COMPARATIVE GENOMICS OF SIX SPECIES OF DELTOCEPHALINE LEAFHOPPERS TO DISCOVER VECTOR COMPETENCE-RELATED GENES

By

CHRISTIAN ORLANDO AYALA-ORTIZ

Bachelor of Science in Biotechnology Engineering

University of the Armed Forces – ESPE

Sangolqui, Ecuador

2016

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2019

COMPARATIVE GENOMICS OF SIX SPECIES OF DELTOCEPHALINE LEAFHOPPERS TO DISCOVER VECTOR COMPETENCE-RELATED GENES

Thesis Approved:

Dr. Astri Wayadande

Thesis Adviser

Dr. Trenna Blagden

Dr. Stephen Marek

AKNOWLEDGEMENTS

I would like to express my immense gratitude to all the people that helped me in the completion of my Master's program. I would like to thank my advisor Dr. Astri Wayadande for giving me the opportunity to join Oklahoma State University and for her continuous support, guidance and encouragement throughout all my research and thesis writing.

I would like to extend my deepest thanks to the other members of my advisory committee: Dr. Trenna Blagden who was always there to support me and giving me advice that would help me to move forward with my project. And Dr. Stephen Marek, for his guidance and insightful comments.

A special thanks to the professors, students and staff of the Department of Entomology and Plant Pathology at Oklahoma State University for providing a pleasant yet challenging environment that encouraged me to keep improving myself.

I would like to thank my fiancée, Viviana Freire, whose invaluable love, support, help and patience have been essential to reach all my goals and without whom none of this will make sense. Last but not least, I would like to thank my parents Orlando Ayala and Elizabeth Ortiz and my brother Daniel Ayala for all their support and encouragement to go abroad and keep studying.

Name: CHRISTIAN ORLANDO AYALA ORTIZ

Date of Degree: DECEMBER, 2019

Title of Study: COMPARATIVE GENOMICS OF SIX SPECIES OF DELTOCEPHALINE LEAFHOPPERS TO DISCOVER VECTOR-COMPETENCE-RELATED GENES

Major Field: ENTOMOLOGY AND PLANT PATHOLOGY

Abstract: Insects have been widely used in genomics research and reference genomes and transcriptomes are available for several model species. However, for non-model organisms, such as leafhoppers, this kind of information is still lacking. The Deltocephalinae (Hemiptera: Cicadellidae), is a leafhopper subfamily that comprises several vectors of economically important plant pathogens, as well as non-vector species. The ability of some species to acquire and transmit a pathogen is called vector competence. Since genomics studies can lead to insight into the mechanisms underlying biological processes, comparisons between the genomes and transcriptomes of vector and non-vector leafhoppers can lead to a better understanding of the genes involved in vector competence. However, identifying those genes is not enough to understand their function. Functional studies, involving the mutation or silencing of those genes are required to fully understand the biological processes in which they are involved. During this study the draft genomes and transcriptomes of six species of Deltocephaline leafhoppers, four vectors and two non-vectors, were sequenced, annotated and compared looking to discover vector-competence related genes. Draft genomes were assembled using MEGAHIT and transcriptomes were assembled using Trinity. The resulting draft genomes had sizes between 0.85 and 1 Gbp and completeness ranging from 53% to 77% according to the Arthropoda BUSCO genes. Genomes were annotated using PASA and transcriptomes were annotated using the Trinotate pipeline. A clustering analysis of the peptide sequences obtained from the transcriptomes identified actin, ATP synthase, vinculin, paramyosin and other genes as possible candidate genes involved in vector competence and insect vector transmission. Furthermore, the use of RNAi as a tool to study gene function was demonstrated by knocking down the laccase-1 gene of *Dalbulus maidis*. Third instar nymphs of *D. maidis* were fed 50 μL of laccase-1 dsRNA for six consecutive days. A reduction in the relative expression of the laccase-1 transcript was observed at day 4 and 6. However, because none of the insects survived and developed into adults, more analyses are required to confirm the RNAi silencing of this gene.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

Insects have been recognized as the largest class of animals present in our planet, representing approximately the 53% of all the known living species (Wilson, 1992). They have a great impact in several human affairs. Some can be beneficial, such as bees, flies, beetles and other insect orders that pollinate around 70% of the world's crops (Fijen et al., 2018); while others are detrimental and can be considered pests. It has been reported that some species of insects can reduce agricultural yields 10%-16% before harvest, and keep producing similar losses during post-harvest (Bradshaw et al., 2016).

The Hemiptera are an order of insects that consists of aphids, thrips, booklice, whiteflies, scale insects, planthoppers, leafhoppers, cicadas, true bugs among other. They can be devastating pests of crops because they can have a wide host range, are able to reproduce quickly and can be vectors of several plant pathogens (Perilla-Henao and Casteel, 2016). In order to penetrate their host tissues, hemipterans use modified mouthparts composed of two mandibular and two maxillary stylets. Their ability to acquire and inoculate plant pathogens is closely related to their feeding mode and target tissue (Mitchell, 2004).

The leafhoppers are plant feeding hemipterans that belong to the family Cicadellidae. This is one of the largest families of insects (Dietrich et al., 2001), and it includes more than 20,000 described species. Many of the most important vectors of plant pathogens are part of this family (Takiya et al., 2006). They feed primarily from plant vascular fluid using piercing/sucking mouthparts, although members of one subfamily, the Typhlocybinae, feed primarily from mesophyll tissue. All leafhoppers oviposit inside living host plant tissues (Dietrich, 1999). The most current taxonomic revision divides the Cicadellidae into around 43 subfamilies (Dietrich, 2006).

One Cicadellid subfamily, the Deltocephalinae contains many insect vectors of economically important crops (Chakravarthy, 2015), so correct identification and determination of their phylogenetic relationships is important. Despite efforts by several systematists, the classification and relationships among many genera are not yet well defined (Zahniser and Dietrich, 2013). Recent advances in genetic studies now permit systematists to use sequence data in addition to morphological characteristics to understand phylogenetic relationships. However, genomic data of the Deltocephaline leafhoppers is currently limited (Du et al., 2017b).

Other research fields, like biomedical research, have moved forward using data obtained from next generation sequencing. Similarly, the study of genomics can expand our understanding of insect vector biology and improve our ability to manage and control insects that are a risk to our food supply and economic security (Robinson et al., 2011).

The decreasing costs of next generation sequencing (NGS) technologies are resulting in the sequencing of more and more organisms every day. This includes, not

only model organisms, but also some species that have not yet been extensively studied. Throughout the last several years, there has been an increase in the number of genomic sequences available for non-model organisms. Access to this data has revolutionized the study of life sciences and pushed forward our understanding of several fields such as evolutionary biology or functional genomics by uncovering biological patterns that were previously unknown (Ellegren, 2014).

For insect vectors, NGS technologies have become a versatile and strong tool that allows researchers to analyze these genomes and transcriptomes to investigate the biological mechanisms that drive the transmission of pathogens, which factors can influence such transmission, and even try to determine whether an insect is a vector or not (Kaur et al., 2016). Moreover, all entomological research has benefitted from the availability of the many genomes and transcriptomes of insects (Yin et al., 2015).

Comparative genomics is a discipline that becomes more powerful as more sequence data accumulates (Hardison, 2003). Thus, it is important to keep generating genomic data for many different organisms and make it available to be used in comparative studies. For example, it has been inferred that the ability to effectively acquire and transmit a pathogen is genetically regulated in insect vectors (Cassone et al., 2014a; Gray et al., 2007). Hence, the comparison of the genomes from vector and non-vector species can produce some information about the genes and processes involved in vector competence allowing us to better understand the genetic differences and/or similarities among vector and non-vector species (Welch et al., 2011).

The information provided through the construction of high-quality reference genomes for insects can be very valuable to identify genes and pathways that may be associated with pathogen transmission (Kaur et al., 2016). With the many insect species that are being sequenced these days, a large collection of insect genes with unknown function have been identified and need to be studied, in order to decipher these genes' functions (Bellés, 2009).

Functional genomics was initially studied by choosing a function of interest and then finding which gene(s) determined that function (a process known as forward functional genomics). Now, the availability of genomic data has presented researchers with many genes of unknown function, so it is possible to choose a gene of interest and then study its function (a process known as reverse functional genomics). In this regard, *Drosophila melanogaster* became a model species because of ease of genetic transformation through conventional techniques. Thus, it allows the study of mutant phenotypes (Bellés, 2009).

For organisms that are not easily transformed other techniques are required to link a function to a particular gene. A technique that can be useful for this task is the RNA interference (RNAi) mechanism (Scott et al., 2013). RNAi is a tool that can be used to knock-down a gene in order to analyze its function. It has been widely used especially when it is not feasible to generate mutants, such as in non-model organisms or nontransformable species (Tomoyasu et al., 2008).

This mechanism starts with the delivering of double-stranded RNA (dsRNA) which will be cleaved to produce small RNAs, into the test organism. The small RNAs

will then be used by the RNA-induced silencing complex to locate target mRNAs through base pairing (Hammond, 2005). RNAi is a highly versatile tool that is used to silence the expression of a particular gene; thus any observed phenotype can be directly associated to the DNA sequence used as a template for the synthesis of the delivered dsRNA (Tomoyasu et al., 2008).

The gene silencing effect is sequence specific and not locus specific, so mRNAs from multiple loci or alleles can be silenced at the same time by designing dsRNAs targeting conserved regions in common to all of them. However, RNAi also can be used to silence only a particular isoform of a gene by designing a more specific dsRNA. Since RNAi produces a partial and quantitative reduction in the expression of the target genes, it can be used to discover previously unknown functions for genes that have lethal effects when knocked out (Bellés, 2009).

In conclusion, insects and their arthropod relatives are the most diverse branch of the metazoan organisms with millions of described species. Thus, it is important to keep obtaining their genomic sequences. However, at the same time it is necessary to take one step further and use the data collected to try and assign functions to newly discovered genes. For that reason, this research includes three main objectives.

The first objective is to provide an initial step to generate reference genomes and transcriptomes to a group on non-model organisms, the Deltocephaline leafhoppers. The species to be studied include four vectors of economic relevance: *Dalbulus maidis*, *Graminella nigrifrons* and *Macrosteles quadrilineatus* as well as two non-vector species of interest: *Balclutha neglecta* and *Balclutha rubrostriata*. The second objective is to

compare the genomes and transcriptomes of those vector and non-vector leafhoppers in order to identify possible genes that may be related to vector competence. To complement this work, the third objective involves the use of interference RNA (RNAi) to silence the laccase-1 gene in *Dalbulus maidis*, as a proof of concept of the mechanisms to determine its utility in the functional genomics of the leafhoppers.

CHAPTER II

LITERATURE REVIEW

OVERVIEW OF THE LEAFHOPPER SUBFAMILY DELTOCEPHALINAE

Leafhoppers (Hemiptera: Cicadellidae), belong to a family of piercing sucking insects that feed from the fluids of vascular plants. This group of insects in one of the most abundant and diverse inhabitants of the grasslands (Hamilton and Whitcomb, 2010). Leafhoppers can be found in many different habitats, however these insects are most abundant in the grasslands and are one of the most dominant groups of prairie herbivores. The association of certain subfamilies and tribes of leafhoppers with the grasslands has seemed to have played an important role in their evolution and diversification (Dietrich, 1999). The Cicadellidae consists of 33 subfamilies (Dietrich, 2005). Looking into the phylogenetic tree of the Membracoidea (leafhoppers and treehoppers) it is possible to observe that several of the tribes and subfamilies that have species able to transmit plant pathogens tend to cluster together at the bottom of the tree (Figure 1), suggesting that they may share a common feature involved in vector competence.

