anop 850	peripheral-type benzodiazepine receptor	1056	*Drosophila grimshawi*	3.22E-40	3
Anop 857	N/A	731			0
Anop 877	suppressor of g2 allele of skp1 homolog	1438	*Gorilla gorilla gorilla*	8.00E-47	6
Anop 878	inter-alpha-trypsin inhibitor heavy chain h4 precursor	2250	*Acyrthosiphon pisum*	9.23E-130	2
Anop 886	pancreatic triacylglycerol lipase	1779	*Acyrthosiphon pisum*	0	1
Anop 902	cathepsin b-like proteinase	1396	*Daphnia pulex*	1.62E-157	3
Anop 906	heat shock protein 70	2713	*Lycorma delicatula*	0	5
Anop 908	guanine nucleotide-binding protein subunit beta-like	1042	*Blattella germanica*	0	2
Anop 911	nadh dehydrogenase subunit 4	2548	*Riptortus pedestris*	2.50E-119	5
Anop 958	arginine kinase	1742	*Anasa tristis*	0	3
Anop 967	N/A	323			0
Anop 971	N/A	818			0
Anop 982	N/A	558			0
Anop 1000	thiamin pyrophosphokinase 1	1131	*Tribolium castaneum*	1.82E-55	1
Anop 1017	gelsolin precursor	2022	*Culex quinquefasciatus*	1.14E-175	2
Anop 1027	N/A	636			0
Anop 1050	phosphoenolpyruvate isoform a	2429	*Tribolium castaneum*	0	4
Anop 1084	pneumolysin	567	*Streptococcus mitis SK597*	1.69E-19	10
Anop 1085	serine rich protein	2575	*Nematostella vectensis*	5.49E-54	5

APPENDIX I1. (cont.)

Anop 1109	unknown	1096	*Lygus lineolaris*	3.06E-10	0
Anop 1125	eukaryotic translation initiation factor x-chromosomal-like	999	*Triatoma infestans*	2.40E-78	7
Anop 1126	arylphorin subunit a4	708	*Calliphora vicina*	1.25E-140	2
Anop 1136	N/A	558			0
Anop 1163	sugar phosphate exchanger 2-like isoform 2	2076	*Megachile rotundata*	3.04E-179	3
Anop 1166	N/A	407			0
Anop 1170	cathepsin d	1027	*Callosobruchus maculatus*	8.02E-97	1
Anop 1173	N/A	363			0
Anop 1189	vitellogenin	624	*Riptortus clavatus*	5.41E-65	2
Anop 1193	phosphoserine aminotransferase	1277	*Pediculus humanus corporis*	8.67E-168	4
Anop 1201	apolipoprotein d-like	919	*Nasonia vitripennis*	1.28E-95	1
Anop 1202	chemosensory protein 1	414	*Apolygus lucorum*	2.14E-40	0
Anop 1207	cathepsin b	1027	*Branchiostoma floridae*	3.92E-107	3
Anop 1217	ribosomal protein l13	787	*Xenopsylla cheopis*	8.21E-86	3
Anop 1229	cytosolic malate dehydrogenase	1484	*Pediculus humanus corporis*	2.23E-164	8
Anop 1233	citrate synthase	1848	*Aedes aegypti*	0	5
Anop 1236	ribosomal protein s15e	445	*Diaphorina citri*	1.33E-66	6
Anop 1237	N/A	650			0
Anop 1243	N/A	703			0
Anop 1268	translationally controlled tumor protein	866	*Graphocephala atropunctata*	2.33E-103	1

APPENDIX I1. (cont.)

Anop 1269	atp synthase-like protein	1078	*Papilio xuthus*	2.13E-86	8
Anop 1276	N/A	1450			0
Anop 1277	N/A	1292			1
Anop 1294	ribosomal protein l28	506	*Triatoma brasiliensis*	2.64E-74	3
Anop 1295	ankyrin repeat protein	954	*Synechococcus sp. JA-3-3Ab*	7.39E-25	0
Anop 1302	cg31997 cg31997-pa	540	*Acyrthosiphon pisum*	7.78E-47	1
Anop 1312	N/A	727			0
Anop 1349	maltase a3	1179	*Nilaparvata lugens*	8.19E-96	1
Anop 1357	N/A	843			0
Anop 1361	N/A	938			0
Anop 1375	N/A	1357			0
Anop 1384	N/A	493			0
Anop 1411	vitellogenin	1922	*Riptortus clavatus*	0	2
Anop 1426	nadh dehydrogenase subunit 5	295	*Hydaropsis longirostris*	1.02E-49	5
Anop 1518	nadh dehydrogenase subunit 5	962	*Hydaropsis longirostris*	6.06E-76	5
Anop 1524	elongation factor 1 beta	772	*Triatoma infestans*	1.17E-108	3
Anop 1526	N/A	450			0
Anop 1550	N/A	441			0
Anop 1553	N/A	939			0
Anop 1563	ribosomal protein l3	1276	*Bombus terrestris*	0	3
Anop 1627	small heat shock protein	823	*Maconellicoccus hirsutus*	2.50E-46	1
Anop 1666	chemosensory protein 1	558	*Apolygus lucorum*	3.24E-51	0

APPENDIX I1. (cont.)

Anop 1686	odorant binding protein 2	843	*Apolygus lucorum*	6.23E-11	0
Anop 1692	cellular retinaldehyde-binding protein	1260	*Tribolium castaneum*	4.14E-120	3
Anop 1703	N/A	467			0
Anop 1766	phosphoglycerate mutase	991	*Acyrthosiphon pisum*	1.66E-140	2
Anop 1776	cytochrome p450 4g15	1857	*Acromyrmex echinatior*	0	6
Anop 1794	atp synthase-coupling factor mitochondrial	809	*Aedes albopictus*	2.43E-33	4
Anop 1815	pyrimidine-specific ribonucleoside hydrolase riha-like	1189	*Bombus impatiens*	4.91E-65	1
Anop 1823	lyzozyme m1	844	Wolbachia endosymbiont of *Drosophila simulans* wNo	5.29E-74	2
Anop 1835	acyl-protein thioesterase	1828	*Camponotus floridanus*	2.65E-96	3
Anop 1842	40s ribosomal protein s3a	879	*Triatoma infestans*	4.32E-164	3
Anop 1844	serine proteinase stubble	809	*Megachile rotundata*	1.69E-43	2
Anop 1855	chkov1	468	*Drosophila persimilis*	1.80E-12	4
Anop 1856	N/A	849			0
Anop 1896	aminopeptidase -like	1463	*Apis mellifera*	2.74E-177	2
Anop 1909	N/A	430			0
Anop 1945	heat shock protein 60	2658	*Apis mellifera*	0	13
Anop 1989	catalase	1763	*Schistocerca gregaria*	0	0
Anop 2003	heat shock protein	963	*Schistocerca gregaria*	1.63E-53	2
Anop 2142	ribosomal protein l7	912	*Papilio polytes*	1.52E-103	2

APPENDIX I1. (cont.)

Anop 2149	heat shock protein 90	419	*Apis mellifera*	1.70E-70	4
Anop 2178	thiol-activated cytolysin	1072	*Streptococcus mitis*	2.69E-12	1
Anop 2212	glutathione s-transferase	417	*Aphis gossypii*	5.51E-18	1
Anop 2312	hemi_pyrap ame: full=hemiptericin	395	*Pyrrhocoris apterus*	8.00E-24	3
Anop 2378	brain protein 44-like	450	*Ixodes scapularis*	1.30E-34	3
Anop 2385	N/A	302			0
Anop 2440	ribosomal protein l11	679	*Triatoma infestans*	6.41E-120	7
Anop 2477	puromycin-sensitive aminopeptidase	2848	*Acromyrmex echinatior*	1.90E-68	0
Anop 2534	cytochrome p450	1594	*Culex quinquefasciatus*	2.25E-45	8
Anop 2559	cytochrome b-c1 complex subunit mitochondrial-like	1535	*Tribolium castaneum*	5.04E-67	6
Anop 2571	AGAP001981-PB	818	*Anopheles gambiae str. PEST*	1.85E-09	0
Anop 2593	N/A	513			0
Anop 2635	cathepsin l	868	*Litopenaeus vannamei*	3.20E-87	1
Anop 2636	fatty acid binding protein	529	*Lygus lineolaris*	1.02E-59	3
Anop 2804	N/A	468			0
Anop 2828	N/A	673			0
Anop 2833	GJ21142	344	*Drosophila virilis*	4.70E-23	0
Anop 2861	dihydrolipoamide dehydrogenase e3 subunit	1732	*Pediculus humanus corporis*	0	5
Anop 2926	kda midgut protein	332	*Lygus lineolaris*	6.26E-11	0
Anop 3353	probable maltase l-like	376	*Nilaparvata lugens*	6.25E-18	1

APPENDIX I1. (cont.)

Anop 3481	small heat shock protein	946	*Lygus hesperus*	2.82E-65	2	3
Anop 3688	N/A	321			0	6
Anop 3823	3-hydroxyacyl-coa dehyrogenase	1100	*Papilio xuthus*	6.73E-134	4	2
Anop 4243	N/A	339			0	1
Anop 5253	N/A	411			0	3
Anop 6789	N/A	659			0	3
Anop 8555	N/A	534			0	1

APPENDIX I2. *A. craccivora.*

Contig ID	Sequence Description	Length (bp)	Organism	Minimum E-value	#GOs	#SNPs
Aphis 41	adp atp translocase	1789	*Acyrthosiphon pisum*	0	4	2
Aphis 45	enolase	1701	*Acyrthosiphon pisum*	0	5	1
Aphis 85	mitochondrial atp synthase f chain	877	*Acyrthosiphon pisum*	1.73E-66	5	2
Aphis 90	60s acidic ribosomal protein p2	454	*Acyrthosiphon pisum*	1.16E-33	3	3
Aphis 102	60s ribosomal protein l11-like	682	*Acyrthosiphon pisum*	1.86E-133	7	6
Aphis 103	ribosomal protein s2	968	*Acyrthosiphon pisum*	8.25E-154	5	2
Aphis 111	myosin light chain 2	1283	*Acyrthosiphon pisum*	1.41E-91	2	1
Aphis 117	acypi000079	453	*Toxoptera citricida*	2.55E-65	8	2
Aphis 133	ribosomal protein l10	762	*Acyrthosiphon pisum*	6.87E-159	3	2
Aphis 200	PREDICTED: hypothetical protein LOC100169357 isoform 1	1943	*Acyrthosiphon pisum*	4.44E-68	0	3
Aphis 212	muscle actin	1670	*Acyrthosiphon pisum*	0	3	1
Aphis 215	elongation factor 1 alpha	2548	*Acyrthosiphon pisum*	0	6	2
Aphis 242	elongation factor 2	2425	*Toxoptera citricida*	0	5	1
Aphis 255	zinc finger protein 512b-like	917	*Acyrthosiphon pisum*	2.55E-49	0	11

APPENDIX I2. (cont.)

Aphis 273	40s ribosomal protein s8-like	972	*Hordeum vulgare* subsp. *vulgare*	1.37E-131	3	4
Aphis 309	ribosomal protein s16	743	*Acyrthosiphon pisum*	2.57E-91	3	1
Aphis 321	h+ transporting atp synthase subunit g	745	*Acyrthosiphon pisum*	1.46E-62	4	9
Aphis 359	malate cytoplasmic-like	2642	*Acyrthosiphon pisum*	0	8	10
Aphis 365	cytochrome oxidase subunit i	2550	*Aphis nerii*	6.23E-154	10	7
Aphis 384	N/A	442			0	13
Aphis 433	ribosomal protein s13	811	*Acyrthosiphon pisum*	2.32E-101	4	2
Aphis 480	ribosomal protein l21	585	*Acyrthosiphon pisum*	2.50E-93	3	1
Aphis 525	60s ribosomal protein l4-like	1633	*Hordeum vulgare* subsp. *vulgare*	0	3	1
Aphis 645	cyclophilin 1	935	*Acyrthosiphon pisum*	1.44E-135	9	2
Aphis 668	ribosomal protein s7	723	*Acyrthosiphon pisum*	3.81E-136	3	1
Aphis 704	ribosomal protein l18a	629	*Acyrthosiphon pisum*	1.55E-117	3	3
Aphis 906	PREDICTED: hypothetical protein LOC100570527	393	*Acyrthosiphon pisum*	1.92E-36	0	1
Aphis 1221	cg12324 protein	831	*Acyrthosiphon pisum*	5.08E-87	5	1
Aphis 1468	isoform a	495	*Acyrthosiphon pisum*	2.32E-36	0	1

APPENDIX I2. (cont.)

Aphis 1725	tpa: cuticle protein	746	*Acyrthosiphon pisum*	4.30E-59	1	1

APPENDIX I3. ***C. tomentosicollis.***

Contig ID	Sequence Description	Length (bp)	Organism	Minimum E-value	#GOs	#SNPs
Clavig 2	hexamerin 1	1165	*Riptortus clavatus*	0	6	2
Clavig 6	N/A	407			0	4
Clavig 7	serine rich protein	389	*Oncopeltus fasciatus*	2.35E-10	0	6
Clavig 9	N/A	439			0	6
Clavig 17	tropomyosin 1	1786	*Lethocerus indicus*	3.78E-158	0	1
Clavig 18	fk506-binding protein	445	*Daphnia pulex*	1.12E-65	3	1
Clavig 19	fk506-binding protein	412	*Daphnia pulex*	8.07E-48	3	2
Clavig 20	rrna intron-encoded homing endonuclease	4211	*Oxytricha trifallax*	2.21E-55	1	2
Clavig 26	N/A	717			0	1
Clavig 37	cathepsin l1	319	*Ornithorhynchus anatinus*	7.19E-43	4	2
Clavig 39	serine rich protein	393	*Oncopeltus fasciatus*	1.50E-07	0	2
Clavig 45	cathepsin l	310	*Nematostella vectensis*	1.19E-48	6	6
Clavig 46	ribosomal protein s6	808	*Pediculus humanus corporis*	4.56E-136	3	3
Clavig 50	apolipophorin-iii precursor	464	*Riptortus clavatus*	4.79E-48	3	2
Clavig 51	h+ transporting atp synthase subunit d	809	*Papilio polytes*	4.82E-71	10	3
Clavig 53	N/A	437			0	1
Clavig 66	N/A	491			0	2
Clavig 67	heat shock protein 90	1349	*Lygus hesperus*	4.57E-177	4	3
Clavig 69	N/A	1089			0	8
Clavig 70	gluten hydrolyzing proteinase	668	*Triatoma brasiliensis*	4.11E-35	3	1
Clavig 75	elongation factor 1-alpha	1607	*Riptortus pedestris*	0	7	3
Clavig 76	elongation factor 1 alpha	1434	*Riptortus pedestris*	0	7	2

APPENDIX I3. (cont.)

Clavig 81	nadh dehydrogenase subunit 1	1531	*Stictopleurus subviridis*	8.91E-24	4	1
Clavig 85	ribosomal protein s4e	890	*Lygus lineolaris*	1.66E-179	4	5
Clavig 89	lipoyltransferase mitochondrial-like	2480	*Bombus terrestris*	9.43E-91	1	1
Clavig 91	odorant binding protein 24	619	*Anopheles funestus*	1.26E-07	3	1
Clavig 103	cathepsin l	689	*Aedes aegypti*	3.99E-96	3	2
Clavig 106	cathepsin l	497	*Dermacentor variabilis*	5.99E-68	3	9
Clavig 107	ankyrin repeat protein	1315	*Candidatus Amoebophilus asiaticus* 5a2	2.14E-60	3	5
Clavig 111	cathepsin l-like	483	*Apis florea*	2.55E-60	3	1
Clavig 115	hexamerin 1	2153	*Riptortus clavatus*	0	6	2
Clavig 116	hexamerin 1	2098	*Riptortus clavatus*	0	6	2
Clavig 126	gluten hydrolyzing proteinase	1035	*Lygus lineolaris*	1.61E-63	1	1
Clavig 127	N/A	246			0	1
Clavig 129	cathepsin l	948	*Triatoma brasiliensis*	3.59E-107	3	3
Clavig 131	transferrin	2166	*Riptortus clavatus*	0	4	1
Clavig 142	N/A	819			0	1
Clavig 149	N/A	351			0	1
Clavig 151	N/A	336			0	1
Clavig 158	N/A	382			0	3
Clavig 159	lysozyme	728	*Anopheles gambiae*	1.01E-46	2	1
Clavig 166	tubulin beta-1 chain	1550	*Manduca sexta*	0	13	3
Clavig 167	N/A	1319			0	5
Clavig 168	unknown	1008	*Lygus lineolaris*	9.72E-13	0	4
Clavig 172	atp synthase subunit mitochondrial-like	1949	*Tribolium castaneum*	0	9	3
Clavig 174	N/A	1534			0	2

APPENDIX I3. (cont.)

Clavig 176	heat shock protein 70	3830	*Riftia pachyptila*	0	37	1
Clavig 177	protein disulfide isomerase	1026	*Anopheles gambiae* str. PEST	1.90E-138	14	2
Clavig 178	peritrophic matrix protein 1-b precursor	430	*Anopheles gambiae* str. PEST	8.15E-08	3	5
Clavig 180	peritrophic matrix protein 1-b precursor	384	*Anopheles gambiae* str. PEST	7.20E-08	3	1
Clavig 187	cuticle protein 34	426	*Dendroctonus ponderosae*	1.36E-24	1	5
Clavig 191	60s ribosomal protein l4-like	1360	*Bombus impatiens*	0	3	2
Clavig 192	probable maltase l-like	665	*Nilaparvata lugens*	2.79E-26	2	2
Clavig 196	ribosomal protein l31	466	*Acromyrmex echinatior*	1.00E-53	3	1
Clavig 201	apolipophorins	3607	*Nilaparvata lugens*	3.75E-121	14	2
Clavig 206	cg31997 cg31997-pa	618	*Megachile rotundata*	8.50E-44	3	1
Clavig 214	myosin heavy muscle isoform 1	6145	*Acyrthosiphon pisum*	0	25	8
Clavig 215	ribosomal protein l13	760	*Xenopsylla cheopis*	1.68E-84	3	1
Clavig 227	N/A	493			0	5
Clavig 238	N/A	613			0	1
Clavig 239	alpha-amylase	1615	*Blattella germanica*	3.91E-178	3	9
Clavig 242	maltase a3	775	*Nilaparvata lugens*	2.63E-41	3	1
Clavig 243	N/A	1804			0	2
Clavig 247	N/A	301			0	1
Clavig 251	serine rich protein	488	*Oncopeltus fasciatus*	4.72E-19	0	4
Clavig 256	atp-citrate synthase-like	3916	*Acyrthosiphon pisum*	0	10	5
Clavig 261	ribosomal protein l14	562	*Lygus lineolaris*	5.10E-68	3	1
Clavig 266	ribosomal protein l15	739	*Pediculus humanus corporis*	6.99E-118	3	1
Clavig 284	fatty acid synthase	3168	*Nasonia vitripennis*	0	5	1

APPENDIX I3. (cont.)

Clavig 293	serine rich protein	1438	*Oncopeltus fasciatus*	8.70E-24	0	2
Clavig 296	alpha amylase catalytic region	439	*Fusarium oxysporum* Fo5176	2.59E-45	3	12
Clavig 297	alpha partial	492	*Acromyrmex echinatior*	1.73E-42	3	1
Clavig 298	probable maltase h-like	577	*Tribolium castaneum*	1.99E-71	3	5
Clavig 300	maltase 1	2668	*Nilaparvata lugens*	6.02E-144	3	2
Clavig 301	N/A	335			0	8
Clavig 318	N/A	719			0	3
Clavig 320	ribosomal protein l23a	1016	*Tribolium castaneum*	2.30E-82	7	2
Clavig 322	ribosomal protein l7	894	*Papilio polytes*	3.66E-127	1	3
Clavig 332	N/A	511			0	7
Clavig 345	cytochrome c	647	*Graphocephala atropunctata*	1.65E-62	5	1
Clavig 349	atp synthase-coupling factor mitochondrial	482	*Tribolium castaneum*	3.73E-36	10	2
Clavig 356	60s acidic ribosomal protein p2-like protein	478	*Bombyx mori*	3.88E-25	4	1
Clavig 366	probable phosphoserine aminotransferase-like	1161	*Harpegnathos saltator*	3.28E-180	4	1
Clavig 372	conserved hypothetical protein	432	*Culex quinquefasciatus*	4.59E-13	3	1
Clavig 378	N/A	334			0	3
Clavig 384	cathepsin l-like	656	*Artemia salina*	1.29E-80	3	1
Clavig 386	N/A	988			0	4
Clavig 398	N/A	849			0	2
Clavig 399	accessory gland protein	1512	*Gryllus firmus*	5.08E-10	0	2

APPENDIX I3. (cont.)

Clavig 406	abp2_ripcl ame: full=probable antibacterial peptide flags: precursor	444	*Riptortus clavatus*	2.64E-34	3	4
Clavig 408	N/A	1245			0	2
Clavig 410	cytochrome c oxidase subunit iii	2075	*Hydaropsis longirostris*	1.12E-128	5	2
Clavig 426	aminopeptidase n-like	3964	*Strongylocentrotus purpuratus*	7.23E-39	4	1
Clavig 433	50 kda midgut protein	737	*Nasonia vitripennis*	1.24E-20	0	1
Clavig 456	AGAP004851-PA	376	*Anopheles gambiae* str. PEST	3.62E-10	3	2
Clavig 466	heat shock protein 70	2783	*Culex quinquefasciatus*	0	6	9
Clavig 475	vitellogenin	435	*Riptortus clavatus*	1.59E-44	6	1
Clavig 484	N/A	365			0	2
Clavig 487	cathepsin l-like	432	*Camponotus floridanus*	9.24E-58	3	1
Clavig 490	N/A	424			0	3
Clavig 497	N/A	596			0	2
Clavig 501	vitellogenin	3404	*Riptortus clavatus*	0	2	8
Clavig 509	odorant-binding protein partial	371	*Rhodnius prolixus*	3.25E-36	1	2
Clavig 513	cytochrome b	625	*Hydaropsis longirostris*	2.71E-96	6	1
Clavig 518	cytochrome b	626	*Hydaropsis longirostris*	1.43E-92	6	2
Clavig 524	odorant-binding protein partial	586	*Adelphocoris lineolatus*	4.22E-14	1	3
Clavig 533	actin	1748	*Drosophila melanogaster*	0	15	5
Clavig 544	odorant binding protein 19d	698	*Apolygus lucorum*	1.17E-15	1	4
Clavig 564	N/A	377			1	1
Clavig 576	peripheral-type benzodiazepine receptor	1114	*Drosophila pseudoobscura pseudoobscura*	3.85E-36	4	1

APPENDIX I3. (cont.)

Clavig 587	isoform a	597	*Drosophila willistoni*	7.77E-18	2	1
Clavig 590	N/A	406			0	5
Clavig 591	hypothetical protein DAPPUDRAFT_70492	597	*Daphnia pulex*	2.39E-15	1	1
Clavig 605	N/A	761			0	1
Clavig 606	af373879_1peritrophin-like protein 1	1921	*Anopheles gambiae* str. PEST	1.53E-12	3	1
Clavig 607	serine rich protein	417	*Oncopeltus fasciatus*	4.33E-20	0	23
Clavig 614	ornithine decarboxylase	1037	*Pediculus humanus corporis*	8.89E-44	2	2
Clavig 615	hypothetical protein AaeL_AAEL015254	1063	*Aedes aegypti*	6.46E-12	3	2
Clavig 625	nadh dehydrogenase subunit i	827	*Aeschyntelus notatus*	2.17E-78	4	3
Clavig 626	fructose -bisphosphate aldolase	1457	*Daphnia pulex*	1.97E-157	19	11
Clavig 640	gamma-interferon-inducible lysosomal thiol reductase-like	1580	*Maconellicoccus hirsutus*	3.11E-23	0	1
Clavig 643	acetyl- mitochondrial	1139	*Acyrthosiphon pisum*	0	14	4
Clavig 670	gluten hydrolyzing proteinase	617	*Triatoma brasiliensis*	2.16E-37	3	2
Clavig 686	ankyrin repeat protein	1050	*Oncopeltus fasciatus*	2.88E-49	0	5
Clavig 692	beta-tubulin	1737	*Manduca sexta*	0	13	5
Clavig 711	nucleoside diphosphate kinase	787	*Aedes aegypti*	1.19E-80	15	1
Clavig 716	elongation factor 1-gamma	1340	*Maconellicoccus hirsutus*	2.52E-137	13	1
Clavig 726	N/A	523			0	3
Clavig 745	N/A	304			0	1
Clavig 757	N/A	521			0	1
Clavig 765	ankyrin repeat protein	1330	*Trichomonas vaginalis* G3	1.41E-39	1	1

APPENDIX I3. (cont.)

Clavig 768	counting factor associated protein d-like	1722	*Periplaneta americana*	0	6	4
Clavig 775	apolipoprotein d	1114	*Pediculus humanus corporis*	4.88E-85	1	2
Clavig 784	N/A	645			0	1
Clavig 799	N/A	654			0	2
Clavig 817	serine rich protein	1196	*Oncopeltus fasciatus*	4.24E-19	0	1
Clavig 858	ribosomal protein l17	596	*Pediculus humanus corporis*	3.37E-97	3	1
Clavig 861	ankyrin repeat protein	913	*Diplorickettsia massiliensis* 20B	1.06E-17	1	2
Clavig 869	endocuticle structural glycoprotein bd-1	2627	*Anopheles gambiae* str. PEST	8.82E-25	1	1
Clavig 891	phosphoenolpyruvate carboxykinase	2001	*Tribolium castaneum*	0	5	1
Clavig 936	spike protein	4244	*Hana virus*	5.94E-15	0	8
Clavig 939	N/A	1018			0	1
Clavig 945	heat shock 70 kda protein cognate 3	2376	*Nasonia vitripennis*	0	6	1
Clavig 953	N/A	746			0	3
Clavig 954	N/A	448			0	2
Clavig 959	ribosomal protein l10	298	*Acyrthosiphon pisum*	8.66E-48	3	1
Clavig 960	ribosomal protein l10	396	*Acyrthosiphon pisum*	1.03E-89	3	2
Clavig 964	N/A	972			0	3
Clavig 965	mitochondrial atp synthase gamma-subunit	808	*Graphocephala atropunctata*	8.95E-142	8	1
Clavig 984	mitochondrial-processing peptidase subunit beta-like	1569	*Anopheles gambiae* str. PEST	0	12	1

APPENDIX I3. (cont.)

Clavig 992	polyadenylate-binding protein 1-like isoform 1	3350	*Bombus terrestris*	0	0	1
Clavig 993	N/A	1126			0	10
Clavig 996	thiamin pyrophosphokinase 1	1356	*Tribolium castaneum*	1.65E-56	4	3
Clavig 1003	ribosomal protein l22	748	*Danaus plexippus*	1.54E-30	3	1
Clavig 1007	mdl1	592	*Acromyrmex echinatior*	3.38E-33	0	1
Clavig 1029	malic enzyme	480	*Pediculus humanus corporis*	3.40E-29	4	2
Clavig 1038	phosphate carrier mitochondrial-like	1349	*Aedes aegypti*	1.98E-158	2	1
Clavig 1049	fatty acid desaturase	1866	*Acheta domesticus*	3.52E-164	15	2
Clavig 1052	N/A	1094			0	4
Clavig 1054	ribosomal protein l26e	456	*Triatoma infestans*	1.10E-80	3	1
Clavig 1055	isoform c	1333	*Drosophila ananassae*	4.28E-25	3	8
Clavig 1058	hexamerin 1	576	*Riptortus clavatus*	1.36E-99	6	1
Clavig 1075	mitochondrial cytochrome c oxidase subunit 5b isoform 1	565	*Tribolium castaneum*	2.35E-53	6	3
Clavig 1079	serine protease	613	*Triatoma infestans*	5.38E-32	3	8
Clavig 1087	ribosomal protein s9	665	*Graphocephala atropunctata*	6.36E-123	4	1
Clavig 1088	ribosomal protein s9	641	*Graphocephala atropunctata*	4.65E-123	4	2
Clavig 1125	chitin binding peritrophin-a domain-containing partial	1714	*Drosophila ananassae*	5.52E-45	3	2
Clavig 1126	mitochondrial cytochrome c oxidase subunit 5b isoform 1	533	*Triatoma infestans*	7.91E-51	6	1
Clavig 1167	nucleoplasmin-like protein	1088	*Maconellicoccus hirsutus*	1.64E-55	1	1

APPENDIX I3. (cont.)

Clavig 1182	N/A	217			0	1
Clavig 1187	N/A	369			0	1
Clavig 1213	N/A	568			0	8
Clavig 1243	60s ribosomal protein l7a	840	*Solenopsis invicta*	1.82E-136	2	1
Clavig 1251	aminopeptidase -like	957	*Solenopsis invicta*	1.91E-105	5	1
Clavig 1261	N/A	850			0	2
Clavig 1280	N/A	976			0	1
Clavig 1286	40s ribosomal protein s16	591	*Apis florea*	3.29E-91	3	1
Clavig 1292	cathepsin l precursor	365	*Triatoma brasiliensis*	1.08E-18	1	1
Clavig 1319	ribosomal protein s12	422	*Manduca sexta*	5.41E-61	3	1
Clavig 1321	ribosomal protein l27e	514	*Hister sp.* APV-2005	1.61E-54	3	1
Clavig 1347	ribosomal protein l13a	650	*Bombus terrestris*	2.78E-113	3	1
Clavig 1350	imaginal disc growth factor	1070	*Oncometopia nigricans*	1.76E-133	3	2
Clavig 1364	15-hydroxyprostaglandin dehydrogenase	1045	*Acromyrmex echinatior*	1.04E-53	2	5
Clavig 1401	atp synthase subunit mitochondrial	2063	*Megachile rotundata*	0	9	4
Clavig 1418	cg12324 protein	468	*Triatoma infestans*	2.56E-84	7	1
Clavig 1429	N/A	899			0	6
Clavig 1431	kininogen-1 isoform 2 precursor	635	*Oncopeltus fasciatus*	2.74E-19	4	6
Clavig 1501	odorant-binding protein	972	*Apolygus lucorum*	5.66E-16	1	1
Clavig 1509	nadh dehydrogenase subunit 5	1560	*Hydaropsis longirostris*	3.27E-160	4	1
Clavig 1522	atp-dependent rna helicase-like protein	797	*Trypanosoma brucei gambiense* DAL972	1.78E-22	4	3
Clavig 1573	guanine nucleotide-binding protein subunit beta-like	1073	*Blattella germanica*	0	2	1

APPENDIX I3. (cont.)