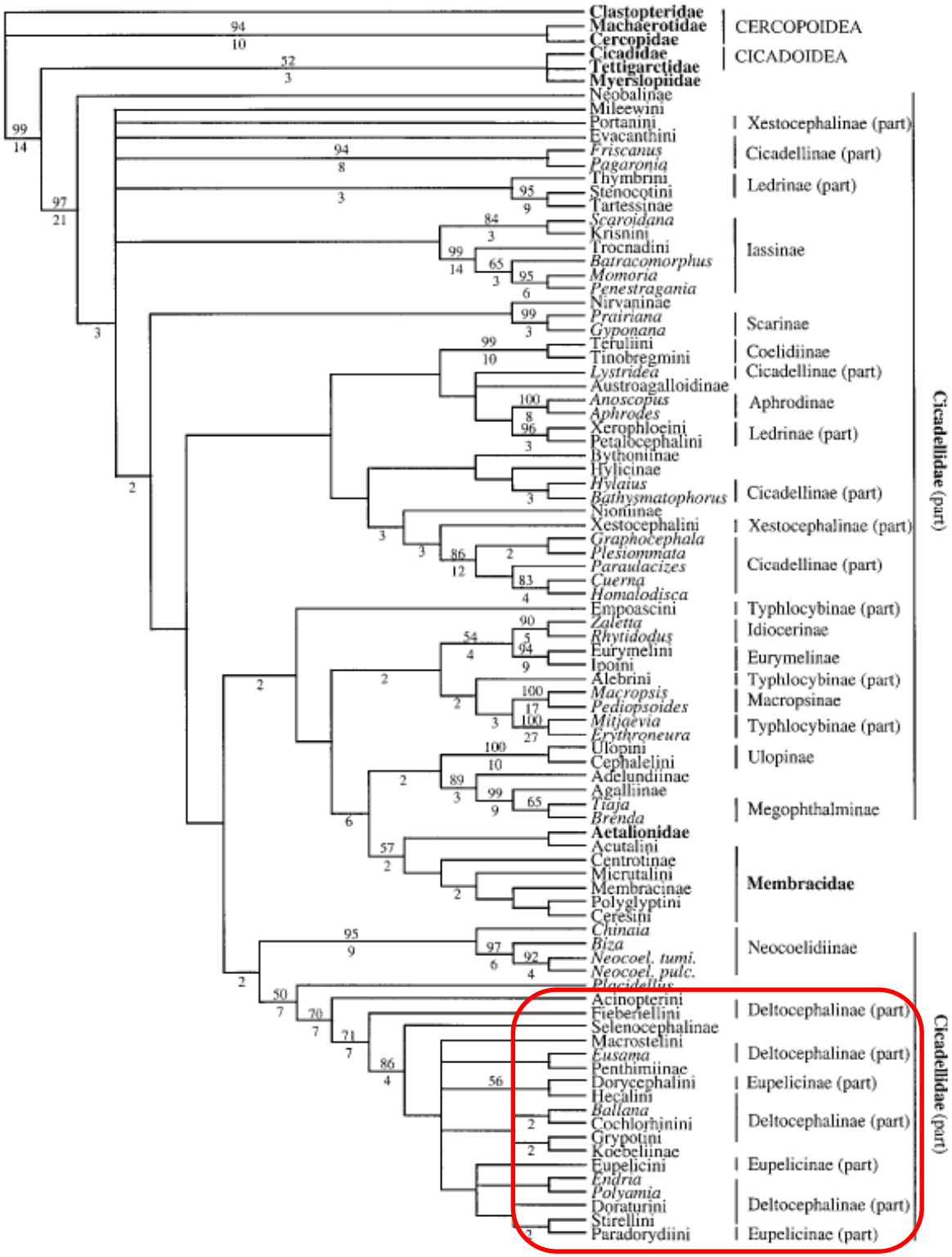

Figure 1. Maximun-parsimony trees of several tribes within the Membracoidea (leafhoppers and treehoppers) based on their 28S sequences (Dietrich, et al., 2001).

One of this subfamilies, the Deltocephalinae, is the largest and most diverse subfamily of the Cicadellidae based upon the number of described species. It consists of 6683 valid species divided in 923 genera and 39 tribes (Zahniser, 2017). The classification and phylogeny of Deltocephalinae has been subject to many changes and revisions over the last 50 years (Zahniser and Dietrich, 2013). Many classifications have been proposed by various authors; the one proposed by Oman, *et al* (1990) is the most accepted and has formed the base for some of the more recent studies. Two main reasons for this lack of a stable classification are: most studies have focused on regional faunas which do not provide a big enough scope to understand the phylogenetic relationships of these worldwide distributed leafhoppers (Zahniser, 2008), and an apparent lack of morphological characters useful in understanding the relationships between leafhopper species (Dietrich et al., 1997).

In spite of challenges related to morphologically identifying leafhoppers, some authors (Dietrich and Dmitriev, 2003; Dietrich and Rakitov, 2002) use the morphological methodology proposed by Oman, *et al* (1990), noting that not all of the features are present in all Deltocephalinae and that some are not unique to this subfamily. Based on phylogenetic studies using molecular and morphological data, the subfamily Deltocephalinae, as defined by Oman, *et al* (1990), was found to be paraphyletic (Dietrich et al., 2001; Zahniser, 2008; Zahniser and Dietrich, 2008, 2010), which means that all the members evolved from the same ancestor, but that not all its descendants are part of this subfamily (Donoghue and Cantino, 1988). This indicates that some of the other subfamilies within Cicadellidae were derived from the Deltocephalinae, to the point that in recent taxonomic studies some subfamilies such as Eupelicinae, Koebeliinae,

Paraboloponinae, Penthimiinae and Selenocephalinae were treated as synonyms of Deltocephalinae, meaning that these studies consider those other subfamilies as part of the Deltocephalinae (Zahniser and Dietrich, 2010)

Members of the Deltocephalinae can be found worldwide. Many tribes are cosmopolitan, such as Balcluthini or Macrostelini (Zahniser and Dietrich, 2008), while others are locally distributed. Deltocephaline leafhoppers can inhabit tropical, subtropical and temperate forest ecosystems. However, they are usually more known for their diversity and abundance in grasslands (Biedermann et al., 2005). Because of their great numbers and the functional importance of Deltocephalinae in their ecosystems, this subfamily represents an appropriate group to study as indicators of grassland biodiversity, conservation and quality (Biedermann et al., 2005; Hamilton, 1995) The close association of some leafhopper tribes with the grasslands suggests that this subfamily first appeared during the Oligocene, in Asia or North America where grasslands first originated (Dietrich, 1999).

Similar to other Cicadellid leafhoppers, the Deltocephalinae possess piercing-sucking mouthparts that allow them to feed from plant vascular tissue (Backus, 1988). Their mouthparts include a needle-like stylet bundle composed of two mandibular and two maxillary stylets that are interlocked forming a pair of canals. The narrower salivary canal is used to deliver saliva into the plant tissues while the wider food canal takes up plant sap into the cibarium (Ammar, 1985). More accurately, they are a subfamily of primarily phloem-feeders (Wayadande, 1994). Phloem sap is a nutrient-rich substance that has a high concentration of sugars, that can be used as a source of carbon and energy, and nitrogen in the form of free amino acids. Generally, it is free of toxins and any other

kind of feeding deterrents (Douglas, 2006). Despite this, phloem sap it is only used as a dominant or sole diet by only one group of organisms, the Hemiptera. The reason is that phloem sap present two major nutritional problems to other groups of animals: the "nitrogen barrier" and the "sugar barrier" (Douglas, 2006).

The "sugar barrier" occurs because of the high concentration of sucrose that is present in the phloem fluid. The concentration often exceeds 1M, resulting in an osmotic pressure up to five times higher than the osmotic pressure inside the insect's body (Douglas, 2006). Phytophagous hemipterans overcome this barrier through osmoregulation by transforming excess sugar into long-chain oligosaccharides that are eliminated as honeydew. This is attributed to the presence of enzymes with sucrose-transglucosidase activity in the guts of some hemipterans (Ashford et al., 2000)

In the case of the "nitrogen barrier" it refers to the "low" nitrogen quality of the phloem sap. The development of phytophagous insects depends on the availability of enough quantity of nitrogen as well as in its quality, or composition (Douglas, 2006). Animals in general lack the ability to synthetize nine of the 20 amino acids, usually called "essential" amino acids (Wu et al., 2013). In this regard, phloem sap has low nitrogen quality because its ratio of essential amino acids: non-essential amino acids is 1:4-1:20, which significantly lower than the usual ratio of 1:1 in animal protein (Douglas, 2006). For that reason, Deltocephaline leafhoppers like other phytophagous hemipterans, are host to obligate symbiotic microorganisms, either bacteria or yeast-like symbionts, that have the ability to synthetize those amino acids that are not being supplied by their diet (Baumann, 2005).

Previous studies have demonstrated that hemipterans harbor at least two obligate endosymbiotic bacteria (Buchner, 1965). These bacteria are usually located inside host cells called bacteriocytes which aggregate together to form bacteriomes and are vertically transmitted from the mother to the egg (Nault and Rodriguez, 1985). Typically, the endosymbionts have been classified into primary symbionts and accessory symbionts (Baumann, 2005; Buchner, 1965). However, there are times in which two or more endosymbionts are obligate and ancient bacteriome-associated. For those situations, the use of the term co-primary symbionts has been suggested (Takiya et al., 2006)

Primary endosymbionts originated from the infection of the ancestor of a group of insects, so they are present in all members of a particular taxa (Kobiałka et al., 2018a). Molecular studies have confirmed the role of this organisms in supplying their host insects with the amino acids, cofactors and vitamins that are missing in their diet (Baumann, 2005).

Secondary symbionts have a more recent origin. In consequence, some taxa or even some populations may possess different secondary symbionts (Kobiałka et al., 2018a). These endosymbionts can be present in the bacteriocytes, in other insect cells or free in the hemolymph. They can be transmitted transovarially but also horizontally between insects of the same population (Kobiałka et al., 2016). The function of these secondary endosymbionts is still being studied; research on aphids suggest that they may provide the insect with protection against heat stress, parasitic hymenopterans or pathogenic fungi (Łukasik et al., 2013; Montllor et al., 2002; Oliver et al., 2003).

Deltocephaline leafhoppers possess at least two co-primary symbionts. The first is an ancient symbiont of all auchenorrhynchans, a member of the Bacteroidetes called *Candidatus Sulcia muelleri*, which is believed to have infected an ancestor of the Auchenorrhyncha more that 260 million years ago (Moran et al., 2005). The other symbiont is a member of the class Betaproteobacteria, however the exact species varies between different groups of insects. For the Deltocephalinae, the betaproteobacterial symbiont is *Candidatus Nasuia deltocephalinicola* (Kobiałka et al., 2018b).