Clavig 1605	N/A	358			0	2
Clavig 1606	N/A	565			0	1
Clavig 1626	N/A	581			0	1
Clavig 1697	nadh dehydrogenase subunit 4	455	*Riptortus pedestris*	3.18E-34	4	1
Clavig 1708	p8 nuclear protein	860	*Amblyomma variegatum*	1.83E-26	2	1
Clavig 1724	N/A	396			0	2
Clavig 1772	calmodulin	1707	*Drosophila melanogaster*	3.64E-96	35	1
Clavig 1782	hypothetical protein	558	*Triatoma brasiliensis*	2.02E-12	0	1
Clavig 1785	midline fasciclin	1385	*Tribolium castaneum*	3.08E-58	1	3
Clavig 1842	af414430_1trypsin precursor	601	*Lygus lineolaris*	4.91E-48	3	1
Clavig 1849	cathepsin l	1503	*Triatoma brasiliensis*	9.22E-133	3	5
Clavig 1913	lyzozyme m1	832	Wolbachia endosymbiont of *Drosophila ananassae*	2.41E-76	2	5
Clavig 1979	translocon-associated protein subunit beta	1392	*Maconellicoccus hirsutus*	5.91E-68	9	1
Clavig 1983	endocuticle structural glycoprotein bd-	637	*Acyrthosiphon pisum*	8.20E-22	1	1
Clavig 2008	cathepsin l	2153	*Harpegnathos saltator*	1.27E-109	3	1
Clavig 2035	serine protease	674	*Ranatra unicolor*	7.66E-45	3	2
Clavig 2045	ribosomal protein l30	426	*Daphnia pulex*	1.52E-63	3	1
Clavig 2054	15-hydroxyprostaglandin dehydrogenase	1011	*Acromyrmex echinatior*	1.69E-43	2	13
Clavig 2077	cathepsin d	1339	*Triatoma infestans*	8.43E-100	3	1
Clavig 2094	60s ribosomal protein l18	627	*Camponotus floridanus*	2.24E-103	3	2
Clavig 2104	sorbitol dehydrogenase	1123	*Pyrrhocoris apterus*	0	4	1

APPENDIX I3. (cont.)

Clavig 2132	N/A	262			0	11
Clavig 2209	N/A	338			0	1
Clavig 2326	elongation factor 1 delta	825	*Graphocephala atropunctata*	9.39E-61	8	4
Clavig 2422	salivary secreted peptide	491	*Lygus lineolaris*	2.08E-15	0	3
Clavig 2438	serine 3-dehydrogenase	861	*Triatoma infestans*	2.43E-47	2	3
Clavig 2677	cytochrome c oxidase polypeptide iv	624	*Locusta migratoria*	2.37E-71	8	1
Clavig 2823	10 kda heat shock mitochondrial-like	819	*Lygus hesperus*	7.81E-55	4	3
Clavig 2828	ubiquinol-cytochrome c reductase complex core protein	1515	*Tribolium castaneum*	6.13E-98	6	1
Clavig 2851	N/A	583			0	2
Clavig 2856	protein 5nuc-like	2073	*Camponotus floridanus*	5.06E-121	3	1
Clavig 2888	PREDICTED: uncharacterized protein LOC100863228	1252	*Apis florea*	1.48E-37	0	7
Clavig 3671	voltage-dependent anion-selective channel protein 2	557	*Homalodisca vitripennis*	1.70E-78	7	1
Clavig 4404	v-type proton atpase subunit g-like	475	*Acyrthosiphon pisum*	2.19E-27	5	2

APPENDIX I4. ***M. sjostedti.***

Contig ID	Sequence Description	Length (bp)	Organism	Minimum E-value	#GOs	#SNPs
Megal 1	elongation factor 2	1780	*Schistocerca gregaria*	0	6	1
Megal 11	N/A	1006			2	9
Megal 13	vitellogenin	4590	*Trigonotylus caelestialium*	0	2	31
Megal 25	storage protein 1	2148	*Chilo suppressalis*	0	3	5
Megal 34	vitellogenin	2855	*Trigonotylus caelestialium*	0	2	2
Megal 47	arylphorin-type storage protein	462	*Omphisa fuscidentalis*	2.77E-71	3	1
Megal 49	mitochondrial aldehyde dehydrogenase	624	*Danaus plexippus*	8.33E-107	4	3
Megal 68	ribosomal protein l27ae	517	*Camponotus floridanus*	5.94E-64	3	3
Megal 72	ribosomal protein l4	1323	*Biphyllus lunatus*	0	3	2
Megal 76	60s ribosomal protein l5-like	1257	*Helianthus annuus*	6.70E-165	5	1
Megal 85	arylphorin precursor	1485	*Omphisa fuscidentalis*	0	1	2
Megal 92	ribosomal protein s12	543	*Apis florea*	2.42E-74	3	2
Megal 107	vitellogenin	379	*Lethocerus deyrollei*	7.45E-33	4	2
Megal 143	N/A	423			0	3
Megal 157	N/A	987			0	1
Megal 183	actin	1612	*Ornithodoros moubata*	0	13	1
Megal 191	heat shock protein 90	2413	*Megachile rotundata*	0	4	4
Megal 193	40s ribosomal protein s15	545	*Diaphorina citri*	1.09E-81	6	1
Megal 199	mitochondrial atp synthase coupling factor 6	537	*Tribolium castaneum*	4.06E-25	5	2

APPENDIX I4. (cont.)

Megal 202	N/A	1411			0	4
Megal 215	60s ribosomal protein l12	865	*Drosophila ananassae*	2.02E-98	5	1
Megal 236	hexamerin 2 beta	2534	*Helicoverpa armigera*	0	3	10
Megal 238	vitellogenin	974	*Lethocerus deyrollei*	4.17E-80	4	7
Megal 312	ribosomal protein s7	650	*Carabus granulatus*	2.00E-113	3	2
Megal 346	cuticle protein 1	349	*Lonomia obliqua*	1.10E-35	1	3
Megal 360	40s ribosomal protein s14	569	*Dascillus cervinus*	8.78E-73	4	2
Megal 362	cytochrome c oxidase subunit iii	398	*Frankliniella intonsa*	1.84E-41	4	1
Megal 370	heat shock protein 70	1555	*Frankliniella occidentalis*	0	3	2
Megal 376	troponin i	871	*Loxostege sticticalis*	1.48E-85	1	1
Megal 378	ribosomal protein s8	713	*Megachile rotundata*	1.45E-124	3	1
Megal 408	N/A	216			0	1
Megal 419	ribosomal protein l35	512	*Chrysomela tremula*	1.03E-50	3	1
Megal 471	atp synthase-like protein	1067	*Culex quinquefasciatus*	9.70E-64	2	1
Megal 475	actin	688	*Diaphorina citri*	1.76E-159	3	1
Megal 491	ribosomal protein s3	781	*Scarabaeus laticollis*	4.38E-156	16	2
Megal 533	N/A	662			1	2
Megal 537	partial	721	*Trigonotylus caelestialium*	3.39E-103	2	3
Megal 551	ribosomal protein s28	408	*Biphyllus lunatus*	1.09E-27	4	1
Megal 569	N/A	806			1	2
Megal 572	N/A	475			0	2
Megal 591	tubulin alpha-1 chain	1518	*Pediculus humanus corporis*	0	12	1
Megal 603	ribosomal protein s18	523	*Cicindela campestris*	2.71E-74	4	1
Megal 608	N/A	533			0	4

APPENDIX I4. (cont.)

Megal 624	60s ribosomal protein l23a-like	716	*Tribolium castaneum*	3.42E-73	7	2
Megal 656	endocuticle structural glycoprotein bd-8-like	784	*Papilio xuthus*	1.42E-28	1	1
Megal 675	ribosomal protein l34	437	*Spodoptera frugiperda*	1.04E-74	3	3
Megal 716	cytochrome oxidase subunit viic	447	*Ixodes pacificus*	4.41E-09	1	1
Megal 749	apolipoprotein d-like	672	*Aedes aegypti*	9.06E-39	6	1
Megal 829	tpa: cuticle protein	374	*Papilio xuthus*	1.95E-17	1	2
Megal 836	female neotenic-specific protein 3	1303	*Trigonotylus caelestialium*	1.58E-11	2	1
Megal 861	glutathione s-transferase	633	*Choristoneura fumiferana*	2.54E-100	2	6
Megal 909	alo2_acrlo ame: full=antimicrobial peptide alo-2	338	*Acrocinus longimanus*	1.42E-14	3	9
Megal 952	ribosomal protein l21	534	*Euphydryas aurinia*	9.58E-81	3	1
Megal 1068	N/A	2803			0	1
Megal 1087	histone h2a	809	*Crassostrea gigas*	1.18E-75	5	1
Megal 1114	N/A	723			4	2
Megal 1118	pyruvate dehydrogenase	1656	*Acromyrmex echinatior*	0	4	2
Megal 1263	trypsin-like serine protease	676	*Ostrinia nubilalis*	1.55E-101	3	4
Megal 1348	elongation factor-1alpha partial	366	*Blasticotoma filiceti*	3.34E-79	5	1
Megal 1482	ribosomal protein s11	492	*Bombyx mori*	1.03E-92	3	2
Megal 1626	60s acidic ribosomal protein p0	650	*Blaptica dubia*	3.76E-135	5	1
Megal 1634	odorant-binding protein	638	*Danaus plexippus*	6.46E-44	1	6
Megal 1739	translocon-associated protein subunit delta	718	*Acromyrmex echinatior*	2.53E-52	2	1

APPENDIX J

List of all SNPs including consensus positions on contigs, alleles, coverage, and frequency

APPENDIX J1. *A. curvipes.*

Serial Number	Contig	Consensus Position	Consensus	Allele	Coverage	Frequency
1	Anop 7	904	C	T	365	35.3
2	Anop 13	186	C	T	70	35.7
3	Anop 14	616	C	T	36	44.4
4	Anop 14	649	T	A	73	37.0
5	Anop 14	708	C	G	85	35.3
6	Anop 14	745	T	A	58	37.9
7	Anop 15	166	T	C	135	36.3
8	Anop 15	197	A	G	138	37.0
9	Anop 15	285	G	A	193	45.6
10	Anop 15	297	A	T	174	48.3
11	Anop 15	310	A	G	207	44.0
12	Anop 15	396	A	G	229	49.3
13	Anop 15	428	C	T	239	47.3
14	Anop 15	433	T	G	244	47.5
15	Anop 15	458	T	C	231	47.6
16	Anop 15	469	A	G	212	46.7
17	Anop 17	431	C	T	118	40.7
18	Anop 17	557	A	G	105	46.7
19	Anop 23	1301	C	T	684	36.6
20	Anop 25	2181	G	A	63	39.7
21	Anop 25	2216	A	G	53	37.7
22	Anop 25	2230	A	G	64	35.9
23	Anop 25	2232	A	G	64	35.9
24	Anop 25	2244	T	C	65	36.9
25	Anop 25	2247	T	C	65	36.9
26	Anop 25	2282	G	A	70	40.0
27	Anop 26	219	G	A	41	36.6
28	Anop 26	354	C	T	45	48.9
29	Anop 27	173	C	A	96	36.5
30	Anop 27	216	A	G	87	40.2
31	Anop 27	283	G	A	88	45.5
32	Anop 27	286	G	A	87	44.8
33	Anop 31	83	A	G	114	43.9

APPENDIX J1. (cont.)

34	Anop 31	431	A	G	247	38.5
35	Anop 31	863	G	A	71	45.1
36	Anop 31	911	C	T	85	48.2
37	Anop 31	976	A	T	78	47.4
38	Anop 33	375	A	G	1013	39.1
39	Anop 33	568	T	C	73	46.6
40	Anop 33	599	G	A	61	47.5
41	Anop 34	387	T	C	956	45.7
42	Anop 34	1178	A	G	498	48.2
43	Anop 35	92	T	C	162	35.2
44	Anop 35	110	T	A	165	38.8
45	Anop 35	284	T	A	274	41.2
46	Anop 35	709	A	G	135	45.9
47	Anop 50	354	C	T	36	36.1
48	Anop 50	960	A	G	51	45.1
49	Anop 50	1048	C	A	46	41.3
50	Anop 50	1119	G	A	47	42.6
51	Anop 50	1575	G	C	63	36.5
52	Anop 50	1848	G	A	96	37.5
53	Anop 50	1893	C	G	97	47.4
54	Anop 50	1911	G	A	110	41.8
55	Anop 50	1992	C	T	91	53.8
56	Anop 50	2169	T	C	138	37.7
57	Anop 50	2607	T	C	133	42.1
58	Anop 50	2703	A	G	126	37.3
59	Anop 51	77	G	A	151	45.0
60	Anop 51	166	A	G	137	35.8
61	Anop 52	116	G	A	142	38.0
62	Anop 53	72	C	T	227	42.3
63	Anop 53	171	C	T	270	44.1
64	Anop 53	174	G	A	258	39.5
65	Anop 53	278	C	T	241	44.4
66	Anop 54	20	T	C	116	44.0
67	Anop 54	71	A	G	153	37.9
68	Anop 55	1071	C	T	49	49.0
69	Anop 58	386	C	G	77	50.6
70	Anop 64	31	C	T	46	43.5
71	Anop 64	549	A	G	101	49.5
72	Anop 64	588	A	G	69	47.8
73	Anop 64	1419	A	G	58	39.7
74	Anop 64	1491	C	T	51	37.3

APPENDIX J1. (cont.)

75	Anop 64	1507	C	T	58	41.4
76	Anop 71	345	T	A	37	37.8
77	Anop 71	383	C	T	45	40.0
78	Anop 71	390	C	G	52	44.2
79	Anop 71	414	A	G	52	48.1
80	Anop 71	555	G	A	76	47.4
81	Anop 71	696	C	T	88	44.3
82	Anop 71	765	G	A	84	47.6
83	Anop 71	795	A	G	77	46.8
84	Anop 71	1520	A	G	38	39.5
85	Anop 71	1532	G	A	37	43.2
86	Anop 71	1595	T	C	50	48.0
87	Anop 72	102	A	G	414	47.3
88	Anop 72	217	G	A	439	46.9
89	Anop 72	327	T	C	282	46.1
90	Anop 84	369	G	A	260	47.7
91	Anop 84	391	T	A	256	49.2
92	Anop 84	1613	G	A	37	40.5
93	Anop 87	97	A	G	117	37.6
94	Anop 92	446	G	A	41	39.0
95	Anop 92	620	T	C	36	47.2
96	Anop 97	109	G	T	72	40.3
97	Anop 99	176	C	T	41	36.6
98	Anop 106	152	T	C	320	35.6
99	Anop 106	386	A	G	290	35.5
100	Anop 106	419	T	C	257	38.1
101	Anop 106	445	C	T	231	44.2
102	Anop 107	299	A	G	422	35.8
103	Anop 107	452	G	A	685	49.3
104	Anop 107	488	A	G	825	37.2
105	Anop 107	507	T	C	850	35.8
106	Anop 107	596	T	C	770	35.7
107	Anop 112	29	G	C	83	38.6
108	Anop 112	101	C	T	93	50.5
109	Anop 112	428	A	C	112	40.2
110	Anop 112	458	A	G	135	48.9
111	Anop 112	563	T	C	114	49.1
112	Anop 112	794	C	T	107	37.4
113	Anop 112	965	T	C	223	44.4
114	Anop 112	1251	A	G	51	43.1
115	Anop 113	248	G	A	71	35.2

APPENDIX J1. (cont.)