Most of the species of Deltocephaline leafhoppers feed on herbaceous or woody dicotyledonous plants, while 14 of the 39 tribes use only grasses or sedges as food source (Zahniser, 2017). Their host range varies from one host to multiple plant species (Maggi et al., 2014; Zahniser and Dietrich, 2010). The range of specialization goes from species capable of feeding from multiple kinds of plants (polyphagy) to some species that feed from: a single plant family (oligophagy), closely related genera of plants (stenophagy), or that only have one host species (monophagy). Usually, single plant specialization comes together with physical adaptations to the host plant, as an example *Attenuipyga minor* (Osborn) is shaped like the seeds of wheat grass that it inhabits, or *Flexamia areolate* (Ball) that have the same color pattern of its host seed, purple love grass (Hamilton and Whitcomb, 2010).

The chemical and nutritional qualities of the plant, upon which the insect hatches and develops, can have some effects in the morphology of polyphagous leafhoppers, such as changes in size and pigmentation among insects of the same species recovered from different hosts (Gillham and Claridge, 1994). In agricultural settings, polyphagous leafhoppers can be considered a bigger concern because they can acquire different

pathogens from plant vegetation near the crops and introduce a new disease cycle (Trivellone et al., 2017). For example, the beet leafhopper *Circulifer tenellus*, vector of the beet leafhopper-transmitted virescence agent (BLTVA) phytoplasma (Munyaneza et al., 2007) that causes potato purple top disease, has been found to be involved in disease outbreaks in potato crops in Washington and Oregon, as well as several other crops including dry beans, carrots and radish (Munyaneza et al., 2010)

LEAFHOPPERS AS VECTORS OF PLANT PATHOGENS

There are two main classes of plant pathogens that can be transmitted by Deltocephaline leafhoppers: plant viruses and mollicutes.

Plant viruses

Viruses can be defined as a set of one or more nucleic acid molecules, usually encased in a protective coat of proteins or lipoproteins that are able to organize their own replication within a suitable host cell (Hull, 2013). In the case of plant viruses, more than 2000 species from at least 21 families have been described, with many more being discovered everyday (Whitfield et al., 2015).

Unlike animals, plants cannot move on their own, so viruses and other pathogens need an effective way to move between host plants. In that regard, most of the plant pathogens use biological vectors to replicate and be transmitted (Whitfield et al., 2015). Organisms capable of being vectors include fungi, nematodes and several kinds of invertebrates, however insects are the largest class of plant pathogen transmitting vectors that exist (Dietzgen et al., 2016). Among the insects, leafhoppers (Cicadellidae) are one of the largest and best characterized group of vectors (Bragard et al., 2013).

The subfamily Deltocephalinae comprises most of the leafhoppers vectors of economically important plant pathogens (Zahniser and Dietrich, 2013). Around 60% of leafhopper vector genera are located inside Deltocephalinae, and they transmit about 70% of the known phytopathogenic agents present in leafhoppers (Dakhil et al., 2011; Maramorosch, 2014; Nielson, 1979). Some of these vectors include *Graminella nigrifrons* (Forbes), *Dalbulus maidis* (DeLong and Wolcott), *Macrosteles quadrilineatus* (Forbes), *Exitianus exitiosus* (Uhler), among others.

Of all of the recognized subfamilies of Cicadellidae, only eight of them have species that are vectors of plant pathogens, these subfamilies are: Agallinae, Macropsinae, Coelidiniiae, Aphrodinae, Gyponinae, Cicadellinae, Typhlocybinae and Deltocephalinae (Nault and Ammar, 1989; Nielson, 1968). The subfamily Deltocephalinae is the most taxonomically diverse group of leafhoppers that can transmit phytopathogenic viruses or mollicutes (Redak et al., 2003). They are also one of the only two subfamilies of Cicadellidae, the Agallinae and Deltocephalinae, that have species capable of transmitting viruses (Nault and Ammar, 1989).

Insect vector transmission of plant viruses can be described in different modes: non-persistent, semi-persistent and persistent; with the later further dividing into circulative and propagative transmission (Nault and Ammar, 1989). The window for acquiring and then inoculating viruses after feeding on an infected plant vary from seconds to minutes, hours to days and days to weeks, depending on the mode of transmission (Dietzgen et al., 2016; Nault, 1997; Nault and Ammar, 1989). Similarly, the amount of time the viruses stays inside the insect body, is also related to the mode of transmission (Ng and Falk, 2006). Since the number of experimental insects can greatly

influence the apparent minimal and maximal amount of time a virus remains in the vector, it has been recommended to use the retention half-life of viruses. Therefore, the retention half-life for non-persistent viruses is minutes, for semi-persistent viruses it is hours and for circulative and propagative viruses the retention half-life is days, weeks or months (Nault, 1997). The specificity of the virus is also variable. Some can be transmitted by several insect species, like the *Cucumber mosaic virus* which is transmitted by different species of aphids (Whitfield et al., 2015), while others are highly specific such as the Rice dwarf virus (RDV), a plant reovirus transmitted by *Nephotettix cincticeps,* even though it can be transmitted ineffectively by the leafhopper *Recilia dorsalis* (Chen et al., 2015).

In non-persistent transmission, viruses are retained in the stylet of the insect. Even though, there are some similarities between the mouthparts and the feeding behavior of various Hemipteran vectors, only aphids have been found to be capable of transmitting plant viruses in a non-persistent mode. Non-persistent viruses are usually acquired while the insect is probing to find whether or not it has encountered a host plant. During this time epithelial cells are ruptured exposing the stylets of the insects to their contents (Ng and Falk, 2006). These viruses use one of two strategies available for their transmission: capsid strategy and helper strategy (Blanc et al., 2014; Ng and Falk, 2006; Whitfield et al., 2015). For the capsid strategy, only the viral coat protein (CP) and conserved capsid surface domains are required for a successful transmission (Chen and Francki, 1990; Liu et al., 2002). For the helper strategy, other viral proteins in addition to complete virions are required to be acquired by the insect in order for viral transmission to occur. These

other components have been called “helper components” or “virus-induced factors” (Whitfield et al., 2015).

Leafhoppers mostly transmit semi-persistent and persistent viruses (Ng and Falk, 2006). Some examples are: Rice tungro bacilliform virus and Rice tungro spherical virus which are transmitted by *Nephotettix virescens* (Distant) in a semi-persistent manner (Gálvez, 1968; Hibino et al., 1979); the Maize chlorotic dwarf virus (MCDV) which is mainly transmitted by *Graminella nigrifrons* (Nault et al., 1973a; Wayadande and Nault, 1993); or the Rice dwarf virus which is a propagative virus transmitted by *Nephotettix cincticeps* (Uhler) (Nakasuji and Kiritani, 1970; Nasu, 1963).

Unlike non-persistent viruses, semi-persistent viruses are usually acquired from and inoculated into the phloem. They require longer acquisition periods than non-persistent viruses, and can be transmitted by insects other than aphids such as leafhoppers and whiteflies (Ng and Falk, 2006). Similarly, these viruses show longer retention times than non-persistent plant viruses but they are lost during the vector molting (Whitfield et al., 2015). Studies have shown that semi-persistently transmitted viruses are retained in the foregut of the insect (Nault and Ammar, 1989). Because of their localization inside the host vectors, transmission of these viruses require that they detach from the foregut cuticle and food canal, so they can be expelled during extravasation (Wayadande and Nault, 1993)

Persistent viruses can be divided into two categories: circulative and propagative viruses. The main difference between these two categories is that while circulative viruses do not seem to replicate inside the host vector, propagative viruses do replicate.

Even more, many propagative viruses are transmitted to the progeny by infecting embryos or germ cells inside the female insect body (Nault and Ammar, 1989).

Persistently-transmitted viruses enter the insect body and disseminate from the gut lumen to the hemolymph or other tissues until finally reaching the salivary glands. From there they can be inoculated into the host plant when the insect feeds again (Hogenhout et al., 2008a). Persistent transmission of viruses may be affected by physiological and anatomical barriers. Four types of barriers have been proposed: (1) midgut infection barrier, (2) dissemination barrier, (3) salivary gland escape barrier and (4) transovarial transmission barrier (Ammar, 1994).

Mollicutes

Even though most of the plant pathogenic prokaryotes do not need an insect vector to be transmitted, some plant pathogenic bacteria do, such as the phytoplasmas and spiroplasmas (Purcell, 1982). Both of these group of organisms belong to the Mollicutes, a class of bacteria characterized for their lack of cell wall. *Phytoplasma* is a genus of non-cultivable, gram-positive prokaryotes that are plant parasites limited to the phloem, thus can only be acquired and transmitted by phloem-feeders (Weintraub and Beanland, 2006). Similarly, *Spiroplasma* is a genus of cultivable, motile, helical prokaryotes without a cell wall that can cause disease in several organisms including plants (Wayadande et al., 1997).

The transmission of mollicutes is done in a persistent and propagative manner. Similar to propagative viruses, mollicutes navigate across the wall of the gut, colonize the hemolymph and finally reach the salivary glands from where they replicate and can be

inoculated back into a host plant (Wayadande and Fletcher, 1995). Mollicutes may have a limited degree of host-pathogen specificity. This means that single species of leafhopper may transmit one or a few mollicutes, but a specific mollicute may be transmitted by one or few species of leafhoppers (Fletcher et al., 1998).

Phytoplasmas are plant wall-less prokaryotes that are the causal agent of hundreds of plant diseases worldwide. They are capable of infecting a variety of plants including several economically important crops. Infection by phytoplasmas causes a number of symptoms in the plant that are indicative of an abnormal balance in the host hormones. These symptoms include witches' brooming, proliferation, virescence and phyllody as well as stunting and a general decline of the host (Christensen et al., 2005; Lee et al., 2000).

As obligate symbionts of insects and plants, phytoplasmas usually require the presence of both hosts to be able to survive and spread in nature (Hogenhout et al., 2008b). Phytoplasmas are acquired by insects during phloem feeding. The acquisition access period (AAP) is the feeding time necessary to acquire enough quantity of phytoplasmas. It can be a few minutes, but is usually in the range of hours (Weintraub and Beanland, 2006). The period between the acquisition of phytoplasmas and their transmission is called the latent period (LP). It can vary between 7 and 80 days depending on the phytoplasma and plant species (Hogenhout et al., 2008b). During the LP, phytoplasmas multiply in the midgut, then invade the haemocytes and other tissues, before finally reaching the salivary glands. If the insect cannot sustain phytoplasma replication, it does not serve as a vector. However, even if the phytoplasmas can multiply inside the insect, it does not always mean that they are competent vectors (Bosco et al.,

2007). In order to be transmitted, phytoplasmas need to reach and accumulate to high levels in the posterior acinar cells inside the salivary glands (Brown et al., 2007). From here, the phytoplasma must pass through three barriers to be expelled with the saliva: the basal lamina, the basal plasmalemma and the apical plasmalemma. Failure to overcome any of this barriers will prevent phytoplasma transmission (Wayadande et al., 1997; Wayadande and Fletcher, 1995).

Auchenorrhyncha is the most successful insect suborder of phytoplasma vectors. The leafhopper subfamily Deltocephalinae contains 75% of all known phytoplasma vector species to date. These high occurrence of vectors within one subfamily may be due to the fact that many of the known phytoplasma diseases are from economically important crops of the family Poaceae, whose members are host to many species of Deltocephaline leafhoppers (Wilson and Weintraub, 2007). Most of these leafhoppers feed from phloem cells in a non-destructive manner (Markham, 1982) and phytoplasmas can propagate and persist in their bodies. Moreover, auchenorrhynchans have mechanisms to pass obligate symbiotic prokaryotes to their offspring that may be used by phytoplasmas for transmission purposes (Weintraub, 2007).

Spiroplasma is the other genus of mollicutes that can be transmitted by leafhoppers. Most of the spiroplasmas are found as commensals of insects, arachnids, crustaceans or plants, however some species are pathogenic to these same organisms (Cisak et al., 2015). Most of the knowledge about plant pathogenic spiroplasmas comes from studies in *Spiroplasma citri* and the corn stunt spiroplasma, *Spiroplasma kunkelii*. Both pathogens can infect monocotyledonous and dicotyledonous plants (Markham, 1983). However, their host range maybe restricted by the preferred hosts of their main

vectors, *Circulifer tenellus* (Liu et al., 1983), and *Dalbulus maidis* (Alivizatos and Markham, 1986), respectively.

LEAFHOPPER SPECIES TO BE STUDIED IN THIS WORK

Balclutha neglecta

The genus *Balclutha* is one of the most common and abundant group of leafhoppers found in the grasslands. It comprises 111 species including *Balclutha neglecta*, all of which live on grasses and sedges and are widely distributed around the world (Knight, 1987; Zahniser et al., 2010). Even though this genus is easily distinguished from other deltocephaline leafhopper genera, species within the genus *Balclutha* are similar to each other presenting some problems for identification (Knight, 1987). Even though there are no records of *Balclutha* species being capable of transmitting viruses (Knight, 1987), *Balclutha neglecta* status as a vector can be considered undetermined (Gahm, 2017). There is at least one study that found phytoplasma DNA in samples of this species (Olivier et al., 2011), however, its vector competency has not been scientifically resolved.

Balclutha rubrostriata

Balclutha rubrostriata is commonly called the red streaked leafhopper because of the characteristic bright red coloration of it forewings (Knight, 1987). It is an invasive species native to Sri Lanka and India that has spread to Australia, Asian Islands, Southeast Asia, Japan, the Mediterranean, various African countries, Puerto Rico, Central America, Hawaii and more recently into the continental US (Andreason et al., 2015; Knight, 1987; Morgan et al., 2013; Zahniser et al., 2010). Like *B. neglecta*, *B.*

rubrostriata vector status is undetermined. Even though it is not considered to be a competent vector, 30% of the individuals sampled for a study in Thailand tested positive for the presence of the sugarcane white leaf phytoplasma (Hanboonsong et al., 2006). This same study did not test white leaf phytoplasma transmissibility by *B. rubrostriata*.

Dalbulus maidis

Dalbulus maidis, also known as the corn leafhopper, is distributed across all the American continent from southern US, including California, to Argentina (Nault, 1990). It is believed that its wide dispersal is due to its association with cultivated maize (de Oliveira et al., 2004). The corn leafhopper is a specialist insect whose host range is restricted to maize and its relatives (Tsai, 2008). It is considered one of the most important pests of the maize in South America (de Oliveira et al., 2007) because it can transmit viral and mollicute agents to corn, like corn stunt spiroplasma (CSS), maize bushy stunt phytoplasma (MBSP) (Nault, 1980) and maize rayado fino virus (MRFV) (Gámez-Lobo, 1969; González and Gámez-Lobo, 1974).

Graminella nigrifrons

Graminella nigrifrons, the black-faced leafhopper, is a common and abundant leafhopper present in at least 35 states (Chen et al., 2012). It can feed from a variety of host grasses and cereal crops, including: ryegrass, johnsongrass, maize and oats (Cassone et al., 2014b). This insect is a vector of several pathogens such as maize bushy stunt phytoplasma (MBSP), corn stunt spiroplasma (CSS) (Nault, 1980), and maize chlorotic dwarf virus (MCDV) (Nault et al., 1973b; Wayadande and Nault, 1993). It is also the

only known vector of maize fine streak virus (MFSV) (Cassone et al., 2014a; Redinbaugh et al., 2002).

Macrosteles quadrilineatus

Macrosteles quadrilineatus, the aster yellows leafhopper, is considered the primary vector of the aster yellows phytoplasma (AYp) (Kirkpatrick and Smart, 1995; Murral et al., 1996). It is considered a serious pest of vegetable crops because of its ability to transmit phytoplasmas. This is a polyphagous leafhopper that can feed on more than 300 plant species for its survival, many of which are susceptible to AYp (Frost et al., 2011). These migrant leafhoppers move from the south, as far as Texas, into the upper Midwestern states and Canada. Leafhoppers can acquire AYp before or during the migration, and usually they are able to infect crops as soon as they arrive to their migratory destination (Beanland et al., 2005).

Exitianus exitiosus

Exitianus exitiosus, the gray lawn leafhopper, is a pest of cultivated crops that is distributed in the low lands of North and Central America (DeLong and Hershberger, 1947). It feeds on different species of grasses including *Avena sativa, Hordem vulgare*, *Triticum aestivum* (Gustin and Stoner, 1973) and *Cynodon dactylon* (Buntin, 1988). It has been described as one of the vectors of maize chlorotic dwarf virus (MCDV) (Nault and Madden, 1988) as well as an experimental vector of *Spiroplasma citri* (Andreason, 2016).

VECTOR COMPETENCE

Research focused on vector transmission aims to understand why only certain species of insects are able to transmit pathogens while others cannot. In this regard, one of the most important characteristics is vector competence. Vector competence can be defined as the efficiency of a vector to transmit a pathogen over time or per transmission opportunity (Purcell and Almeida, 2005). Factors or traits that are directly associated with the vector-pathogen interaction are comprised in this definition. Some examples are the susceptibility of the insect to be infected by the pathogen or the ability of the insect to transmit the pathogen to a suitable host or to its own progeny (Tabachnick, 1994).

These traits or factors can be divided into intrinsic and extrinsic factors and can affect the efficiency of the vectors to successfully acquire and transmit the pathogen (Nault and Ammar, 1989). Extrinsic factors affect the possibility of the vector to come into contact with the pathogen, meanwhile intrinsic factors influence the probability of the vector to become infected with a pathogen and to successfully transmit it (Hardy et al., 1983). Vector competence can also be affected by innate behavioral traits of the insect such as feeding behavior, probing activity and host preference (Ammar, 1994).

Extrinsic factors can include the distribution and density of the pathogen population in the host plant, the environmental conditions, or the insect vector's plant preference (Almeida et al., 2005). For example, temperature is one of the main environmental factors that can modify the rate of pathogen acquisition, as well as changing the length of the latent period and retention time (Ammar, 1994)

On the other hand, intrinsic factors are primarily related to internal physiological characteristics (Ammar, 1994) such as the rate at which the pathogens can move through the body to the salivary glands, the existence of barriers that can limit this movement (Nagata et al., 2002), or the presence of vector proteins capable of interacting with surface proteins of the pathogen (Galetto et al., 2011).

Despite the devastating effects that some insect-borne plant pathogens can cause in economically important crops, the mechanisms that regulate the transmission process and efficiency are not completely understood (Gray and Banerjee, 1999). However, advances in molecular biology and the ability to genetically manipulate pathogens and their vectors, have led to the discovery of genetic components in pathogens and their vectors that are responsible for successful transmission (Gray et al., 2007). For example, mutations in the functional chitinase gene, *chiA*, of *Xylella fastidiosa*, produces changes in the expression levels of other bacterial genes involved in vector and plant colonization resulting in the diminishing of the bacteria's ability to successfully colonize its insect host (Labroussaa et al., 2017). A study of the transmission of MFSV by *G. nigrifrons* has shown that, depending on the pathosystem, transmission efficiencies can be low and highly variable between leafhoppers, even within the same species, with some not being capable of successfully transmitting the virus. Only a quarter of the insects in this study were able to be infected and only 11-20% of them were capable of transmitting the virus. The differences in transcript expression between the infected and non-infected insects were considerable, especially in genes related with cytoskeleton organization and immunity. However, differences in transcript expression between transmitter and non-transmitter leafhoppers were more limited, showing that the transmitter phenotype was

the result of only a few genetic differences. The difference in competence to transmit the virus was apparently not associated with sequence polymorphism of the virus or differences in the transcript level abundance. (Cassone et al., 2014a).

INSECT GENOMICS AND NEXT GENERATION SEQUENCING

The insects are the largest group of animals in the planet with more than one million species described, and many more to discover. With the publishing of the *Drosophila melanogaster* genome (Adams et al., 2000), the understanding of insect genomics was initiated. It gave scientists a model system with more than 13,000 genes which became the basis of many comparative, evolutionary and population genomics studies. (Chen et al., 2016). Furthermore, with the development of next generation sequencing, an even larger number of studies have been conducted, including not only model species, but also non-model species such as the brown planthopper, *Nilaparvata lugens* (Matsukawa et al., 2018).

Since 2005, next generation sequencing (NGS) has changed the way scientists think about genetic information (Schuster, 2008). Unlike the former capillary electrophoresis-based sequencing, NGS techniques have the ability to process millions of sequence reads in parallel and produce shorter sequences (35-250 bp) which impacts how the data is going to be used and understood (Mardis, 2008).

Several technologies have been developed since the advent of NGS. During recent years, Illumina methodology has dominated most of the sequencing industry (Quail et al., 2012). Their method, called sequence-by-synthesis, uses reversible-terminator

nucleotides labeled with fluorescence, on clonally amplified DNA templates immobilized on a glass flowcell (Illumina, 2010).

Pacific Biosciences has developed a new technology that allows single molecule real time (SMRT) sequencing. This technology produces longer reads and is faster than other NGS methods, but has a lower throughput, higher error rate and is more costly. "PacBio sequencing" uses a double-stranded DNA (dsDNA) which have hairpin adaptors ligated on both ends creating a single-strand circular DNA. This molecule binds to a DNA polymerase attached to the bottom of a well. Each time a new nucleotide is added a different emission spectrum is generated and recorded in real time (Rhoads and Au, 2015).

Thanks to the advancement of NGS techniques, more information about insect genomics is becoming available. As more insect genes are discovered, the knowledge about their mechanisms of development and metamorphosis increases, providing researchers with a basis for novel approaches to pest control or the use of beneficial insects (Chen et al., 2016). Analyses of insect genomes have provided a lot of information regarding genome evolution as well as basic biological processes such as evolution, development, physiology, reproduction and survival (i5K Consortium, 2013).

Initiatives such as the i5K (i5K Consortium, 2013) or the Earth Biogenome project (Lewin et al., 2018) urge the scientific community to keep sequencing more insect genomes (as well as other organisms) as a means to better understand Earth's biodiversity. These, together with the availability of several comprehensive databases focused on different insect groups such as VectorBase (Lawson et al., 2008; Megy et al.,

2011), InsectBase (Yin et al., 2015), or Hymenoptera Genomes Database (Munoz-Torres et al., 2010), are driving the advancement of insect genomics by increasing the amount of data available to study.