116	Anop 113	581	T	G	47	36.2
117	Anop 120	339	G	A	369	42.0
118	Anop 121	200	T	A	176	44.3
119	Anop 123	344	G	A	327	41.0
120	Anop 123	416	C	T	251	49.0
121	Anop 123	506	A	G	158	46.8
122	Anop 125	237	T	C	1011	39.9
123	Anop 125	443	G	A	1378	37.7
124	Anop 133	565	T	C	76	40.8
125	Anop 152	431	A	G	43	41.9
126	Anop 152	1532	G	T	61	44.3
127	Anop 153	5795	A	T	114	48.2
128	Anop 159	339	G	C	61	41.0
129	Anop 161	291	T	C	36	38.9
130	Anop 162	378	G	C	664	47.1
131	Anop 162	1004	A	T	70	48.6
132	Anop 162	1070	A	C	61	52.5
133	Anop 162	1093	G	C	55	45.5
134	Anop 162	1172	G	T	59	45.8
135	Anop 162	1412	A	G	46	45.7
136	Anop 163	866	C	T	549	40.1
137	Anop 164	348	C	T	240	46.3
138	Anop 165	1961	T	C	35	45.7
139	Anop 169	1319	A	C	39	38.5
140	Anop 169	1535	A	G	48	50.0
141	Anop 169	1588	C	T	42	47.6
142	Anop 169	1666	C	T	51	35.3
143	Anop 169	1757	A	C	51	43.1
144	Anop 169	1971	A	T	68	41.2
145	Anop 169	2096	C	T	51	41.2
146	Anop 175	5260	C	A	49	46.9
147	Anop 175	5377	C	G	48	35.4
148	Anop 175	6334	C	T	160	42.5
149	Anop 175	6742	C	T	98	37.8
150	Anop 184	456	C	T	173	48.6
151	Anop 184	515	C	T	121	40.5
152	Anop 187	234	A	T	84	38.1
153	Anop 187	249	C	G	65	40.0
154	Anop 187	306	T	G	109	43.1
155	Anop 187	325	T	C	116	44.0
156	Anop 187	327	G	A	116	45.7

APPENDIX J1. (cont.)

157	Anop 187	368	A	G	120	43.3
158	Anop 187	477	A	T	116	37.9
159	Anop 187	479	T	A	116	37.9
160	Anop 187	498	T	C	105	39.0
161	Anop 187	513	A	G	107	39.3
162	Anop 188	317	A	G	67	52.2
163	Anop 188	347	A	G	68	35.3
164	Anop 188	665	T	C	35	48.6
165	Anop 198	468	G	T	138	44.2
166	Anop 205	843	T	C	136	47.8
167	Anop 213	69	C	T	95	40.0
168	Anop 214	1160	C	A	39	35.9
169	Anop 215	199	T	A	120	40.8
170	Anop 215	207	C	A	122	41.0
171	Anop 215	225	C	T	117	42.7
172	Anop 215	227	C	T	117	42.7
173	Anop 215	228	A	G	117	42.7
174	Anop 215	234	G	A	124	42.7
175	Anop 215	276	G	C	104	41.3
176	Anop 236	297	C	A	170	42.4
177	Anop 236	431	A	C	140	44.3
178	Anop 238	226	T	C	36	41.7
179	Anop 238	315	T	C	36	38.9
180	Anop 241	631	A	G	60	38.3
181	Anop 249	193	A	G	56	39.3
182	Anop 258	165	G	A	111	42.3
183	Anop 258	309	G	A	165	38.2
184	Anop 261	186	T	G	45	40.0
185	Anop 261	187	T	C	45	40.0
186	Anop 261	386	C	T	52	42.3
187	Anop 266	699	G	A	49	38.8
188	Anop 266	768	G	T	45	42.2
189	Anop 266	876	G	A	44	43.2
190	Anop 274	123	G	C	52	40.4
191	Anop 275	44	G	A	62	50.0
192	Anop 275	47	A	C	64	48.4
193	Anop 275	56	C	T	77	45.5
194	Anop 275	155	C	T	94	52.1
195	Anop 275	182	C	T	113	46.0

APPENDIX J1. (cont.)

196	Anop 275	190	A	G	114	37.7
197	Anop 275	786	C	T	81	46.9
198	Anop 277	362	T	C	156	38.5
199	Anop 282	245	T	C	236	46.6
200	Anop 282	357	C	T	234	36.8
201	Anop 282	591	T	C	66	45.5
202	Anop 285	88	C	G	116	37.1
203	Anop 285	195	G	C	104	37.5
204	Anop 290	85	G	C	241	43.6
205	Anop 290	148	A	G	337	48.1
206	Anop 290	325	T	C	247	42.5
207	Anop 291	62	C	T	150	44.7
208	Anop 291	191	C	T	91	41.8
209	Anop 291	316	T	C	63	42.9
210	Anop 291	333	G	A	61	36.1
211	Anop 291	379	T	C	43	41.9
212	Anop 293	245	A	G	60	43.3
213	Anop 293	418	C	T	80	51.3
214	Anop 296	415	G	A	40	35.0
215	Anop 300	85	T	C	85	43.5
216	Anop 304	1717	T	A	57	43.9
217	Anop 304	1971	T	C	40	47.5
218	Anop 316	247	C	T	396	37.1
219	Anop 318	177	G	A	76	47.4
220	Anop 318	297	C	G	78	43.6
221	Anop 324	288	C	A	83	41.0
222	Anop 324	327	T	G	86	44.2
223	Anop 325	1563	A	G	42	45.2
224	Anop 325	1575	C	T	39	38.5
225	Anop 325	1645	C	A	41	48.8
226	Anop 332	395	T	G	41	41.5
227	Anop 332	606	T	G	39	35.9
228	Anop 334	105	C	T	253	49.0
229	Anop 334	142	T	C	244	48.0
230	Anop 334	213	C	T	52	36.5
231	Anop 337	544	T	C	39	56.4
232	Anop 344	32	G	T	52	40.4

APPENDIX J1. (cont.)

233	Anop 344	339	C	T	78	47.4
234	Anop 348	30	A	G	97	41.2
235	Anop 349	423	G	A	73	38.4
236	Anop 351	568	C	T	42	42.9
237	Anop 351	706	C	T	70	51.4
238	Anop 351	712	G	T	70	51.4
239	Anop 351	715	T	C	71	50.7
240	Anop 351	720	G	A	53	39.6
241	Anop 351	741	T	C	64	46.9
242	Anop 351	742	C	G	64	46.9
243	Anop 351	743	T	A	64	46.9
244	Anop 351	758	A	G	75	46.7
245	Anop 351	759	A	C	75	45.3
246	Anop 351	818	C	A	69	37.7
247	Anop 351	843	A	G	66	39.4
248	Anop 351	847	C	A	67	35.8
249	Anop 351	888	G	A	62	38.7
250	Anop 351	901	T	C	62	40.3
251	Anop 351	930	A	G	45	35.6
252	Anop 354	298	A	G	39	43.6
253	Anop 354	340	T	C	41	43.9
254	Anop 358	146	G	A	104	35.6
255	Anop 358	403	C	T	161	35.4
256	Anop 358	430	G	A	119	36.1
257	Anop 358	578	T	C	45	48.9
258	Anop 358	606	A	G	151	35.1
259	Anop 358	678	A	G	144	36.1
260	Anop 364	214	G	A	192	39.1
261	Anop 364	215	T	C	195	40.5
262	Anop 364	292	T	C	210	36.7
263	Anop 364	406	T	C	199	41.2
264	Anop 364	460	C	T	192	41.7
265	Anop 364	654	G	A	86	40.7
266	Anop 365	1350	C	T	52	46.2
267	Anop 376	77	G	A	43	37.2
268	Anop 376	139	A	C	54	46.3
269	Anop 376	314	G	A	53	47.2
270	Anop 376	444	A	G	46	50.0

APPENDIX J1. (cont.)

271	Anop 376	736	G	T	45	37.8
272	Anop 376	832	A	T	44	43.2
273	Anop 377	286	T	C	61	42.6
274	Anop 377	310	C	T	63	39.7
275	Anop 380	90	C	A	39	35.9
276	Anop 380	96	C	T	37	37.8
277	Anop 380	99	C	T	38	42.1
278	Anop 380	105	G	A	43	37.2
279	Anop 380	108	G	A	43	37.2
280	Anop 389	261	A	G	613	41.9
281	Anop 389	426	C	T	737	50.1
282	Anop 392	165	T	A	129	49.6
283	Anop 392	309	A	G	145	45.5
284	Anop 392	345	T	G	110	40.9
285	Anop 395	266	C	T	148	45.3
286	Anop 395	620	G	T	215	49.8
287	Anop 395	743	T	C	218	49.1
288	Anop 422	9	G	A	62	40.3
289	Anop 422	33	G	A	93	39.8
290	Anop 422	124	G	C	122	43.4
291	Anop 422	136	C	T	116	42.2
292	Anop 422	264	G	A	91	45.1
293	Anop 422	301	C	T	87	46.0
294	Anop 422	323	C	A	54	38.9
295	Anop 433	152	A	G	67	46.3
296	Anop 433	164	T	C	74	36.5
297	Anop 447	407	T	C	63	38.1
298	Anop 467	145	A	C	75	45.3
299	Anop 467	250	T	C	89	42.7
300	Anop 467	469	C	T	55	43.6
301	Anop 467	532	G	A	70	47.1
302	Anop 471	281	G	A	55	40.0
303	Anop 473	255	A	G	51	43.1
304	Anop 480	465	T	C	44	43.2
305	Anop 485	56	T	C	112	39.3
306	Anop 485	67	C	T	101	36.6
307	Anop 485	76	C	T	104	35.6
308	Anop 494	159	T	A	59	35.6
309	Anop 494	274	A	C	44	40.9
310	Anop 494	276	C	T	44	50.0

APPENDIX J1. (cont.)

311	Anop 494	281	T	C	42	50.0
312	Anop 494	339	A	G	49	42.9
313	Anop 494	423	T	C	43	37.2
314	Anop 503	75	T	C	76	39.5
315	Anop 503	99	A	G	82	40.2
316	Anop 503	307	C	T	91	36.3
317	Anop 503	399	G	C	82	47.6
318	Anop 503	482	A	T	82	52.4
319	Anop 517	67	T	C	36	38.9
320	Anop 520	301	T	C	127	44.9
321	Anop 523	451	G	T	96	50.0
322	Anop 527	446	G	T	45	35.6
323	Anop 534	350	G	A	113	38.1
324	Anop 534	392	G	C	118	35.6
325	Anop 534	545	T	G	93	44.1
326	Anop 541	381	C	G	420	36.2
327	Anop 552	309	C	T	127	44.9
328	Anop 552	354	C	T	136	42.6
329	Anop 561	405	G	A	47	38.3
330	Anop 568	290	C	G	37	45.9
331	Anop 568	435	T	C	75	41.3
332	Anop 569	368	A	G	46	50.0
333	Anop 576	27	T	A	37	43.2
334	Anop 576	70	C	T	41	43.9
335	Anop 576	215	A	T	44	36.4
336	Anop 576	251	G	A	44	36.4
337	Anop 576	257	C	T	49	42.9
338	Anop 576	266	G	A	47	46.8
339	Anop 576	599	T	C	57	47.4
340	Anop 576	611	C	T	51	43.1
341	Anop 576	632	G	A	48	45.8
342	Anop 576	635	T	C	45	37.8
343	Anop 576	860	T	A	62	41.9
344	Anop 576	1061	C	T	100	41.0
345	Anop 576	1226	T	A	37	43.2
346	Anop 582	430	G	A	92	37.0
347	Anop 597	1191	A	G	39	48.7
348	Anop 597	1737	T	A	45	37.8

APPENDIX J1. (cont.)

349	Anop 602	335	C	G	80	36.3
350	Anop 602	423	T	A	62	43.5
351	Anop 602	431	A	C	66	43.9
352	Anop 602	432	G	A	66	43.9
353	Anop 602	465	C	T	70	37.1
354	Anop 602	486	A	G	62	46.8
355	Anop 602	569	T	C	56	46.4
356	Anop 602	587	A	G	60	46.7
357	Anop 602	588	T	C	60	46.7
358	Anop 602	632	C	T	79	41.8
359	Anop 602	648	T	A	71	47.9
360	Anop 602	731	T	C	54	55.6
361	Anop 609	108	T	C	37	43.2
362	Anop 609	260	C	T	47	46.8
363	Anop 615	72	A	T	197	42.6
364	Anop 622	333	T	C	54	42.6
365	Anop 622	339	G	C	53	47.2
366	Anop 622	369	G	A	45	40.0
367	Anop 622	414	A	G	49	38.8
368	Anop 622	435	C	T	48	47.9
369	Anop 622	444	A	G	52	48.1
370	Anop 622	498	T	C	61	47.5
371	Anop 622	530	G	A	63	46.0
372	Anop 643	22	G	A	75	48.0
373	Anop 643	113	T	C	90	42.2
374	Anop 643	116	T	C	89	41.6
375	Anop 643	130	T	C	87	47.1
376	Anop 643	291	A	C	48	35.4
377	Anop 648	553	A	G	35	40.0
378	Anop 648	566	C	G	36	36.1
379	Anop 648	582	A	G	35	45.7
380	Anop 658	20	A	G	58	36.2
381	Anop 658	21	G	T	58	36.2
382	Anop 658	44	A	C	59	40.7
383	Anop 658	137	A	G	86	41.9
384	Anop 658	355	G	A	49	40.8
385	Anop 662	404	C	T	133	35.3
386	Anop 664	97	G	A	60	43.3
387	Anop 664	250	G	C	71	38.0
388	Anop 664	324	C	T	77	37.7

APPENDIX J1. (cont.)

389	Anop 674	241	A	G	49	44.9
390	Anop 678	1051	A	G	38	44.7
391	Anop 683	51	G	C	139	44.6
392	Anop 683	90	G	C	141	47.5
393	Anop 683	93	A	G	139	48.2
394	Anop 683	103	C	T	123	48.8
395	Anop 683	112	T	C	104	43.3
396	Anop 683	225	A	C	126	46.8
397	Anop 683	277	G	C	106	43.4
398	Anop 683	328	C	G	88	40.9
399	Anop 708	600	T	C	43	44.2
400	Anop 708	687	T	C	45	48.9
401	Anop 708	690	G	C	41	46.3
402	Anop 708	699	T	C	38	55.3
403	Anop 708	705	T	G	40	52.5
404	Anop 708	720	A	G	42	50.0
405	Anop 714	459	C	T	51	45.1
406	Anop 714	478	T	C	46	37.0
407	Anop 729	138	C	A	39	46.2
408	Anop 729	158	G	T	41	46.3
409	Anop 729	273	C	T	61	45.9
410	Anop 744	205	T	C	215	46.0
411	Anop 744	393	G	A	217	47.0
412	Anop 758	32	A	T	38	36.8
413	Anop 758	198	A	T	47	40.4
414	Anop 758	384	T	C	49	36.7
415	Anop 759	146	T	C	222	40.1
416	Anop 759	251	C	T	346	39.9
417	Anop 759	359	G	A	332	35.2
418	Anop 771	581	A	G	124	37.1
419	Anop 771	591	A	T	117	40.2
420	Anop 771	634	T	C	53	43.4
421	Anop 771	695	T	A	47	40.4
422	Anop 771	696	G	A	47	40.4
423	Anop 771	739	A	G	37	45.9
424	Anop 772	315	T	G	87	44.8
425	Anop 772	338	A	T	89	36.0
426	Anop 772	395	G	A	73	54.8

APPENDIX J1. (cont.)