Even though there are more than 430 insect species with complete genomes available (Childers, 2019), only one of those species belong to the family Cicadellidae and it is not a Deltocephaline leafhopper. This genome belongs to the glassy-winged sharpshooter (*Homalodisca vitripennis*). Nevertheless, more studies regarding the Deltocephalinae are being done in order to better understand the relationships among members of this group.

Many of these studies have focused on sequencing the complete mitochondrial genome of different species such as: *M. quadrilineatus* (Mao et al., 2017), *Drabescoides nuchalis* (Wu et al., 2016), *Maiestas dorsalis* and *Japanus hyalinus* (Du et al., 2017b), *Scaphoideus* spp. (Du et al., 2017a), *Abrus expansivus* (Wang and Xing, 2019) and many more. Other studies have focused in the bacterial communities and endosymbionts within these insects like *Dalbulus maidis* (Brentassi et al., 2017), or leafhoppers from the genus *Macrosteles* (Bennett and Moran, 2013; Kobiałka et al., 2016; Moran et al., 2005). While others have focused in the study of transcriptomes (Cassone et al., 2014a).

RNA INTERFERENCE (RNAi)

RNA interference (RNAi) is the process in which double-stranded RNA (dsRNA) is used to block gene expression (Zamore et al., 2000). It is a gene expression regulatory process that has a vital role in the regulation and maintenance of host defenses against viruses (Nandety et al., 2015). This mechanism was first observed in the nematode,

Caenorhabditis elegans (Fire et al., 1998). Since then it has been used in a wide variety of organisms, with *Drosophila melanogaster* becoming the first insect in which RNAi was reported (Kennerdell and Carthew, 1998).

The use of RNAi has greatly advanced the development of insect research because it enables scientists to silence a gene of interest offering the possibility of linking a gene function to a phenotype (Scott et al., 2013). It has pushed forward advances in reverse functional genomics because it allows the study of gene functions in nontransformable species (Bellés, 2009). Other applications of RNAi in insect science include its use for the management of pest populations and the reduction of disease pathogens (Scott et al., 2013).

RNAi is a post-transcriptional gene silencing (PTGS) process (Bass, 2000). Its mechanism initiates when dsRNA is cleaved into short interfering RNA (siRNA), micro RNA (miRNA) and piwi-interacting RNA (piRNA) by Dicer type RNAse II enzymes (Kanakala and Ghanim, 2016). After being cleaved, all of these small RNAs are loaded into the RNA-induced silencing complex (RISC). The siRNA (21-24 bp) is used by the complex to locate mRNA targets by base pairing. Once a target is located, it is subject to nucleolytic degradation mediated by the RNAse H enzyme, Argonaute. Thus, producing gene silencing because of translational inhibition (Hammond, 2005).

The responses to RNAi can be categorized as cell-autonomous or non-cell-autonomous (Kanakala and Ghanim, 2016). In the cell-autonomous response, the effect is limited to the cell where dsRNA is being expressed or was introduced. The non-cell-autonomous response can be divided into environmental and systemic. Environmental

RNAi consists of taking dsRNA from the cell environment, whereas systemic RNAi is the passing of the silencing signal from the tissue where it was first applied to other tissues (Gu and Knipple, 2013). A key component for the systemic RNAi is the transmembrane protein SID-1, which is required to import the RNAi sequencing signal. Homologs of this protein are present in other invertebrates and vertebrates, which shows the existence of an ancient origin and a conserved function (Shih and Hunter, 2011).

Because RNAi targets any mRNA that is complementary to the siRNA that forms part of the activated RISC complex, it has been largely used to study gene functions (Chen, 2013). Many of these studies have used insect systems, and a few have targeted members of the subfamily Deltocephalinae. Examples include: the use of parental RNAi in the green rice leafhopper, *Nephotettix cincticeps*, to cause gene knockdown in the next generation (Matsumoto and Hattori, 2016); and the use of RNAi in *Graminella nigrifrons* to silence two transcripts involved in the immune response against MFSV (Chen, 2013).

Laccase in insects

Laccases are oxido-reductase enzymes that are part of the large group of multicopper enzymes. They were first described in 1883, being one of the oldest described enzymes (Mayer and Staples, 2002). Laccase can be divided in two major groups, those that come from higher plants, and those that come from fungi (Harvey and Walker, 1999). The presence of laccase enzymes has been reported in bacteria as well as in insects (Mayer and Staples, 2002).

Insect laccases differ from those in plant and fungi in that they have a methionine in their T1 copper center, where plants and fungi have either phenylalanine or leucine

(Dittmer et al., 2004). Two main forms of this enzyme have been discovered through cloning: laccase-1 and lacasse-2 (Hattori et al., 2010). Lacasse-1 has been linked to the formation of part of the insect cuticle, and it is believed to be involved in cuticle sclerotization (Sugumaran et al., 1992). Lacasse-2 was found in the midgut and epidermis of *Manduca sexta*, and it is believed to oxidize toxic compounds ingested by the insect (Dittmer et al., 2004).

Studies of the green rice leafhopper, *N. cincticeps* have identified the presence of laccase excreted in its watery saliva, and suggest its role in the oxidization of toxic monolignols and non-toxic polymers during insect feeding (Hattori et al., 2005). More recent studies in the same insect have revealed the presence of a soluble laccase exclusively expressed in the salivary glands of the insect, further suggesting its role during feeding (Hattori et al., 2010), making it an interesting gene to be silenced in order to study its functions.

CHAPTER III

ASSEMBLY OF THE DRAFT GENOMES AND TRANSCRIPTOMES OF SIX DIFFERENT SPECIES OF DELTOCEPHALINE LEAFHOPPERS

ABSTRACT

With the advancements of next generation sequencing, the number of organisms that are being sequenced is increasing exponentially. The most common approach to sequencing the whole genome of an organism is to generate millions of reads and then assemble them as pieces in a puzzle. To assemble genomes, it is assumed that highly similar reads are originated in the same genomic region, however repetitive regions within a genome can further complicate this process. Nevertheless, whole genomes are a very useful resource to infer gene evolution or to develop novel pest management strategies. Insect genomes have become the most studied group of organisms regarding conservation or divergence of functional genomic elements. Of all the available insect genomes, less than 7% belong to the order Hemiptera, and only one of them, *Homalodisca vitripennis*, came from a species within the leafhopper family Cicadellidae.

For that reason, the objective of this study is to obtain high quality genome assemblies from Deltocephaline leafhoppers. To achieve that, DNA and RNA of six

species of Deltocephaline leafhoppers was extracted and sequenced using the Illumina platform. Genomic short reads were assembled *de novo* using MEGAHIT, and RNA reads were assembled using Trinity. Taxonomic partitioning of the genomic datasets was done using BlobTools and contaminant contigs were removed from the assembly. The estimated size of the genomes ranged between 0.82 Gbp to 1 Gbp, and completeness, assessed with BUSCO, was between 53% to 77%. The resulting genomes are highly fragmented due to low sequencing coverage, for a *de novo* genome project, and because DNA was obtained from pooled samples of non-inbred insects. Nevertheless, the now available draft genomes for the subfamily Deltocephalinae can serve as the basis of comparative analyses to better understand this insect subfamily.

INTRODUCTION

Until recently the sequencing of large genomes was carried out primarily by large research laboratories. However, thanks to the growth and spread of next generation sequencing technologies (NGS), whole-genome sequencing (WGS), as well as transcriptome sequencing, can be done by many laboratories and has become a key research component to study genome structure and expression (Nagarajan and Pop, 2013).

However, none of the sequencing technologies currently available are capable of reading the complete sequence of a whole-genome in only one glance (Simpson and Pop, 2015). Thus, sequencing is done by generating millions of small sequences (known as reads) that need to be assembled in the correct order. For example, Illumina sequencing generates millions of short reads with a size of 150 bp or 250 bp depending on how the library was prepared. The use of short reads increases the assembly difficulty, because

short reads are unable to span and cover completely the repetitive and polymorphic regions of the genome (Richards and Murali, 2015). The challenges to *de novo* assembly of large genomes are even greater if their inherent complexity and nonrandomness is taken into account (Sohn and Nam, 2016).

Most assembly methods rely on the underlying assumption that DNA with highly similar sequences come from the same location in the genome (Nagarajan and Pop, 2013). Consequently, DNA reads are joined together to create larger fragments known as contigs (contiguous sequences) and then the contigs are joined together to form scaffolds. However, this assumption is not always true; genomes have repetitive segments of DNA which generate those highly similar reads. To resolve and assemble those repetitive segments paired reads that span a repeat with enough unique sequences at any side of the repeat and paired reads with only one end on the repeat are needed (Miller et al., 2010). To assemble a genome is easier when all the repeats are shorter than the reads length and it becomes more difficult when the length of the repetitive content is longer than the length of the reads (Nagarajan and Pop, 2009).

Nevertheless, the use of whole-genome sequencing has allowed researchers to more comprehensively examine genomic information (Goodwin et al., 2016). Whole genomes are a resource that can be used as a foundation to understand a single species or for comparative studies. They can be used to infer gene evolution, population genetics analyses and can even be the basis to develop novel pest management strategies (Papanicolaou et al., 2016).

Because of the generally small genome size and the fact that they inhabit almost all ecological niches of earth, insects are the best group of organisms to study conservation or divergence of functional genomic elements (Waterhouse, 2015). *Drosophila melanogaster* was the first eukaryotic organisms to have its full genome sequenced and annotated, resulting in its becoming a model organism for animal genomics (Adams et al., 2000). The success of the *D. melanogaster* genome resulted in more insect species being chosen to be sequenced as part of research projects belonging to different fields namely agriculture, human health, economic or environmental sciences.

These first sequencing efforts focused on model insects from different orders for example the malaria mosquito, *Anopheles gambiae* (Holt et al., 2002), the silk moth, *Bombyx mori* (Xia et al., 2004), another fruit fly, *Drosophila pseudoobscura* (Richards et al., 2005), the honey bee, *Apis mellifera* (Weinstock et al., 2006), a mosquito, *Aedes aegypti* (Nene et al., 2007) and the red flour beetle, *Triboleum castaneum* (Richards et al., 2008). These efforts significantly increased the breadth of known insect genomes.

One of the main objectives of genome sequencing is obtaining high quality reference genomes needed to produce high quality gene model annotation. Because the size of the average gene locus ranges between 12 kb to 25 kb in *D. melanogaster*, and longer in larger insects, a N50 of at least 10 kb is required for high quality gene annotation (Richards and Murali, 2015). However, achieving this type of assembly in insects is challenging due to various reasons. Insects tend to have highly heterozygous genomes; their small size makes it difficult to obtain enough DNA from a single insect (Love et al., 2016); and their large diversity does not allow the use of available high quality genome assemblies from other species to aid in the process (Richards and Murali, 2015).

Despite these complications, there are genome assemblies for at least 430 species of arthropods representing 163 families of 38 different orders (Childers, 2019), which are archived in an International Nucleotide Sequence Database Collaboration (INSDC). There is a large disproportion in the number of species sequenced in different insect orders. While more than 30% of the sequenced arthropod genomes belong to the order Diptera, the Hemiptera, the fifth most sequenced group of insects, only represents about 7% of all the arthropod genomes available, with a total of 30 sequenced species (Figure 2). Of those 30 hemipteran genomes available only one belongs to a member of the family Cicadellidae.

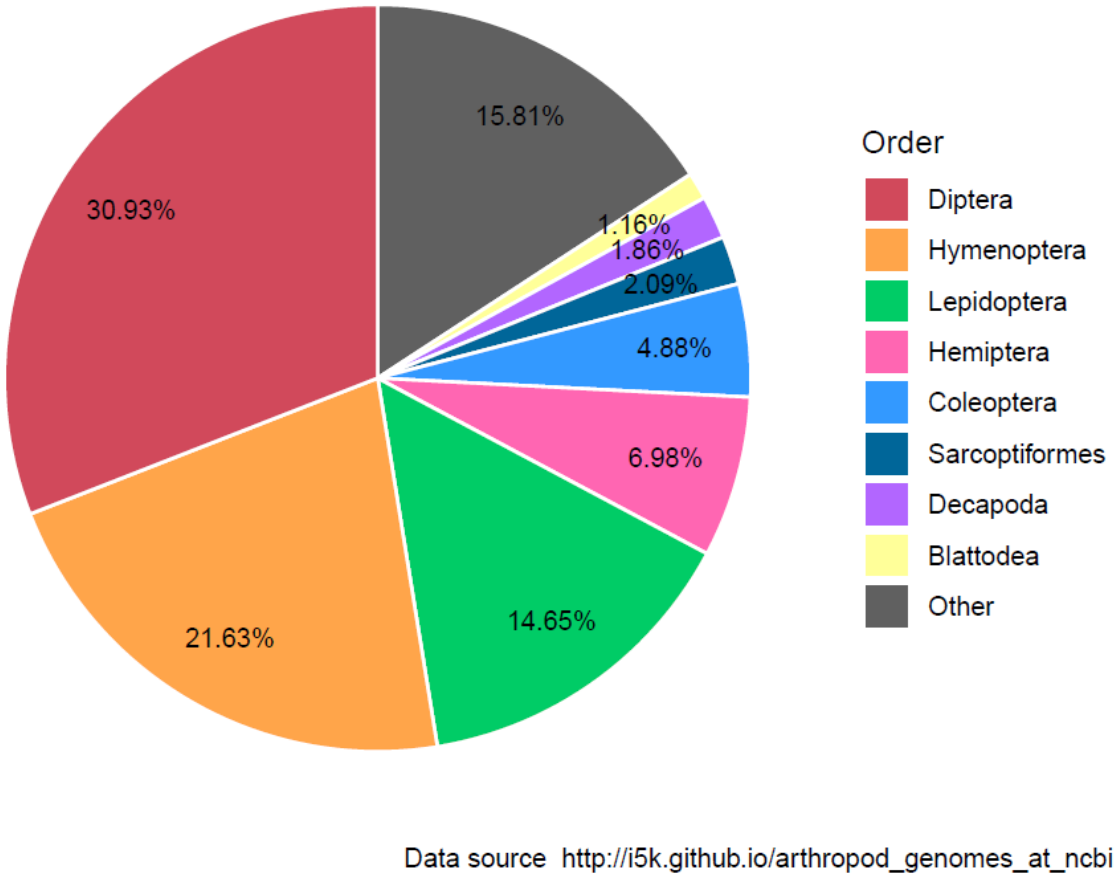

Figure 2. Percentage of number of arthropod species with sequenced genomes available at the NCBI database.

The leafhoppers (Hemiptera: Cicadellidae) are a group of plant feeding insects that can be found worldwide (Dietrich et al., 2001). One of the cicadellid subfamilies, the Deltocephalinae, contains the majority of leafhopper vectors of economically important plant pathogens (Chakravarthy, 2015). Even given the importance of this group of insects in agricultural, the classification and relationship among many of their genera is not yet well defined (Zahniser and Dietrich, 2013). For that reason, nucleic acids for six different species of Deltocephaline leafhoppers: *Balclutha neglecta, Balclutha rubrostriata, Dalbulus maidis, Exitianus exitiosus, Graminella nigrifrons* and *Macrosteles quadrilineatus* were sequenced to obtain draft genomes and transcriptomes. The overall goal is to obtain high quality genome assemblies that will serve as the basis for some comparative analyses that will allow a better understanding of the relationships among the members of this insect subfamily.

MATERIAL AND METHODS

Insect sources

Samples of *Balclutha rubrostriata, Dalbulus maidis, Graminella nigrifrons* and *Macrosteles quadrilineatus* were obtained from colonies maintained at Oklahoma State University. Colonies were maintained in insect-proof cages (10 x19 x 18 inches), at 24±1 °C with a photoperiod of 18-h light: 7-h dark. *Balclutha rubrostriata, Graminella nigrifrons* and *Macrosteles quadrilineatus* were fed with a combination of corn and oats, while colonies of *Dalbulus maidis* were maintained exclusively on corn.

The other two species, *Balclutha neglecta* and *Exitianus exitiosus* were collected from patches of mixed grasses in Stillwater, Oklahoma.

Nucleic acid extraction and sequencing

For all of the species used in this study, with the exception of *E. exitiosus*, genomic DNA was extracted from 5-6 adults using the E.Z.N.A.® Insect DNA Kit (Omega Bio-Tek Inc., Norcross, GA), with the following modification to the manufacturer instructions: the volume of all reagents was scaled down to one quarter of the amount detailed in the product manual (e.g. 87.5 μL of the lysis buffer was used instead of 350 μL). The extracted DNA was quantified using a Qubit 3.0 spectrophotometer using the Qubit dsDNA Broad Range Assay Kit (Life Technologies, Carlsbad, CA). The quality of the DNA was measured with a Nanodrop 3000 (Thermo Fisher Scientific, Waltham, MA) and by visualizing on an electrophoresis gel. High quality DNA for the above species was frozen at -80 °C, then shipped overnight to Novogene (Novogene Co., Ltd., Beijing, China). For *E. exitiosus*, the DNA extraction, quantification and quality control was done by Novogene. Around 1000 frozen *E. exitiosus*, primarily female, were shipped to Novogene for DNA and RNA extraction and quality control.

Similarly, total RNA of all of the species was obtained by extracting pooled samples of 5-6 adults using the E.Z.N.A.® Mollusk RNA Kit (Omega Bio-Tek Inc., Norcross, GA), with the following modification to the manufacturer instructions: the volume of all reagents was scaled down to one quarter of the amount detailed in the product manual. The quantification of the RNA was done using a Qubit 3.0 spectrophotometer using the Qubit RNA Broad Range Assay Kit (Life Technologies, Carlsbad, CA). Quality of the RNA was determined using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). High quality RNA for all species was frozen at -80 °C, then shipped overnight to Novogene.

DNA and RNA libraries construction and sequencing of the samples was completed by Novogene. The nucleic acids were used to construct paired-end 150 base pair (PE150 bp) TruSeq Illumina genomic libraries. Sequencing was done in an Illumina HiSeq 4000 platform. Raw sequencing data was received from Novogene through their FTP site.

Genome assembly

All bioinformatics analyses were done using the resources available at the High Performance Computing Center of Oklahoma State University. The quality of the raw sequencing data obtained from Novogene was checked using FastQC, version 0.11.3 (Andrews, 2010). Poor quality reads and adapters were trimmed using Trimmomatic, version 0.38 (Bolger et al., 2014).

The draft genome for each species was assembled using MEGAHIT, version 1.1.2, (Li et al., 2015) using the quality trimmed raw reads. Because there is not a reference genome available for this group of insects, the genome size estimate was calculated by counting unique k-mers (Marçais and Kingsford, 2011; Pucker, 2019). Jellyfish, version 2.2.6, (Marçais and Kingsford, 2011) was used to count the k-mers for each set of raw reads and R, version 3.6.1, was used to visualize and determine k-mer peak position and estimate the genome size.

Because the genomic DNA was extracted without performing any purification step on the insects, it can be assumed that some contaminants, such as bacterial endosymbionts, were likely sequenced. To clean up the draft genome, a taxonomic partition of the dataset was performed using BlobTools, version 1.0, (Laetsch and Blaxter, 2017) to assign each contig a taxonomic ID based on the comparison of their sequences against the nucleotide

database of the National Center for Biotechnology Information (NCBI). Lastly, the software P_RNA_scaffolder, version 1.0, (Zhu et al., 2018) was used to join contigs into scaffolds using information obtained from the transcriptome of each species.

Statistics of the assembled contigs and scaffolds were obtained using QUAST, version 4.6.3 (Gurevich et al., 2013). The completeness of the genome assembly was assessed using BUSCO, version 3.0.2, with the Arthropoda ODB9 dataset (Simão et al., 2015; Waterhouse et al., 2017).

Transcriptome assembly

Similar to the DNA data, the quality of the RNA raw reads was checked using FastQC, version 0.11.3 (Andrews, 2010). The transcriptome of each of the species was assembled using Trinity, version 2.5.1 (Grabherr et al., 2011; Haas et al., 2013). The assembly statistics of the transcriptomes was checked using QUAST, version 4.6.3 (Gurevich et al., 2013).

RESULTS

DNA extraction

Genomic DNA of each species was extracted from multiple samples. The best sample for each species (Table 1), based on their concentration and quality, was selected to be sent to Novogene for sequencing. Genomic DNA of *E. exitiosus* was extracted by Novogene.

Table 1. Concentration and quality of DNA samples used for sequencing

Sample name	Species	Concentration (ng/uL)	260/280	260/230
Bn1	*Balclutha neglecta*	23.5	1.91	2.1
Br2	*Balclutha rubrostriata*	24.7	2.07	2.1
D2	*Dalbulus maidis*	22.4	2.08	2.37
G1	*Graminella nigrifrons*	27.6	1.94	1.66
M1	*Macrosteles quadrilineatus*	21.2	1.99	1.70

Genome assembly

Between 27.48 and 60.95 Gbp of raw sequencing data was obtained from Novogene. After quality filtering and trimming the remaining amount of raw reads data ranged between 24.38 and 56.22 Gbp (Table 2).

Table 2. Amount of DNA sequencing data obtained for each species

Sample	Original raw reads (Gbp)	Quality trimmed raw reads (Gbp)
B. neglecta	32.50	27.78
B. rubrostriata	28.74	24.38
D. maidis	31.63	25.62
E. exitiosus	60.95	56.22
G. nigrifrons	32.51	28.24
M. quadrilineatus	27.48	24.29

Draft genomes were assembled using the quality-trimmed data using MEGAHIT and had the following metrics: the cumulative length of the draft genomes ranged between

0.82 to 1.5 Gbp. The total number of contigs in each draft assembly is between 41,6451 and 1,489,938; and the N50 of the assemblies ranges from 1363 to 3155 bp (Table 3).

Table 3. Statistics of the genome draft assemblies obtained with MEGAHIT

Species	**# contigs**	**Largest contig (bp)**	**Cumulative length (Gbp)**	**N50 (bp)**
B. neglecta	694903	176128	0.82	1363
B. rubrostriata	608598	182434	0.87	1861
D. maidis	416451	190373	0.86	3155
E. exitiosus	1489938	191606	1.5	1118
G. nigrifrons	917351	208779	1.3	1700
M. quadrilineatus	957700	176510	1.2	1383

Because the cumulative length does not truly represent the genome size, it was necessary to calculate the overall genome size using other methods. Using the k-mer counting method, the genome size estimates for each of the species ranges between 0.82 to 1 Gbp (Table 4)

Table 4. Genome size estimates using k-mer counting

Species	**Estimated genome size (Gbp)**
B. neglecta	0.85
B. rubrostriata	0.85
D. maidis	1.00
E. exitiosus	0.96
G. nigrifrons	0.98
M. quadrilineatus	0.82

During the taxonomic partition of the genomic datasets, the majority of contigs in each assembly, between 84 to 86%, did not produce any significant hit against the NCBI Nucleotide database. From the contigs that did produce a hit against the Nucleotide database, it was found that the three most represented taxa were Arthropoda, Bacteria and Chordata (Figure 3).