427	Anop 772	1103	C	T	156	36.5
428	Anop 782	101	C	A	45	44.4
429	Anop 782	138	T	G	55	41.8
430	Anop 782	175	G	C	56	37.5
431	Anop 782	211	A	G	42	35.7
432	Anop 782	213	C	T	50	38.0
433	Anop 782	297	T	A	37	40.5
434	Anop 799	402	T	C	115	36.5
435	Anop 800	266	A	G	172	47.1
436	Anop 805	67	A	G	65	47.7
437	Anop 805	130	G	C	82	40.2
438	Anop 805	204	A	C	86	43.0
439	Anop 805	266	A	G	73	45.2
440	Anop 805	278	T	C	53	45.3
441	Anop 805	383	C	T	48	43.8
442	Anop 807	378	C	T	223	55.6
443	Anop 809	278	C	T	139	38.1
444	Anop 809	545	T	C	99	36.4
445	Anop 818	205	G	C	36	36.1
446	Anop 818	244	C	T	36	36.1
447	Anop 818	265	C	G	39	38.5
448	Anop 818	342	G	A	38	42.1
449	Anop 818	661	C	T	45	40.0
450	Anop 825	133	C	T	49	36.7
451	Anop 825	661	T	A	44	45.5
452	Anop 828	157	C	T	88	45.5
453	Anop 829	516	A	G	36	38.9
454	Anop 843	239	T	A	96	39.6
455	Anop 843	284	C	G	97	41.2
456	Anop 849	108	T	A	84	42.9
457	Anop 850	346	C	T	39	46.2
458	Anop 857	264	A	G	71	42.3
459	Anop 877	378	A	G	35	51.4
460	Anop 877	379	C	T	35	37.1
461	Anop 878	783	C	G	35	37.1
462	Anop 878	851	T	A	40	35.0
463	Anop 878	1657	C	T	56	39.3
464	Anop 878	1666	G	T	51	39.2
465	Anop 878	1674	A	T	40	37.5

APPENDIX J1. (cont.)

466	Anop 878	1678	A	C	46	52.2
467	Anop 878	1702	C	G	63	39.7
468	Anop 886	148	A	G	58	46.6
469	Anop 886	238	T	C	59	44.1
470	Anop 886	289	C	G	54	44.4
471	Anop 886	391	A	G	52	48.1
472	Anop 902	489	T	C	46	43.5
473	Anop 906	183	T	G	144	42.4
474	Anop 906	252	C	A	147	42.9
475	Anop 906	336	G	A	141	43.3
476	Anop 906	707	A	G	102	43.1
477	Anop 906	1199	A	G	55	38.2
478	Anop 906	1262	A	G	39	38.5
479	Anop 906	1700	G	A	78	41.0
480	Anop 906	2552	G	A	50	48.0
481	Anop 908	233	C	T	39	41.0
482	Anop 908	302	T	C	50	48.0
483	Anop 908	677	A	G	81	43.2
484	Anop 911	343	T	C	38	39.5
485	Anop 911	370	G	A	41	36.6
486	Anop 911	1170	G	A	161	45.3
487	Anop 911	1230	G	A	179	49.7
488	Anop 911	1461	G	A	165	38.8
489	Anop 911	1503	A	G	177	41.8
490	Anop 911	1606	G	A	48	39.6
491	Anop 911	1630	C	T	69	47.8
492	Anop 911	1954	C	T	56	42.9
493	Anop 958	1107	T	C	36	50.0
494	Anop 958	1191	T	A	44	38.6
495	Anop 967	29	A	G	67	37.3
496	Anop 967	50	G	C	68	47.1
497	Anop 967	97	C	A	115	40.9
498	Anop 967	140	A	C	120	44.2
499	Anop 971	335	T	A	40	35.0
500	Anop 971	358	C	T	42	38.1
501	Anop 971	393	A	C	36	47.2
502	Anop 982	55	C	T	52	36.5
503	Anop 982	150	C	G	62	35.5

APPENDIX J1. (cont.)

504	Anop 982	162	C	G	76	38.2
505	Anop 982	277	A	G	75	46.7
506	Anop 982	393	G	C	74	43.2
507	Anop 1000	1076	G	A	76	42.1
508	Anop 1017	502	G	A	42	54.8
509	Anop 1017	598	A	G	53	47.2
510	Anop 1017	784	T	C	36	38.9
511	Anop 1017	937	G	A	47	36.2
512	Anop 1017	943	C	T	47	44.7
513	Anop 1027	93	C	G	36	41.7
514	Anop 1027	95	A	G	36	41.7
515	Anop 1027	110	C	T	39	43.6
516	Anop 1027	111	G	T	39	41.0
517	Anop 1027	131	C	G	39	41.0
518	Anop 1027	150	C	T	36	47.2
519	Anop 1027	154	A	T	36	44.4
520	Anop 1027	359	A	T	37	48.6
521	Anop 1050	195	G	A	51	47.1
522	Anop 1050	240	T	C	51	41.2
523	Anop 1050	252	C	T	50	42.0
524	Anop 1050	383	A	G	56	41.1
525	Anop 1050	384	G	A	56	41.1
526	Anop 1050	609	T	C	41	43.9
527	Anop 1050	621	C	G	38	47.4
528	Anop 1050	645	C	G	37	37.8
529	Anop 1084	343	G	A	377	47.2
530	Anop 1085	2116	A	G	35	42.9
531	Anop 1085	2344	C	T	36	36.1
532	Anop 1109	333	C	T	232	36.6
533	Anop 1125	641	A	G	46	47.8
534	Anop 1126	20	A	G	63	49.2
535	Anop 1126	176	A	G	84	39.3
536	Anop 1126	188	A	G	82	39.0
537	Anop 1126	197	G	A	87	37.9
538	Anop 1136	279	A	G	60	48.3
539	Anop 1163	1920	C	A	59	39.0
540	Anop 1166	165	A	C	43	39.5
541	Anop 1166	214	C	G	43	44.2

APPENDIX J1. (cont.)

542	Anop 1170	213	C	T	64	45.3
543	Anop 1170	598	C	T	36	38.9
544	Anop 1173	20	T	C	43	51.2
545	Anop 1173	26	T	C	45	48.9
546	Anop 1173	32	C	T	51	45.1
547	Anop 1173	344	G	A	67	47.8
548	Anop 1189	376	T	C	453	47.7
549	Anop 1189	529	C	A	271	45.0
550	Anop 1193	465	C	T	72	48.6
551	Anop 1193	466	C	T	72	48.6
552	Anop 1201	322	C	T	46	41.3
553	Anop 1201	433	A	G	55	43.6
554	Anop 1201	466	C	T	54	40.7
555	Anop 1201	499	A	G	50	48.0
556	Anop 1202	317	A	G	130	49.2
557	Anop 1202	343	T	C	112	41.1
558	Anop 1207	457	A	T	40	42.5
559	Anop 1207	554	A	G	56	50.0
560	Anop 1207	621	G	C	52	40.4
561	Anop 1217	339	T	C	97	45.4
562	Anop 1217	698	C	T	56	35.7
563	Anop 1229	788	G	A	75	44.0
564	Anop 1229	1358	A	C	36	36.1
565	Anop 1233	430	C	T	77	48.1
566	Anop 1233	955	A	G	75	48.0
567	Anop 1236	294	A	T	166	36.1
568	Anop 1237	286	G	C	54	38.9
569	Anop 1237	430	C	T	53	41.5
570	Anop 1237	454	G	A	55	47.3
571	Anop 1243	336	G	A	93	38.7
572	Anop 1243	429	T	C	90	35.6
573	Anop 1243	573	G	A	72	36.1
574	Anop 1268	144	T	C	104	49.0
575	Anop 1269	437	A	G	95	46.3
576	Anop 1276	951	T	C	63	36.5
577	Anop 1276	1012	G	T	45	42.2
578	Anop 1276	1013	T	C	45	42.2
579	Anop 1276	1032	A	T	51	45.1
580	Anop 1277	574	G	A	39	35.9

APPENDIX J1. (cont.)

581	Anop 1294	160	G	A	185	35.1
582	Anop 1294	175	A	C	202	44.1
583	Anop 1295	302	G	A	37	45.9
584	Anop 1295	311	T	A	36	41.7
585	Anop 1295	534	A	T	93	38.7
586	Anop 1295	536	T	G	93	38.7
587	Anop 1295	707	A	T	57	47.4
588	Anop 1295	846	C	T	67	41.8
589	Anop 1295	859	G	A	66	42.4
590	Anop 1302	272	G	A	78	42.3
591	Anop 1302	288	C	T	77	35.1
592	Anop 1312	488	C	A	41	46.3
593	Anop 1312	498	G	C	37	40.5
594	Anop 1312	531	G	A	36	36.1
595	Anop 1312	569	G	T	42	47.6
596	Anop 1349	30	A	G	35	37.1
597	Anop 1349	51	G	A	35	37.1
598	Anop 1357	98	G	A	63	39.7
599	Anop 1361	603	C	A	64	42.2
600	Anop 1361	604	C	G	64	35.9
601	Anop 1361	775	C	A	36	47.2
602	Anop 1375	455	T	G	39	48.7
603	Anop 1384	203	T	A	35	42.9
604	Anop 1384	287	C	G	35	42.9
605	Anop 1411	317	A	G	79	39.2
606	Anop 1426	150	T	C	35	40.0
607	Anop 1518	573	T	C	56	48.2
608	Anop 1524	115	C	T	44	43.2
609	Anop 1526	114	G	C	37	43.2
610	Anop 1526	126	C	T	37	43.2
611	Anop 1526	255	T	C	38	36.8
612	Anop 1526	258	G	A	38	36.8
613	Anop 1526	304	T	C	35	37.1
614	Anop 1550	136	T	G	58	43.1
615	Anop 1553	338	A	T	44	38.6
616	Anop 1553	426	A	T	46	45.7
617	Anop 1563	245	C	A	39	41.0
618	Anop 1627	419	G	C	42	47.6
619	Anop 1627	437	G	A	64	37.5

APPENDIX J1. (cont.)

620	Anop 1627	440	G	T	62	51.6
621	Anop 1627	506	G	A	46	41.3
622	Anop 1627	616	G	A	49	40.8
623	Anop 1627	625	G	T	41	36.6
624	Anop 1666	201	C	T	86	37.2
625	Anop 1686	109	A	T	56	42.9
626	Anop 1686	265	G	A	64	40.6
627	Anop 1686	381	A	G	89	51.7
628	Anop 1686	514	G	A	74	45.9
629	Anop 1692	506	A	C	60	41.7
630	Anop 1692	640	C	A	64	37.5
631	Anop 1692	1018	A	G	54	46.3
632	Anop 1703	20	A	G	52	38.5
633	Anop 1703	414	T	G	78	43.6
634	Anop 1703	460	T	C	70	40.0
635	Anop 1766	311	T	C	69	46.4
636	Anop 1776	1451	A	G	47	38.3
637	Anop 1794	472	A	G	102	45.1
638	Anop 1815	468	T	C	43	39.5
639	Anop 1815	518	C	A	36	38.9
640	Anop 1823	458	A	G	51	41.2
641	Anop 1823	472	T	C	49	46.9
642	Anop 1835	199	C	T	42	42.9
643	Anop 1835	205	A	C	43	37.2
644	Anop 1842	372	A	G	168	38.1
645	Anop 1844	131	C	T	39	48.7
646	Anop 1844	158	G	A	41	41.5
647	Anop 1844	218	T	C	47	40.4
648	Anop 1844	272	C	T	40	45.0
649	Anop 1844	296	A	T	47	42.6
650	Anop 1855	233	A	G	35	45.7
651	Anop 1856	475	G	C	50	42.0
652	Anop 1856	491	C	T	44	43.2
653	Anop 1856	493	G	C	44	43.2
654	Anop 1856	498	G	C	54	35.2
655	Anop 1856	578	A	T	39	56.4
656	Anop 1856	608	T	C	45	40.0
657	Anop 1856	743	T	G	35	37.1

APPENDIX J1. (cont.)

658	Anop 1896	783	T	C	38	39.5
659	Anop 1896	1023	T	C	48	56.3
660	Anop 1896	1026	C	T	59	35.6
661	Anop 1909	220	A	T	94	43.6
662	Anop 1909	221	A	C	94	43.6
663	Anop 1945	2065	A	T	43	46.5
664	Anop 1989	1383	A	G	39	46.2
665	Anop 2003	415	C	T	80	42.5
666	Anop 2142	615	T	G	87	40.2
667	Anop 2149	396	T	A	37	40.5
668	Anop 2178	361	C	T	45	46.7
669	Anop 2178	596	T	C	52	40.4
670	Anop 2178	732	G	C	44	43.2
671	Anop 2212	206	G	A	38	39.5
672	Anop 2312	20	C	G	59	47.5
673	Anop 2312	29	T	A	61	49.2
674	Anop 2378	263	G	A	39	35.9
675	Anop 2385	100	A	G	38	47.4
676	Anop 2385	166	C	T	40	50.0
677	Anop 2385	255	C	G	38	36.8
678	Anop 2385	274	G	A	39	46.2
679	Anop 2385	283	A	G	38	42.1
680	Anop 2440	516	C	T	99	37.4
681	Anop 2440	600	A	G	59	35.6
682	Anop 2477	199	G	T	36	50.0
683	Anop 2534	1254	C	T	35	45.7
684	Anop 2559	1182	T	C	67	40.3
685	Anop 2571	67	C	T	39	46.2
686	Anop 2571	106	T	C	45	46.7
687	Anop 2571	301	A	C	56	44.6
688	Anop 2593	350	G	C	42	42.9
689	Anop 2635	643	C	T	37	45.9
690	Anop 2635	739	C	T	40	47.5
691	Anop 2636	263	C	T	59	45.8
692	Anop 2636	281	A	G	62	38.7
693	Anop 2636	449	G	T	55	43.6
694	Anop 2804	321	C	T	87	40.2
695	Anop 2828	152	G	A	42	38.1

APPENDIX J1. (cont.)

696	Anop 2828	323	G	A	58	48.3
697	Anop 2833	171	A	C	36	36.1
698	Anop 2833	173	G	A	36	36.1
699	Anop 2833	181	C	A	36	36.1
700	Anop 2833	198	A	C	37	37.8
701	Anop 2833	201	C	T	37	37.8
702	Anop 2833	222	T	A	38	39.5
703	Anop 2833	263	T	C	40	35.0
704	Anop 2833	264	A	T	39	38.5
705	Anop 2833	283	A	G	39	35.9
706	Anop 2861	1141	T	C	40	47.5
707	Anop 2926	86	G	A	72	40.3
708	Anop 2926	117	C	T	72	36.1
709	Anop 3353	39	G	C	43	37.2
710	Anop 3353	74	C	A	51	35.3
711	Anop 3353	75	T	A	51	35.3
712	Anop 3353	99	C	T	55	36.4
713	Anop 3353	105	A	G	55	36.4
714	Anop 3353	148	G	A	55	36.4
715	Anop 3353	158	C	T	57	40.4
716	Anop 3353	170	C	G	54	37.0
717	Anop 3353	172	T	A	53	35.8
718	Anop 3353	174	G	A	54	35.2
719	Anop 3353	184	C	G	59	39.0
720	Anop 3353	186	T	A	59	39.0
721	Anop 3353	188	C	A	59	39.0
722	Anop 3353	233	T	G	61	37.7
723	Anop 3353	246	T	A	58	36.2
724	Anop 3481	591	G	C	74	43.2
725	Anop 3481	597	C	A	77	46.8
726	Anop 3481	744	T	A	61	45.9
727	Anop 3688	185	G	A	49	40.8
728	Anop 3688	190	A	C	44	47.7
729	Anop 3688	196	C	T	40	40.0
730	Anop 3688	202	T	A	49	46.9
731	Anop 3688	272	A	G	47	40.4
732	Anop 3688	303	T	C	44	40.9
733	Anop 3823	637	G	C	38	47.4

APPENDIX J1. (cont.)