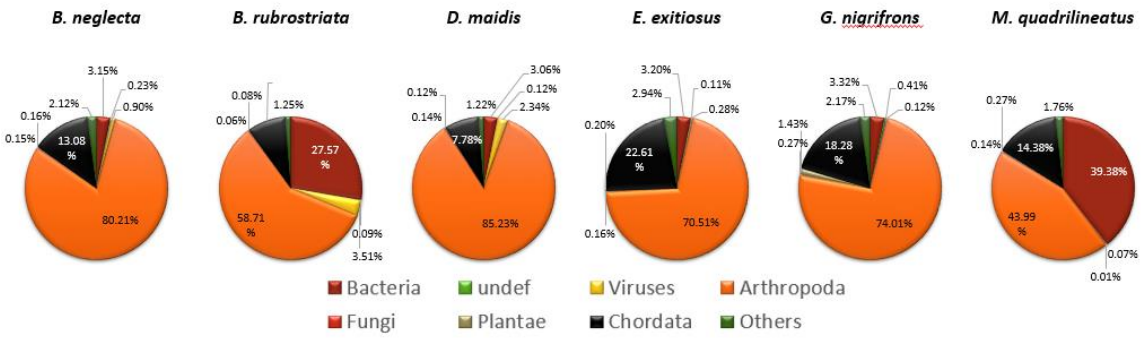

Figure 3. Percentage of contigs assigned to each taxon using BlobTools

All of the contigs not belonging to any taxon within the kingdom Animalia were removed from the assembly. The remaining contigs were used with P_RNA_scaffolder to improve the assembly showing and slight increase in the N50 for all of the six species draft genomes (Table 5).

The completeness of the draft genomes, based on the presence of conserved orthologs for the Arthropoda ranged between 53% to 77% (Figure 4). Only orthologs whose complete gene was found in the assemblies, either single copy or duplicated, are taken into account to assess genome completeness.

Table 5. Draft genomes statistics after P_RNA_scaffolder

Species	Largest contig (bp)	Cumulative length (Gbp)	N50 (bp)
B. neglecta	81569	0.82	1424
B. rubrostriata	115649	0.86	2043
D. maidis	75377	0.84	3461
E. exitiosus	86996	1.5	1119
G. nigrifrons	93985	1.3	1784
M. quadrilineatus	54000	1.1	1446

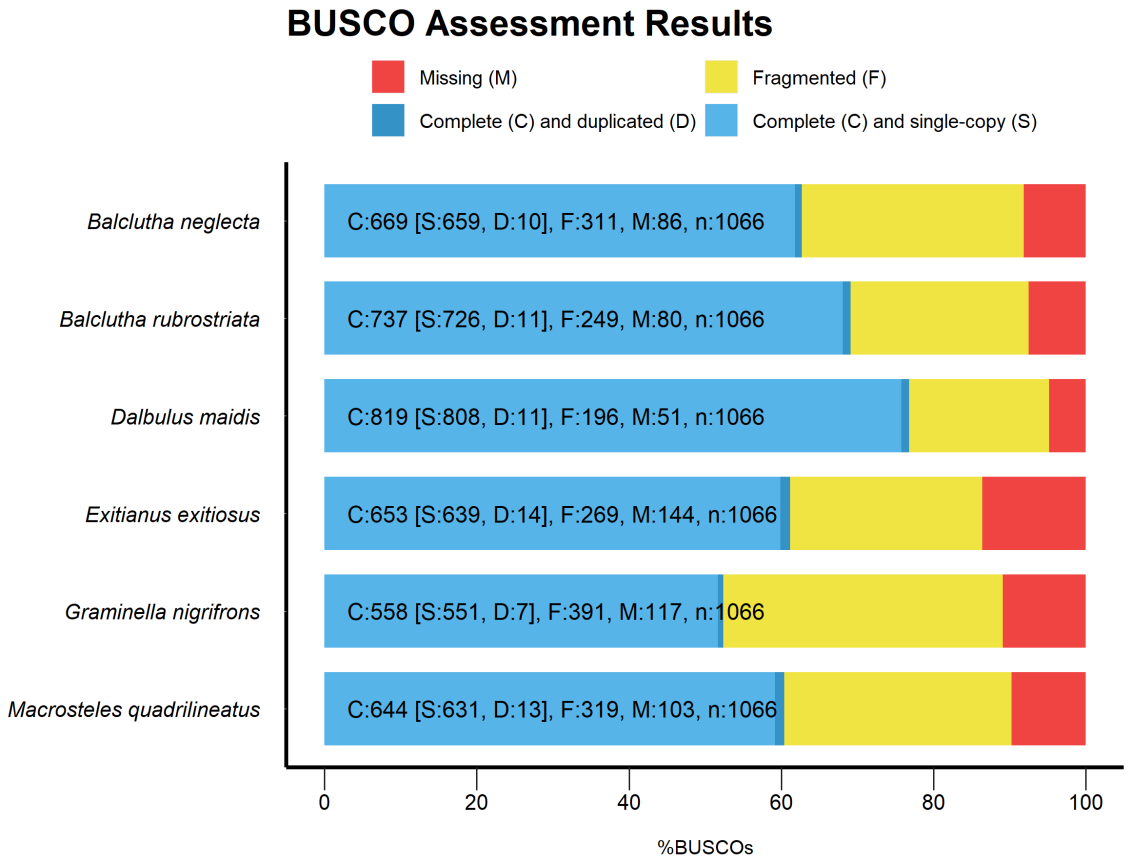

Figure 4. Completeness assessment of the leafhopper draft genomes using BUSCO

RNA extraction

Similar to the DNA, total RNA of each species was extracted from multiple samples, and the best one was chosen (Table 6) based on their concentration and quality.

Table 6. Concentration and quality of RNA samples used for sequencing

Sample name	Species	Concentration (ng/uL)	RIN
RBn1	*Balclutha neglecta*	120	8.1
RBr2	*Balclutha rubrostriata*	120	7.2
RD2	*Dalbulus maidis*	120	7.1
RG1	*Graminella nigrifrons*	120	8.4
RM1	*Macrosteles quadrilineatus*	120	6.7

Transcriptome assembly

Between 27.48 and 60.95 Gbp of raw sequencing data was obtained from Novogene (Table 7). Quality trimming was not performed in the raw data because the reads for all species were of high enough quality and were free of sequencing adapters.

Table 7. Amount of RNA sequencing data obtained for each species

Sample	Original raw reads (Gbp)
Balclutha neglecta	6.19
Balclutha rubrostriata	7.37
Dalbulus maidis	6.90
Exitianus exitiosus	14.99
Graminella nigrifrons	6.74
Macrosteles quadrilineatus	6.78

Raw reads were used to assemble a transcriptome using Trinity. The assembled transcriptome metrics can be seen in Table 8.

Table 8. Statistics of the transcriptome assemblies obtained with Trinity

Species	**# contigs**	**Largest contig (bp)**	**N50 (bp)**
B. neglecta	209600	19702	1365
B. rubrostriata	121626	16958	1564
D. maidis	102849	17949	2020
E. exitiosus	115350	31526	1599
G. nigrifrons	117882	18051	1696
M. quadrilineatus	117442	12731	1452

DISCUSSION

A total of six draft genomes and transcriptomes, one for each species, were obtained. To our knowledge, these are the first draft genomes available for any species within the leafhopper subfamily Deltocephalinae. Nonetheless, more work is required to improve these draft genomes, as they appear to be highly fragmented and discontinuous.

N50 is a value that is typically used to determine the success of an assembly. This value represents the length of the contig or scaffold that overlaps the mid-point of the length-arranged concatenation of contigs (or scaffolds) (Mäkinen et al., 2012). For high quality insect assemblies this value is expected to be of at least 10 kb (Richards and Murali, 2015).

The N50 values obtained in this study, together with the high number of contigs in each assembly, demonstrate that the obtained assemblies are not highly contiguous (Alhakami et al., 2017). Many reasons can explain a discontinuous assembly, from poor quality DNA samples to excessively low coverage.

High quality DNA is very important in order to obtain high quality sequencing data (Dominguez del Angel et al., 2018) so selecting the appropriate extraction method and quantity of samples to use is important. In order to prepare the libraries for sequencing, high molecular weight DNA is usually preferred or required (Ekblom and Wolf, 2014). Having an absorbance A260/A280 ratio between 1.8 to 2 is generally considered "pure" for DNA. The DNA extracted from the leafhoppers used in this study, with the exception of *B. rubrostriata* and *D. maidis*, are within this ratio and thus can be considered pure (Table 1). In these two species, their DNA A260/A280 ratio is a little bit higher, 2.07 and 2.08 respectively, however high 260/280 ratios are not usually considered problematic (Chovatia and Sharma, 2014). In the case of A260/A230 ratio, values between 1.8 to 2.2 are considered "pure" for nucleic acids. The A260/A230 ratio of the two *Balclutha* species is within the accepted range, the A260/A230 ratio of *D. maidis* is higher than 2.2, and the A260/A230 ratio of *G. nigrifrons* and *M. quadrilineatus* is lower than 1.8. While high A260/A230 ratios does not constitute an issue it may be indicative of a dirty pedestal at the moment of measuring, low A260/A230 ratios may indicate the presence of contaminants within those two samples (Chovatia and Sharma, 2014).

Pooling individuals may increase the heterozygosity present in the sample, leading to a more fragmented assembly (Dominguez del Angel et al., 2018). However, in subjects such as the leafhoppers, for whom a single individual may not yield enough DNA for

sequencing, pooled samples can be used but it does require having highly inbred individuals. For some of the species that were used in this study, *B. rubrostriata*, *D. maidis*, *G. nigrifrons* and *M. quadrilineatus*, DNA was extracted from pooled samples obtained from relatively new established colonies, whose members cannot be considered highly inbred. Furthermore, for *B. neglecta* and *E. exitiosus*, insects were collected from the field further increasing the diversity between them.

Most of the analysis of sequencing data assumes that the data being used is trustworthy. However, this is not always the case with the raw reads that are obtained from automated sequencing machines (Chou and Holmes, 2001). The presence of poor quality sequences among the raw reads, or sequencing adapters that were not removed can interfere with the downstream analysis of the data (Bolger et al., 2014). Consequently, it is usually necessary to pre-process the raw data and perform a quality-trimming of the raw reads (Dominguez del Angel et al., 2018). To assure the trustworthiness of the data, quality trimming of the raw reads was completed resulting in the loss of 14%, on average, of the overall amount of data available for each species (Table 2).