734	Anop 3823	643	A	G	39	41.0
735	Anop 4243	124	C	A	53	39.6
736	Anop 5253	227	C	G	43	41.9
737	Anop 5253	312	A	G	37	37.8
738	Anop 5253	328	A	G	36	38.9
739	Anop 6789	324	C	T	38	42.1
740	Anop 6789	327	A	G	38	42.1
741	Anop 6789	339	G	T	39	48.7
742	Anop 8555	451	C	T	311	35.0

APPENDIX J2. *A. craccivora.*

Serial Number	Contig	Consensus Position	Consensus	Allele	Coverage	Frequency
1	Aphis 41	98	A	T	39	35.9
2	Aphis 41	918	C	T	88	38.6
3	Aphis 45	1329	T	C	45	48.9
4	Aphis 85	138	C	T	62	48.4
5	Aphis 85	289	T	C	70	42.9
6	Aphis 90	144	G	A	67	49.3
7	Aphis 90	164	C	T	67	47.8
8	Aphis 90	221	C	T	68	50.0
9	Aphis 102	282	A	G	59	49.2
10	Aphis 102	333	T	C	79	36.7
11	Aphis 102	429	T	A	74	36.5
12	Aphis 102	468	C	T	72	36.1
13	Aphis 102	480	C	T	64	39.1
14	Aphis 102	489	A	G	71	39.4
15	Aphis 103	215	G	A	51	45.1
16	Aphis 103	437	T	A	112	43.8
17	Aphis 111	462	T	C	167	49.1
18	Aphis 117	101	T	C	54	44.4
19	Aphis 117	445	C	T	56	44.6
20	Aphis 133	565	G	A	53	47.2
21	Aphis 133	595	A	G	40	52.5
22	Aphis 200	245	A	C	43	41.9
23	Aphis 200	317	C	G	56	41.1
24	Aphis 200	488	G	A	40	47.5
25	Aphis 212	1286	C	G	84	48.8
26	Aphis 215	1480	T	A	165	49.1
27	Aphis 215	1534	T	C	152	40.8
28	Aphis 242	1886	G	A	57	36.8
29	Aphis 255	72	G	A	37	37.8
30	Aphis 255	81	C	T	41	41.5
31	Aphis 255	93	C	T	41	43.9
32	Aphis 255	120	A	G	44	40.9
33	Aphis 255	123	G	A	44	40.9
34	Aphis 255	126	C	T	45	40.0
35	Aphis 255	132	A	T	46	39.1
36	Aphis 255	136	A	G	44	36.4
37	Aphis 255	231	A	G	46	52.2

APPENDIX J2. *A. craccivora.*

38	Aphis 255	255	G	A	52	40.4
39	Aphis 255	701	A	G	54	48.1
40	Aphis 273	67	T	C	72	43.1
41	Aphis 273	171	C	T	87	47.1
42	Aphis 273	468	A	T	94	46.8
43	Aphis 273	489	G	A	98	42.9
44	Aphis 309	319	A	G	50	36.0
45	Aphis 321	372	A	G	41	46.3
46	Aphis 321	462	G	A	44	50.0
47	Aphis 321	471	G	A	43	48.8
48	Aphis 321	498	A	G	44	47.7
49	Aphis 321	507	T	A	44	47.7
50	Aphis 321	570	C	T	43	48.8
51	Aphis 321	579	A	G	43	48.8
52	Aphis 321	632	T	C	41	48.8
53	Aphis 321	636	G	T	42	50.0
54	Aphis 359	553	A	G	73	49.3
55	Aphis 359	556	T	C	79	46.8
56	Aphis 359	562	G	A	77	35.1
57	Aphis 359	649	A	C	85	55.3
58	Aphis 359	706	T	A	113	38.9
59	Aphis 359	712	A	G	94	43.6
60	Aphis 359	714	C	T	94	43.6
61	Aphis 359	715	A	T	94	43.6
62	Aphis 359	725	T	C	121	35.5
63	Aphis 359	730	T	C	123	44.7
64	Aphis 365	1829	T	G	97	41.2
65	Aphis 365	1837	G	A	95	42.1
66	Aphis 365	1850	A	T	75	41.3
67	Aphis 365	1852	G	A	75	41.3
68	Aphis 365	1856	C	T	80	43.8
69	Aphis 365	1865	G	A	69	44.9
70	Aphis 365	1874	G	A	68	47.1
71	Aphis 384	16	A	G	37	35.1
72	Aphis 384	41	A	T	45	37.8
73	Aphis 384	104	G	A	51	37.3
74	Aphis 384	110	C	G	51	37.3
75	Aphis 384	133	G	C	42	47.6

APPENDIX J2. (cont.)

76	Aphis 384	137	G	A	50	38.0
77	Aphis 384	147	A	T	50	38.0
78	Aphis 384	152	T	G	35	54.3
79	Aphis 384	153	T	A	35	54.3
80	Aphis 384	176	C	A	53	37.7
81	Aphis 384	179	G	A	53	35.8
82	Aphis 384	220	G	T	53	35.8
83	Aphis 384	257	C	A	53	35.8
84	Aphis 433	293	T	G	173	47.4
85	Aphis 433	440	A	T	176	48.3
86	Aphis 480	157	A	G	44	45.5
87	Aphis 525	993	T	A	62	43.5
88	Aphis 645	570	C	T	48	43.8
89	Aphis 645	645	G	A	48	41.7
90	Aphis 668	637	A	T	64	48.4
91	Aphis 704	419	C	T	59	39.0
92	Aphis 704	431	C	T	53	37.7
93	Aphis 704	447	C	A	45	44.4
94	Aphis 906	122	T	A	44	36.4
95	Aphis 1221	723	G	A	111	45.9
96	Aphis 1468	302	T	C	37	45.9
97	Aphis 1725	250	C	A	38	47.4

APPENDIX J3. *C. tomentosicollis*.

Serial Number	Contig	Consensus Position	Consensus	Allele	Coverage	Frequency
1	Clavig 2	393	A	G	37	45.9
2	Clavig 2	482	A	C	35	45.7
3	Clavig 6	42	T	A	182	38.5
4	Clavig 6	47	C	T	184	38.0
5	Clavig 6	104	G	C	229	35.8
6	Clavig 6	299	T	A	176	35.2
7	Clavig 7	276	T	C	154	42.9
8	Clavig 7	279	C	T	157	35.0
9	Clavig 7	315	T	C	166	41.6
10	Clavig 7	326	C	G	162	43.2
11	Clavig 7	333	C	T	161	43.5
12	Clavig 7	340	T	C	157	44.6
13	Clavig 9	24	A	G	219	53.4
14	Clavig 9	36	T	C	259	46.7
15	Clavig 9	47	C	G	262	48.1
16	Clavig 9	54	C	T	258	49.2
17	Clavig 9	61	T	C	259	51.0
18	Clavig 9	420	C	T	35	45.7
19	Clavig 17	726	C	T	56	42.9
20	Clavig 18	89	T	C	57	36.8
21	Clavig 19	268	T	C	55	45.5
22	Clavig 19	349	G	A	53	39.6
23	Clavig 20	1709	C	T	1244	43.2
24	Clavig 20	2517	T	C	100	36.0
25	Clavig 26	213	T	A	156	42.3
26	Clavig 37	144	G	A	193	42.0
27	Clavig 37	248	G	A	211	45.5
28	Clavig 39	211	T	C	78	43.6
29	Clavig 39	374	T	C	62	38.7
30	Clavig 45	67	T	G	302	43.0
31	Clavig 45	68	T	C	302	43.0
32	Clavig 45	125	T	C	321	39.9
33	Clavig 45	191	T	C	319	46.1
34	Clavig 45	266	A	C	329	45.6
35	Clavig 45	278	T	C	277	39.4

APPENDIX J3. (Cont.)

36	Clavig 46	224	A	G	116	35.3
37	Clavig 46	314	A	G	114	36.8
38	Clavig 46	389	T	A	141	47.5
39	Clavig 50	258	G	A	495	36.4
40	Clavig 50	276	A	T	465	43.0
41	Clavig 51	372	A	G	90	35.6
42	Clavig 51	396	T	C	96	46.9
43	Clavig 51	678	C	T	68	48.5
44	Clavig 53	325	G	A	278	39.9
45	Clavig 66	91	A	G	244	43.9
46	Clavig 66	202	A	T	302	48.7
47	Clavig 67	349	G	A	44	45.5
48	Clavig 67	403	C	T	53	41.5
49	Clavig 67	574	A	G	54	40.7
50	Clavig 69	142	C	T	56	35.7
51	Clavig 69	301	A	C	108	46.3
52	Clavig 69	302	C	A	108	46.3
53	Clavig 69	519	G	A	102	42.2
54	Clavig 69	575	G	A	121	43.0
55	Clavig 69	855	C	T	75	46.7
56	Clavig 69	874	A	G	68	38.2
57	Clavig 69	875	T	G	68	38.2
58	Clavig 70	261	A	G	36	50.0
59	Clavig 75	795	G	A	345	42.3
60	Clavig 75	1005	G	A	228	38.6
61	Clavig 75	1413	G	T	158	35.4
62	Clavig 76	580	C	T	47	44.7
63	Clavig 76	601	C	T	52	42.3
64	Clavig 81	366	A	G	240	35.4
65	Clavig 85	259	C	T	60	35.0
66	Clavig 85	571	C	G	68	47.1
67	Clavig 85	604	A	T	64	39.1
68	Clavig 85	610	A	G	52	40.4
69	Clavig 85	679	T	C	36	36.1
70	Clavig 89	2195	A	T	39	38.5
71	Clavig 91	164	C	T	62	37.1
72	Clavig 103	256	T	A	67	49.3
73	Clavig 103	365	A	T	66	37.9

APPENDIX J3. (Cont.)

74	Clavig 106	72	C	T	410	50.7
75	Clavig 106	109	C	T	392	37.8
76	Clavig 106	147	T	A	594	38.4
77	Clavig 106	151	G	T	594	38.2
78	Clavig 106	162	C	T	614	44.6
79	Clavig 106	222	G	A	721	44.1
80	Clavig 106	299	T	G	577	39.5
81	Clavig 106	300	T	C	577	39.5
82	Clavig 106	423	T	C	290	54.5
83	Clavig 107	259	A	G	48	47.9
84	Clavig 107	303	A	G	50	44.0
85	Clavig 107	321	C	T	45	53.3
86	Clavig 107	609	T	C	55	43.6
87	Clavig 107	693	T	A	35	48.6
88	Clavig 111	106	A	G	180	35.6
89	Clavig 115	137	T	A	362	45.0
90	Clavig 115	538	T	C	338	41.7
91	Clavig 116	354	T	G	115	41.7
92	Clavig 116	1733	T	A	40	45.0
93	Clavig 126	192	A	G	42	45.2
94	Clavig 127	31	G	A	65	36.9
95	Clavig 129	152	G	A	168	44.0
96	Clavig 129	241	C	T	236	36.4
97	Clavig 129	439	A	T	406	45.3
98	Clavig 131	492	A	G	36	50.0
99	Clavig 142	147	A	G	58	43.1
100	Clavig 149	20	C	T	38	42.1
101	Clavig 151	268	G	A	262	47.7
102	Clavig 158	16	G	C	528	40.3
103	Clavig 158	60	G	A	569	37.1
104	Clavig 158	90	G	A	493	47.9
105	Clavig 159	379	C	T	246	47.6
106	Clavig 166	462	T	C	66	36.4
107	Clavig 166	966	G	A	45	40.0
108	Clavig 166	1158	A	T	85	38.8
109	Clavig 167	378	T	C	36	47.2
110	Clavig 167	577	T	C	58	37.9
111	Clavig 167	603	G	A	60	36.7
112	Clavig 167	662	G	C	63	47.6

APPENDIX J3. (Cont.)

113	Clavig 167	967	C	G	46	37.0
114	Clavig 168	454	A	G	908	39.3
115	Clavig 168	574	A	T	1015	40.2
116	Clavig 168	799	A	G	610	47.2
117	Clavig 168	966	G	C	315	46.7
118	Clavig 172	136	T	C	36	44.4
119	Clavig 172	980	G	A	73	35.6
120	Clavig 172	1350	C	A	121	38.8
121	Clavig 174	353	C	A	40	47.5
122	Clavig 174	1132	G	A	39	41.0
123	Clavig 176	3000	T	G	35	37.1
124	Clavig 177	789	A	T	56	48.2
125	Clavig 177	909	A	G	45	48.9
126	Clavig 178	68	T	A	468	37.4
127	Clavig 178	71	T	C	464	45.0
128	Clavig 178	103	T	C	442	47.1
129	Clavig 178	290	A	T	448	41.5
130	Clavig 178	296	C	T	395	39.7
131	Clavig 180	169	T	A	89	42.7
132	Clavig 187	103	C	T	62	50.0
133	Clavig 187	104	C	A	62	50.0
134	Clavig 187	124	T	C	72	45.8
135	Clavig 187	142	A	G	69	39.1
136	Clavig 187	250	G	A	79	39.2
137	Clavig 191	374	C	T	144	43.8
138	Clavig 191	908	G	A	125	42.4
139	Clavig 192	61	A	G	37	48.6
140	Clavig 192	296	C	T	163	46.6
141	Clavig 196	202	C	T	50	48.0
142	Clavig 201	21	G	A	51	39.2
143	Clavig 201	418	G	C	87	47.1
144	Clavig 206	351	C	T	165	35.8
145	Clavig 214	3416	T	C	66	42.4
146	Clavig 214	3724	T	C	50	54.0
147	Clavig 214	3872	G	T	73	35.6
148	Clavig 214	3881	A	G	73	35.6
149	Clavig 214	3944	A	G	76	35.5
150	Clavig 214	4088	T	C	61	44.3

APPENDIX J3. (Cont.)

151	Clavig 214	4115	T	C	59	47.5
152	Clavig 214	6027	T	C	71	45.1
153	Clavig 215	470	G	C	119	37.0
154	Clavig 227	82	G	T	192	37.5
155	Clavig 227	99	C	A	189	37.0
156	Clavig 227	100	A	C	189	37.0
157	Clavig 227	214	G	A	205	47.3
158	Clavig 227	368	C	T	248	35.1
159	Clavig 238	365	T	G	141	36.2
160	Clavig 239	208	C	T	95	44.2
161	Clavig 239	352	T	C	78	43.6
162	Clavig 239	536	C	A	75	37.3
163	Clavig 239	670	A	G	84	47.6
164	Clavig 239	745	A	G	74	40.5
165	Clavig 239	757	G	A	80	35.0
166	Clavig 239	802	T	A	80	42.5
167	Clavig 239	1276	T	C	89	49.4
168	Clavig 239	1312	T	C	73	41.1
169	Clavig 242	260	G	T	41	41.5
170	Clavig 243	533	T	A	191	49.7
171	Clavig 243	566	T	A	173	37.0
172	Clavig 247	177	C	T	66	45.5
173	Clavig 251	114	T	C	67	37.3
174	Clavig 251	118	T	C	58	43.1
175	Clavig 251	244	C	G	169	47.9
176	Clavig 251	268	T	A	167	36.5
177	Clavig 256	2843	C	T	36	44.4
178	Clavig 256	2879	T	G	39	48.7
179	Clavig 256	2882	T	C	39	38.5
180	Clavig 256	3044	A	G	39	46.2
181	Clavig 256	3080	C	T	38	47.4
182	Clavig 261	323	T	A	101	43.6
183	Clavig 266	72	G	C	86	45.3
184	Clavig 284	625	G	A	55	47.3
185	Clavig 293	996	G	A	66	45.5
186	Clavig 293	1049	A	T	66	50.0
187	Clavig 296	65	T	A	103	36.9
188	Clavig 296	74	G	T	94	47.9
189	Clavig 296	145	C	G	109	37.6

APPENDIX J3. (Cont.)

190	Clavig	296	146	A	C	109	37.6
191	Clavig	296	152	G	C	98	42.9
192	Clavig	296	166	A	G	83	43.4
193	Clavig	296	167	G	A	83	43.4
194	Clavig	296	196	A	G	84	42.9
195	Clavig	296	197	A	T	84	42.9
196	Clavig	296	200	T	C	92	38.0
197	Clavig	296	215	A	T	94	35.1
198	Clavig	296	218	A	T	96	38.5
199	Clavig	297	170	G	A	69	46.4
200	Clavig	298	289	C	A	68	35.3
201	Clavig	298	345	C	T	46	41.3
202	Clavig	298	360	A	G	39	43.6
203	Clavig	298	363	C	T	39	43.6
204	Clavig	298	372	C	G	35	42.9
205	Clavig	300	690	A	T	41	39.0
206	Clavig	300	948	T	G	40	40.0
207	Clavig	301	12	T	C	55	40.0
208	Clavig	301	279	T	A	50	46.0
209	Clavig	301	285	T	C	71	35.2
210	Clavig	301	288	A	T	64	35.9
211	Clavig	301	297	T	A	63	38.1
212	Clavig	301	304	G	T	57	38.6
213	Clavig	301	312	A	C	55	38.2
214	Clavig	301	327	A	T	44	36.4
215	Clavig	318	577	A	G	81	40.7
216	Clavig	318	580	T	C	87	36.8
217	Clavig	318	700	T	A	84	40.5
218	Clavig	320	446	G	A	87	42.5
219	Clavig	320	656	T	C	64	42.2
220	Clavig	322	113	G	A	43	37.2
221	Clavig	322	159	T	C	45	40.0
222	Clavig	322	161	G	A	45	40.0
223	Clavig	332	24	G	A	309	44.0
224	Clavig	332	30	A	G	318	44.3
225	Clavig	332	60	C	T	345	44.3
226	Clavig	332	87	T	C	349	41.0

APPENDIX J3. (Cont.)

227	Clavig	332	291	T	C	285	38.2
228	Clavig	332	303	T	G	267	35.6
229	Clavig	332	402	T	C	148	37.8
230	Clavig	345	385	A	G	56	42.9
231	Clavig	349	256	A	T	84	46.4
232	Clavig	349	283	T	C	77	42.9
233	Clavig	356	58	T	C	149	45.0
234	Clavig	366	979	G	A	55	41.8
235	Clavig	372	128	G	A	252	48.4
236	Clavig	378	71	T	A	70	44.3
237	Clavig	378	72	A	G	70	44.3
238	Clavig	378	101	A	T	77	41.6
239	Clavig	384	352	A	C	118	35.6
240	Clavig	386	152	T	C	94	43.6
241	Clavig	386	359	A	G	83	37.3
242	Clavig	386	425	G	A	132	45.5
243	Clavig	386	456	G	A	151	50.3
244	Clavig	398	320	C	T	72	37.5
245	Clavig	398	509	A	G	41	36.6
246	Clavig	399	1011	T	A	41	36.6
247	Clavig	399	1044	A	G	41	36.6
248	Clavig	406	201	G	T	47	44.7
249	Clavig	406	211	C	T	46	47.8
250	Clavig	406	283	C	T	42	42.9
251	Clavig	406	312	C	G	36	44.4
252	Clavig	408	739	T	C	35	40.0
253	Clavig	408	936	G	C	40	35.0
254	Clavig	410	849	C	T	228	37.7
255	Clavig	410	1817	T	C	399	46.6
256	Clavig	426	1283	G	A	41	39.0
257	Clavig	433	186	G	A	275	37.1
258	Clavig	456	165	A	G	43	37.2
259	Clavig	456	285	A	T	50	36.0
260	Clavig	466	1806	C	T	86	37.2
261	Clavig	466	1845	G	A	86	40.7
262	Clavig	466	2325	C	T	175	45.1
263	Clavig	466	2373	A	G	161	46.6
264	Clavig	466	2411	G	A	139	46.0

APPENDIX J3. (Cont.)

265	Clavig 466	2447	C	T	177	44.6
266	Clavig 466	2480	T	A	189	49.7
267	Clavig 466	2679	T	C	73	35.6
268	Clavig 466	2685	A	C	74	37.8
269	Clavig 475	35	G	T	78	35.9
270	Clavig 484	297	T	C	69	39.1
271	Clavig 484	333	G	A	78	38.5
272	Clavig 487	92	A	G	50	36.0
273	Clavig 490	237	A	G	57	35.1
274	Clavig 490	351	C	T	58	36.2
275	Clavig 490	355	G	A	57	36.8
276	Clavig 497	218	C	T	392	37.8
277	Clavig 497	495	C	T	132	37.1
278	Clavig 501	1806	G	A	49	40.8
279	Clavig 501	2163	C	T	58	39.7
280	Clavig 501	2331	T	C	58	36.2
281	Clavig 501	2391	G	A	62	37.1
282	Clavig 501	2577	G	A	51	39.2
283	Clavig 501	2718	C	T	77	36.4
284	Clavig 501	2832	G	A	97	35.1
285	Clavig 501	3162	A	G	125	50.4
286	Clavig 509	148	T	C	54	40.7
287	Clavig 509	313	A	G	50	50.0
288	Clavig 513	261	G	A	279	35.1
289	Clavig 518	393	T	C	164	37.8
290	Clavig 518	515	A	G	177	42.4
291	Clavig 524	254	A	G	188	45.7
292	Clavig 524	293	A	T	202	45.1
293	Clavig 524	431	A	G	119	48.7
294	Clavig 533	71	C	T	106	36.8
295	Clavig 533	77	T	C	112	35.7
296	Clavig 533	92	C	T	111	41.4
297	Clavig 533	314	C	T	161	49.1
298	Clavig 533	1055	G	T	156	44.2
299	Clavig 544	42	C	G	67	46.3
300	Clavig 544	287	C	T	181	42.0
301	Clavig 544	481	A	C	51	39.2
302	Clavig 544	499	C	T	59	42.4

APPENDIX J3. (Cont.)

303	Clavig 564	317	A	G	173	42.2
304	Clavig 576	369	A	T	36	47.2
305	Clavig 587	155	T	C	57	35.1
306	Clavig 590	20	T	C	43	46.5
307	Clavig 590	91	G	A	52	38.5
308	Clavig 590	98	A	G	52	38.5
309	Clavig 590	137	G	A	59	49.2
310	Clavig 590	321	G	T	47	42.6
311	Clavig 591	124	C	T	59	44.1
312	Clavig 605	399	T	C	71	38.0
313	Clavig 606	725	G	A	267	38.2
314	Clavig 607	7	T	A	101	41.6
315	Clavig 607	9	T	C	101	41.6
316	Clavig 607	39	A	G	127	40.9
317	Clavig 607	48	A	T	137	40.1
318	Clavig 607	63	C	T	155	42.6
319	Clavig 607	84	C	T	172	40.7
320	Clavig 607	90	A	T	182	41.8
321	Clavig 607	102	C	T	173	44.5
322	Clavig 607	108	T	C	173	45.1
323	Clavig 607	113	C	G	178	43.8
324	Clavig 607	117	A	G	169	42.6
325	Clavig 607	132	T	G	198	37.9
326	Clavig 607	144	C	T	208	37.5
327	Clavig 607	152	A	C	204	36.8
328	Clavig 607	153	T	C	204	36.8
329	Clavig 607	163	G	T	199	43.2
330	Clavig 607	165	C	T	199	43.2
331	Clavig 607	172	A	T	197	40.6
332	Clavig 607	174	C	T	197	40.6
333	Clavig 607	183	T	C	223	46.6
334	Clavig 607	186	C	T	216	47.7
335	Clavig 607	250	A	T	225	36.4
336	Clavig 607	255	T	C	217	42.9
337	Clavig 614	496	T	A	46	39.1
338	Clavig 614	691	G	A	97	47.4
339	Clavig 615	291	T	G	106	47.2
340	Clavig 615	615	T	A	219	42.0

APPENDIX J3. (Cont.)

341	Clavig	625	298	T	C	117	43.6
342	Clavig	625	310	A	G	126	44.4
343	Clavig	625	602	T	C	47	42.6
344	Clavig	626	676	T	C	71	36.6
345	Clavig	626	796	T	C	65	40.0
346	Clavig	626	898	G	A	42	59.5
347	Clavig	626	1009	T	A	54	46.3
348	Clavig	626	1027	G	A	53	49.1
349	Clavig	626	1033	A	G	55	45.5
350	Clavig	626	1039	T	C	52	44.2
351	Clavig	626	1051	G	A	43	46.5
352	Clavig	626	1063	G	T	50	40.0
353	Clavig	626	1084	T	C	45	35.6
354	Clavig	626	1225	G	A	47	36.2
355	Clavig	640	1247	G	T	45	42.2
356	Clavig	643	629	C	T	43	39.5
357	Clavig	643	737	G	A	44	45.5
358	Clavig	643	794	C	T	57	47.4
359	Clavig	643	851	C	T	59	45.8
360	Clavig	670	410	T	C	88	40.9
361	Clavig	670	590	C	T	72	44.4
362	Clavig	686	414	C	A	70	50.0
363	Clavig	686	418	C	G	69	47.8
364	Clavig	686	450	A	G	71	46.5
365	Clavig	686	456	G	A	70	44.3
366	Clavig	686	719	G	A	58	36.2
367	Clavig	692	654	T	C	38	39.5
368	Clavig	692	735	A	G	53	35.8
369	Clavig	692	834	T	C	74	48.6
370	Clavig	692	1206	C	T	62	46.8
371	Clavig	692	1224	A	G	54	46.3
372	Clavig	711	436	A	G	203	40.9
373	Clavig	716	1083	A	G	40	45.0
374	Clavig	726	69	C	A	108	44.4
375	Clavig	726	74	T	C	102	42.2
376	Clavig	726	291	T	G	129	38.0
377	Clavig	745	72	T	C	71	40.8
378	Clavig	757	328	C	G	63	41.3
379	Clavig	765	446	G	A	165	43.0

APPENDIX J3. (Cont.)

380	Clavig 768	744	C	T	54	35.2
381	Clavig 768	873	A	C	67	43.3
382	Clavig 768	1245	C	T	148	50.0
383	Clavig 768	1629	T	C	60	45.0
384	Clavig 775	411	C	T	78	47.4
385	Clavig 775	421	A	G	82	37.8
386	Clavig 784	372	A	G	36	41.7
387	Clavig 799	426	A	T	193	37.3
388	Clavig 799	566	A	G	158	40.5
389	Clavig 817	348	G	A	48	43.8
390	Clavig 858	251	A	G	179	35.8
391	Clavig 861	127	C	T	41	41.5
392	Clavig 861	422	G	A	131	39.7
393	Clavig 869	2293	C	G	69	39.1
394	Clavig 891	1655	A	G	37	35.1
395	Clavig 936	413	A	T	40	42.5
396	Clavig 936	440	A	G	41	39.0
397	Clavig 936	551	T	C	38	39.5
398	Clavig 936	590	G	A	46	47.8
399	Clavig 936	593	G	A	44	50.0
400	Clavig 936	638	A	G	42	38.1
401	Clavig 936	680	G	A	40	37.5
402	Clavig 936	686	A	G	38	42.1
403	Clavig 939	101	A	G	38	47.4
404	Clavig 945	1355	T	C	43	44.2
405	Clavig 953	253	A	G	59	42.4
406	Clavig 953	475	A	T	71	47.9
407	Clavig 953	604	A	G	76	42.1
408	Clavig 954	143	A	T	61	47.5
409	Clavig 954	239	G	A	63	39.7
410	Clavig 959	59	A	G	54	48.1
411	Clavig 960	89	C	T	55	45.5
412	Clavig 960	173	G	A	51	43.1
413	Clavig 964	151	G	A	512	42.4
414	Clavig 964	322	G	A	1016	40.8
415	Clavig 964	649	C	T	334	38.9
416	Clavig 965	269	G	A	43	39.5
417	Clavig 984	1274	C	T	35	51.4

APPENDIX J3. (Cont.)

418	Clavig 989	2004	A	G	41	41.5
419	Clavig 993	435	T	C	147	46.3
420	Clavig 993	471	C	A	149	37.6
421	Clavig 993	519	T	A	152	50.0
422	Clavig 993	565	C	G	181	45.9
423	Clavig 993	566	A	T	181	45.9
424	Clavig 993	585	A	C	182	46.2
425	Clavig 993	694	T	A	197	46.2
426	Clavig 993	708	C	T	191	47.1
427	Clavig 993	810	C	T	91	38.5
428	Clavig 993	846	G	A	83	44.6
429	Clavig 996	60	A	T	48	45.8
430	Clavig 996	457	A	C	67	37.3
431	Clavig 996	692	A	T	53	50.9
432	Clavig 1003	381	T	A	43	48.8
433	Clavig 1007	391	G	C	106	41.5
434	Clavig 1029	235	G	A	36	41.7
435	Clavig 1029	385	A	G	36	47.2
436	Clavig 1038	420	T	G	81	43.2
437	Clavig 1049	981	G	A	48	45.8
438	Clavig 1049	1593	G	A	39	46.2
439	Clavig 1052	754	A	G	38	50.0
440	Clavig 1052	791	T	A	42	47.6
441	Clavig 1052	834	G	A	43	37.2
442	Clavig 1052	929	A	G	43	39.5
443	Clavig 1054	245	T	C	75	46.7
444	Clavig 1055	246	T	C	44	47.7
445	Clavig 1055	393	C	T	73	43.8
446	Clavig 1055	554	G	A	69	44.9
447	Clavig 1055	791	G	C	81	38.3
448	Clavig 1055	800	G	A	81	48.1
449	Clavig 1055	842	A	G	75	49.3
450	Clavig 1055	977	C	T	40	35.0
451	Clavig 1055	982	T	C	44	40.9
452	Clavig 1058	338	C	T	37	48.6
453	Clavig 1075	45	T	C	53	35.8
454	Clavig 1075	138	A	G	144	43.8
455	Clavig 1075	144	C	T	151	43.7
456	Clavig 1079	330	G	C	39	35.9

APPENDIX J3. (Cont.)

457	Clavig	1079	332	T	A	39	35.9
458	Clavig	1079	341	C	T	40	40.0
459	Clavig	1079	342	G	A	40	35.0
460	Clavig	1079	344	C	A	40	35.0
461	Clavig	1079	412	G	A	43	37.2
462	Clavig	1079	413	A	G	43	37.2
463	Clavig	1079	437	A	T	42	38.1
464	Clavig	1087	356	G	A	135	39.3
465	Clavig	1088	317	A	T	39	43.6
466	Clavig	1088	338	C	T	36	41.7
467	Clavig	1125	1505	C	G	35	48.6
468	Clavig	1125	1535	T	C	37	40.5
469	Clavig	1126	153	C	T	37	43.2
470	Clavig	1167	685	A	G	98	51.0
471	Clavig	1182	104	A	G	41	46.3
472	Clavig	1187	121	G	A	42	42.9
473	Clavig	1213	25	C	T	47	38.3
474	Clavig	1213	40	C	T	58	44.8
475	Clavig	1213	41	T	C	58	44.8
476	Clavig	1213	141	T	A	80	38.8
477	Clavig	1213	336	A	C	81	46.9
478	Clavig	1213	339	A	T	80	47.5
479	Clavig	1213	340	C	G	80	47.5
480	Clavig	1213	351	T	C	78	48.7
481	Clavig	1243	389	A	G	114	44.7
482	Clavig	1251	518	A	T	44	36.4
483	Clavig	1261	497	G	A	39	38.5
484	Clavig	1261	504	A	C	41	39.0
485	Clavig	1280	541	A	T	41	51.2
486	Clavig	1286	280	G	A	135	44.4
487	Clavig	1292	287	G	A	52	48.1
488	Clavig	1319	306	T	A	94	43.6
489	Clavig	1321	175	T	A	79	36.7
490	Clavig	1347	591	C	G	44	45.5
491	Clavig	1350	185	C	T	38	39.5
492	Clavig	1350	890	C	T	43	41.9
493	Clavig	1364	400	C	T	51	49.0
494	Clavig	1364	417	T	C	52	48.1

APPENDIX J3. (Cont.)

495	Clavig 1364	493	T	C	81	46.9
496	Clavig 1364	703	G	T	54	61.1
497	Clavig 1364	803	T	C	48	45.8
498	Clavig 1401	169	G	A	70	44.3
499	Clavig 1401	187	C	T	63	42.9
500	Clavig 1401	787	C	T	106	39.6
501	Clavig 1401	1120	C	T	109	45.0
502	Clavig 1418	416	A	T	77	45.5
503	Clavig 1429	404	C	G	58	36.2
504	Clavig 1429	513	T	C	75	44.0
505	Clavig 1429	552	T	C	68	51.5
506	Clavig 1429	576	A	G	70	52.9
507	Clavig 1429	582	G	C	73	49.3
508	Clavig 1429	675	C	T	45	37.8
509	Clavig 1431	152	G	A	37	48.6
510	Clavig 1431	155	C	G	37	48.6
511	Clavig 1431	170	C	T	35	48.6
512	Clavig 1431	180	A	T	37	43.2
513	Clavig 1431	186	A	G	37	40.5
514	Clavig 1431	280	C	T	37	43.2
515	Clavig 1501	344	C	T	40	50.0
516	Clavig 1509	854	A	G	57	43.9
517	Clavig 1522	201	T	A	44	43.2
518	Clavig 1522	268	G	C	82	48.8
519	Clavig 1522	557	T	C	59	49.2
520	Clavig 1573	243	A	G	63	38.1
521	Clavig 1605	324	C	T	68	44.1
522	Clavig 1605	333	A	G	65	46.2
523	Clavig 1606	177	T	C	35	37.1
524	Clavig 1626	178	T	C	49	46.9
525	Clavig 1697	363	T	C	35	45.7
526	Clavig 1708	334	C	T	44	36.4
527	Clavig 1724	95	T	C	36	36.1
528	Clavig 1724	212	A	C	36	44.4
529	Clavig 1772	759	C	T	36	47.2
530	Clavig 1782	325	A	G	56	46.4
531	Clavig 1785	786	G	A	36	36.1
532	Clavig 1785	945	T	A	44	43.2

APPENDIX J3. (Cont.)

533	Clavig 1785	1084	A	G	42	38.1
534	Clavig 1842	384	A	G	35	40.0
535	Clavig 1849	925	T	C	55	40.0
536	Clavig 1849	961	A	C	57	42.1
537	Clavig 1849	962	T	C	57	42.1
538	Clavig 1849	1006	C	T	63	41.3
539	Clavig 1849	1129	A	G	42	45.2
540	Clavig 1913	117	C	A	42	40.5
541	Clavig 1913	151	G	A	49	44.9
542	Clavig 1913	223	C	T	55	36.4
543	Clavig 1913	388	T	G	96	44.8
544	Clavig 1913	484	T	G	77	36.4
545	Clavig 1979	521	C	T	41	39.0
546	Clavig 1983	500	G	A	142	47.2
547	Clavig 2008	1727	G	A	68	41.2
548	Clavig 2035	196	C	T	38	36.8
549	Clavig 2035	304	A	G	39	38.5
550	Clavig 2045	164	C	T	62	45.2
551	Clavig 2054	112	T	C	65	36.9
552	Clavig 2054	181	T	C	79	38.0
553	Clavig 2054	283	T	C	71	43.7
554	Clavig 2054	301	T	C	75	41.3
555	Clavig 2054	331	T	C	51	39.2
556	Clavig 2054	367	G	A	70	40.0
557	Clavig 2054	406	A	G	68	47.1
558	Clavig 2054	412	T	C	81	38.3
559	Clavig 2054	601	T	A	123	46.3
560	Clavig 2054	862	T	A	88	48.9
561	Clavig 2054	925	G	T	81	49.4
562	Clavig 2054	979	C	G	46	41.3
563	Clavig 2054	980	G	C	46	41.3
564	Clavig 2077	839	A	G	42	47.6
565	Clavig 2094	251	T	C	153	40.5
566	Clavig 2094	398	T	G	163	36.2
567	Clavig 2104	413	G	A	93	43.0
568	Clavig 2132	42	T	C	102	42.2
569	Clavig 2132	53	T	C	105	52.4
570	Clavig 2132	70	C	A	107	38.3

APPENDIX J3. (Cont.)

571	Clavig 2132	79	A	G	109	41.3
572	Clavig 2132	81	T	A	109	41.3
573	Clavig 2132	91	C	G	107	35.5
574	Clavig 2132	120	G	A	106	35.8
575	Clavig 2132	129	A	T	111	42.3
576	Clavig 2132	135	T	C	111	40.5
577	Clavig 2132	138	T	A	110	40.9
578	Clavig 2132	231	T	C	88	36.4
579	Clavig 2209	139	T	C	37	35.1
580	Clavig 2326	179	G	A	54	46.3
581	Clavig 2326	245	C	T	50	42.0
582	Clavig 2326	290	A	C	48	35.4
583	Clavig 2326	572	T	C	39	38.5
584	Clavig 2422	46	T	A	77	40.3
585	Clavig 2422	65	C	T	83	39.8
586	Clavig 2422	178	A	G	71	38.0
587	Clavig 2438	194	T	C	41	46.3
588	Clavig 2438	287	T	C	50	38.0
589	Clavig 2438	335	A	G	44	38.6
590	Clavig 2677	400	T	C	40	37.5
591	Clavig 2823	384	T	C	54	46.3
592	Clavig 2823	390	G	A	55	38.2
593	Clavig 2823	399	C	T	58	48.3
594	Clavig 2828	693	C	T	39	46.2
595	Clavig 2851	256	T	C	39	46.2
596	Clavig 2851	310	A	G	45	44.4
597	Clavig 2856	1279	C	T	41	46.3
598	Clavig 2888	523	T	C	41	46.3
599	Clavig 2888	573	A	G	40	45.0
600	Clavig 2888	591	T	C	43	41.9
601	Clavig 2888	594	A	G	42	38.1
602	Clavig 2888	603	C	T	38	36.8
603	Clavig 2888	627	C	T	42	40.5
604	Clavig 2888	630	C	A	44	40.9

APPENDIX J3. (Cont.)

605	Clavig 3671	348	G	T	40	35.0
606	Clavig 4404	116	A	G	38	52.6
607	Clavig 4404	160	T	C	47	46.8

APPENDIX J4. ***M. sjostedti.***

Serial Number	Contig	Consensus Position	Consensus	Allele	Coverage	Frequency
1	Megal 1	1197	G	A	63	36.5
2	Megal 11	244	C	T	53	47.2
3	Megal 11	331	C	T	79	45.6
4	Megal 11	342	C	T	82	46.3
5	Megal 11	353	T	G	79	43.0
6	Megal 11	382	C	T	81	50.6
7	Megal 11	442	T	C	95	38.9
8	Megal 11	496	C	T	93	46.2
9	Megal 11	547	C	T	83	38.6
10	Megal 11	548	G	A	83	38.6
11	Megal 13	2980	T	C	43	46.5
12	Megal 13	3020	C	G	50	38.0
13	Megal 13	3161	A	C	90	42.2
14	Megal 13	3173	A	T	82	37.8
15	Megal 13	3174	G	C	82	37.8
16	Megal 13	3175	C	T	82	37.8
17	Megal 13	3202	C	G	69	56.5
18	Megal 13	3208	C	G	63	38.1
19	Megal 13	3209	A	T	63	38.1
20	Megal 13	3210	A	C	63	38.1
21	Megal 13	3217	G	C	84	36.9
22	Megal 13	3275	A	C	80	37.5
23	Megal 13	3303	A	C	83	36.1
24	Megal 13	4102	C	T	47	42.6
25	Megal 13	4153	T	C	49	38.8
26	Megal 13	4233	A	G	96	41.7
27	Megal 13	4262	C	A	115	35.7
28	Megal 13	4368	G	A	153	35.9
29	Megal 13	4371	C	A	125	43.2
30	Megal 13	4433	C	A	141	36.2
31	Megal 13	4448	C	A	118	38.1
32	Megal 13	4484	G	C	105	35.2
33	Megal 13	4526	A	G	77	48.1
34	Megal 13	4528	G	C	77	48.1

APPENDIX J4. (cont.)

35	Megal 13	4539	C	T	101	35.6
36	Megal 13	4540	G	A	101	35.6
37	Megal 13	4545	C	T	101	36.6
38	Megal 13	4557	G	T	100	35.0
39	Megal 13	4559	T	A	100	37.0
40	Megal 13	4561	G	A	100	35.0
41	Megal 13	4572	C	T	91	36.3
42	Megal 25	804	G	A	67	49.3
43	Megal 25	888	C	T	60	38.3
44	Megal 25	1074	A	G	65	47.7
45	Megal 25	1383	G	A	40	35.0
46	Megal 25	1419	G	A	45	42.2
47	Megal 34	2117	G	A	37	43.2
48	Megal 34	2200	T	C	40	40.0
49	Megal 47	254	A	G	168	36.3
50	Megal 49	416	C	T	42	47.6
51	Megal 49	419	A	G	42	47.6
52	Megal 49	438	T	A	39	48.7
53	Megal 68	141	C	T	229	48.9
54	Megal 68	276	T	G	241	45.6
55	Megal 68	285	G	A	233	37.3
56	Megal 72	832	T	A	163	48.5
57	Megal 72	1228	A	G	47	61.7
58	Megal 76	572	G	A	46	37.0
59	Megal 85	72	C	T	52	38.5
60	Megal 85	450	G	A	64	39.1
61	Megal 92	88	T	C	92	37.0
62	Megal 92	121	T	A	113	41.6
63	Megal 107	285	A	G	44	45.5
64	Megal 107	331	C	T	38	39.5
65	Megal 143	61	G	A	52	42.3
66	Megal 143	122	T	C	72	41.7
67	Megal 143	241	A	T	73	38.4
68	Megal 157	763	G	T	40	40.0
69	Megal 183	175	T	C	64	42.2
70	Megal 191	432	G	A	48	41.7
71	Megal 191	652	T	C	56	50.0
72	Megal 191	894	C	T	60	43.3

APPENDIX J4. (cont.)

73	Megal 191	954	T	C	67	46.3
74	Megal 193	178	G	A	256	41.4
75	Megal 199	145	A	G	45	37.8
76	Megal 199	205	G	A	75	36.0
77	Megal 202	708	A	T	50	40.0
78	Megal 202	865	G	A	66	43.9
79	Megal 202	999	T	C	45	44.4
80	Megal 202	1030	G	T	37	37.8
81	Megal 215	725	A	G	50	48.0
82	Megal 236	440	C	A	66	48.5
83	Megal 236	752	A	G	51	41.2
84	Megal 236	803	A	G	47	44.7
85	Megal 236	830	A	G	50	48.0
86	Megal 236	854	G	A	45	46.7
87	Megal 236	866	A	T	47	38.3
88	Megal 236	1043	G	T	42	50.0
89	Megal 236	1133	A	G	38	50.0
90	Megal 236	1181	A	C	40	35.0
91	Megal 236	1225	T	A	36	44.4
92	Megal 238	560	G	A	36	52.8
93	Megal 238	592	C	G	41	46.3
94	Megal 238	646	G	A	50	40.0
95	Megal 238	661	A	T	39	51.3
96	Megal 238	670	C	G	48	47.9
97	Megal 238	792	A	T	37	45.9
98	Megal 238	838	G	A	35	45.7
99	Megal 312	169	T	C	390	37.7
100	Megal 312	391	T	A	421	46.6
101	Megal 346	91	A	G	40	40.0
102	Megal 346	126	C	T	36	38.9
103	Megal 346	198	C	T	37	37.8
104	Megal 360	298	T	C	257	42.4
105	Megal 360	334	C	T	254	44.1
106	Megal 362	198	A	T	39	38.5
107	Megal 370	1151	C	G	67	44.8
108	Megal 370	1166	G	C	79	46.8
109	Megal 376	294	T	C	41	41.5

APPENDIX J4. (cont.)

110	Megal 378	257	A	G	138	39.9
111	Megal 408	33	A	T	110	41.8
112	Megal 419	288	T	A	71	36.6
113	Megal 471	219	A	G	53	43.4
114	Megal 475	306	G	A	36	47.2
115	Megal 491	113	T	A	59	35.6
116	Megal 491	386	A	G	129	42.6
117	Megal 533	374	A	G	67	44.8
118	Megal 533	537	G	T	47	36.2
119	Megal 537	100	T	C	53	35.8
120	Megal 537	249	C	G	55	38.2
121	Megal 537	360	T	C	60	35.0
122	Megal 551	319	C	T	40	47.5
123	Megal 569	351	G	A	152	43.4
124	Megal 569	421	A	C	162	38.9
125	Megal 572	72	C	T	52	44.2
126	Megal 572	139	T	A	56	42.9
127	Megal 591	676	A	G	39	48.7
128	Megal 603	159	T	C	91	40.7
129	Megal 608	259	T	C	85	43.5
130	Megal 608	289	C	T	85	44.7
131	Megal 608	433	T	A	61	39.3
132	Megal 608	441	C	T	74	47.3
133	Megal 624	201	C	T	93	36.6
134	Megal 624	222	T	C	54	40.7
135	Megal 656	198	C	T	37	43.2
136	Megal 675	270	T	C	38	42.1
137	Megal 675	294	C	T	38	42.1
138	Megal 675	342	C	T	38	36.8
139	Megal 716	340	G	A	70	47.1
140	Megal 749	508	A	C	38	36.8
141	Megal 829	133	C	T	50	40.0
142	Megal 829	148	G	A	54	38.9
143	Megal 836	491	C	T	39	38.5
144	Megal 861	17	C	T	40	35.0
145	Megal 861	121	A	C	113	46.0
146	Megal 861	285	A	G	127	35.4
147	Megal 861	432	G	A	114	44.7
148	Megal 861	465	T	C	63	39.7

APPENDIX J4. (cont.)

149	Megal 861	474	T	A	82	35.4
150	Megal 909	20	C	G	40	47.5
151	Megal 909	23	G	T	40	47.5
152	Megal 909	71	T	G	46	43.5
153	Megal 909	107	T	C	48	37.5
154	Megal 909	205	C	T	49	44.9
155	Megal 909	267	G	A	44	43.2
156	Megal 909	301	T	G	41	48.8
157	Megal 909	313	T	C	43	44.2
158	Megal 909	330	C	T	38	39.5
159	Megal 952	114	T	C	97	36.1
160	Megal 1068	346	A	T	35	40.0
161	Megal 1087	477	C	T	57	38.6
162	Megal 1114	107	T	A	44	47.7
163	Megal 1114	135	G	C	47	40.4
164	Megal 1118	625	A	T	49	44.9
165	Megal 1118	793	A	G	49	46.9
166	Megal 1263	122	C	T	49	42.9
167	Megal 1263	419	T	C	66	39.4
168	Megal 1263	587	T	A	60	35.0
169	Megal 1263	610	G	A	48	37.5
170	Megal 1348	182	G	A	43	41.9
171	Megal 1482	178	G	A	45	37.8
172	Megal 1482	256	G	T	46	39.1
173	Megal 1626	136	T	C	87	43.7
174	Megal 1634	238	G	A	55	41.8
175	Megal 1634	256	G	C	56	46.4
176	Megal 1634	354	C	T	38	42.1
177	Megal 1634	360	A	G	39	41.0
178	Megal 1634	375	C	T	40	42.5
179	Megal 1634	465	C	G	40	45.0
180	Megal 1739	401	G	A	46	43.5