Another explanation of the highly fragmented nature of these assemblies is that the sequencing coverage was not as high as expected. Because the genome size of most of the hemipterans ranges between 407 Mbp to 1.230 Gbp (Hanrahan and Johnston, 2011) a small genome size and thus higher coverage was expected. However, based on the estimated genome size, between 0.85 to 1 Gbp, and the amount of raw sequencing data obtained from Novogene, between 27.48 and 32.51 Gbp, the estimated coverage of this sequencing effort was between 27X to 32X. A good assembly requires enough overlap between the sequence reads at every position of the genome. The more times different sequences overlap over

every nucleotide the higher the coverage (Ekblom and Wolf, 2014). For *de novo* assemblies it is usually recommended to have a coverage of between 60X to 80X (Desai et al., 2013).

One interesting and unexpected finding was that the estimated %GC content of the draft genomes was between 33% to 35% which can be considered relatively low compared to other insects. For example, the honeybee, *Apis mellifera*, a species traditionally considered to be AT-rich has a %GC content of 32.7%, the pea aphid, *Acyrthosiphon pisum*, also has a low %GC content, 31.5%. On the other hand, Diptera species, such as *Anopheles gambiae*, *Culex pipiens* and *Drosophila melanogaster*, tend to have higher %GC content with 44.3%, 63.2% and 42.3% respectively (Manoj, 2007). It is known that low GC contents as well as high GC contents can interfere with the sequencing causing low coverage of these regions. GC bias can affect several steps of the sequencing process resulting in the uneven coverage of the genome. One of the main factors in producing GC bias is during the library amplification by PCR, where these GC-rich and GC-poor regions will not amplify at the same rate as the rest of the genome (Chen et al., 2013).

As a way to improve the genome assembly, as well as to get rid of any "contaminants" that can be present in the sample, a taxonomic partition of the dataset was required. When DNA is extracted from a sample it may contain genomic information from organisms other than the target, such as parasites or bacteria. If those extraneous sequences are not identified and removed, it can lead to false assumptions regarding the target biology and metabolic processes (Laetsch and Blaxter, 2017). Because Deltocephaline leafhoppers harbor two obligate endosymbiotic bacteria, *Sulcia muelleri* and *Nasuia deltocephalinicola* (Kobiałka et al., 2018b; Moran et al., 2005; Nault and Rodriguez, 1985), the presence of sequencing data belonging to these bacteria was expected to be among the assemblies.

Unfortunately, it was not possible to obtain the complete genomes of the endosymbionts from the data available, as only a few contigs from these bacterial species were identified in the assemblies. This happens because the predominance of host DNA tends to diminish the coverage of the microbial genomes (Pereira-Marques et al., 2019).

Every contig that belongs to the kingdom Animalia as well as every contig that did not have a hit in the NCBI database were conserved into the assembly. Because fragmented and low-quality assemblies are prone to have low-quality annotation and under-represent the gene content of the genome (Denton et al., 2014), the removal of too much information from the assembly can result in the potential loss of important information.

Despite all efforts, the completeness of the genomes assessed with BUSCO for all of the species was below the minimum standard for a "high quality" draft genome which requires that the completeness of the genome must be at least 90% (Chain et al., 2009). As these standards are most likely to be maintained or increase due to the availability of new technologies for sequencing and assembling genomes, the draft genomes presented here require further improvement. For example, new massive sequencing efforts such as the i5K or the Earth Biogenome project are likely to require genomes to have over 80% complete BUSCO genes, less than 10% partial BUSCO genes and less than 10% missing BUSCO genes (Lewin et al., 2018).

The lack of a reference genome and the fact that leafhoppers are not a model organism for genomic research, makes them a challenging group to work with (da Fonseca et al., 2016). Nevertheless, the study of non-model organisms is important because they allow the exploration of underlying biological diversity that may be lost when working

with model species (Armengaud et al., 2014). For that reason, the availability of draft genomes for the subfamily Deltocephalinae, even though they are not as contiguous and complete as desired, is a valuable resource to better understand the relationship between the members of this group as well as a greater insight into how vector competence may be genetically regulated.

CHAPTER IV

ANNOTATION AND COMPARISON OF THE GENOMES AND TRANSCRIPTOMES OF SIX DIFFERENT DELTOCEPHALINE LEAFHOPPERS TO DISCOVER VECTOR COMPETENCE-RELATED GENES

ABSTRACT

To acquire more knowledge of the biology of an organism using sequencing data is not sufficient to simply make the assemblies available, genomes and transcriptomes also need to be properly annotated. Annotation is the process of identifying features of biological interests, such as genes or regulatory sequences, that are encoded in the genome. The annotation of genomes and transcriptomes can reveal more information about the mechanisms underlying biological processes of interest, such as insect vector transmission and vector competence. In this study, the genomes and transcriptomes of six species of Deltocephaline leafhoppers were annotated to gain insights about the genes that may be involved in vector competence. For the genome annotation, the repetitive content of the genomes was identified and masked using RepeatMasker with custom repeat libraries. Gene models were predicted with the help of the RNA transcripts using PASA.

The transcriptomes were annotated using the Trinotate pipeline that includes homology searches of the transcripts against several biological databases. The amount of repetitive content in the genomes ranged between 12.87% to 32.30%. PASA predicted between 48,233 and 116,558 possible gene structures in the genome. Between 29.95% and 47.78% of the transcripts of each species were annotated to at least one of the biological databases. In order to clarify the differences between vector and non-vector leafhoppers, a clustering analysis of the peptide sequences was performed. This analysis revealed 925 peptide sequences that were shared among the vectors that were less than 85% similar to the non-vectors. Some of these peptide sequences were identified by homology as belonging to actin, ATP synthase, myosin, vinculin and other proteins that are believed to be involved in the process of insect vector transmission.

INTRODUCTION

Next generation sequencing has become a widely used tool in genomics, transcriptomics, epigenetics and transcription factor research among many others. This has resulted in the increase of the number of organisms whose genome has been sequenced (Ellegren, 2014). To be able to efficiently use the genomic and transcriptomic sequences of an organism to answer biological questions, it is necessary to annotate them in a rigorous and efficient manner (Misra et al., 2002).

The objective of annotation is to identify features of biological interest or importance, such as genes, by deciphering the information coded in the genome or transcriptome (Haas et al., 2011). Specifically, genome annotation is the process of pairing specific features to the genomic DNA sequence. These features can be protein-coding

genes model, repetitive sequences or non-protein coding RNA sequences (Misra et al., 2002). To fully utilize the potential information in a genome sequence, biologically relevant information needs to be annotated. This includes the generation of gene models as well as functional information such as gene ontology terms (GO) (The Gene Ontology Consortium, 2004), or the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa and Goto, 2000).

The annotation of protein-coding genes in an accurate manner is one of the earliest and most important steps when analyzing assembled genomes. This can be done using several methods and pipelines that have been developed for genome annotation. Some methods use only the genome itself as evidence to predict genes, called *ab initio* methods, while others can use additional evidence originated from RNA-seq, expressed sequence tags (ETSs), protein databases or from genomes of close relatives (Hoff and Stanke, 2015).

Among the several insect species that have a genome available, *Drosophila melanogaster* and *Anopheles gambiae* are two of the few whose genomes have been extensively sequenced, and whose genes have been manually annotated and are updated regularly. For most insect species, genome annotation is made computationally, which can lead to low-quality models that may hinder future functional studies (Cao and Jiang, 2019).

The annotation of insect genomes can be considered at an intermediate difficulty level because even though insect introns are long enough to permit complex alternative splicing, the size of insect genomes are, usually, smaller than more complex vertebrate genomes (Hoff and Stanke, 2015). Nevertheless, to completely annotate the genome is a

formidable task that requires a lot of effort and proficiency in managing and analyzing bioinformatic data (Ekblom and Wolf, 2014).

The quality of the genome assembly has a very important role in determining the success of the annotation. In order to obtain satisfying results, a genome assembly that is highly contiguous and fairly complete (more than 90%) is needed (Ekblom and Wolf, 2014). Deciding when an assembly is ready to be annotated is an important step of a successful genome annotation. Assembly errors can occur due to many reasons and much of the time determining what is real and what is an artifact is complicated. Nonetheless, there are some statistical tests and tools that can be used to ensure the quality of the genome assembly (Del Angel et al., 2018).

A widely used metric to evaluate genome assemblies is the N50. N50 is the length of the smallest contig that makes up 50% of the total size of all the contigs. It is important to note that N50 does not represent the correctness of the assembly, it is a measure of contiguity, the higher its value the lower level of fragmentation of the assembly (Mäkinen et al., 2012). It can be assumed that an assembly with an N50 larger than the median gene length will produce an acceptable annotation. Since the genome size is roughly proportional to the median gene length, genome size can be used to estimate the minimum N50 that is required for annotation (Yandell and Ence, 2012).

Other tools to evaluate the genome assembly may include: the use of REAPR (Hunt et al., 2013) to identify misassemblies based on the mapping of the paired end reads or BlobTools (Laetsch and Blaxter, 2017) to remove contaminants in the reads. Finally, the completeness of the assembly can be determined using BUSCO, a software that searches

the assembly for the presence of protein coding genes that should always be present in the genome based on its taxonomic lineage (Simão et al., 2015; Waterhouse et al., 2017).

Once the assembly has been determined adequate to move forward, the first step of genome annotation is identifying and masking repetitive content. Repetitive content can be classified into two different types: "low-complexity" sequences, that are homopolymeric runs of the same nucleotide, or transposable elements (TE) (Del Angel et al., 2018). These TEs have been described in virtually all eukaryotic species. They generate plasticity in the genome due to their ability to move and replicate. The TEs are very diverse group with thousands of families and can occur in large numbers. They represent a large proportion of the genome: more than 80% of the wheat genome, 45% of the human genome and around 20% of the *Drosophila melanogaster* genome. Most TEs are autonomous, meaning that they encode for proteins that can mediate their own transposition (Flutre et al., 2011).

A few classification systems have been proposed to classify TEs. One of these hierarchical classifications systems has been developed and implemented into RepBase (Jurka et al., 2005), which is database of eukaryotic repetitive and transposable elements that has been available since 1990. According to this classification system, TEs can be divided into two types of transposable elements based on their transposition intermediate. Type I TEs, or retrotransposons, have an RNA intermediate and its mechanism is commonly called "copy-and-paste". Type II TEs, or DNA transposons, have a DNA intermediate and their mechanism is called "cut-and-paste" (Finnegan, 1989).

These two types comprise five major TEs classes that are: long-terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, cut-and-paste DNA transposons,

rolling-circle DNA transposons (*Helitrons*), and self-synthetizing DNA transposons (*Polintons*). These five classes are based on structural similarity, sequence relationships and enzymology (Kapitonov and Jurka, 2008)

Retrotransposons do not cleave or transfer DNA at the donor site. These TEs transcribe the RNA intermediate using the genomic copy as a template, then a TE-encoded retrotranscriptase will reverse transcribe the intermediate into DNA. Retrotransposons are usually the biggest contributors to the repetitive content of a genome. Long-terminal repeats are more predominant in plants than in animals and seem to be closely related to retroviruses (Wicker et al., 2007). These retrotransposons include the *Gypsi, Copia, BEL* and *DIRS* superfamilies and some families of endogamous retroviruses (Kapitonov and Jurka, 2008). Non-LTR retrotransposons include the short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs), among other superfamilies. SINEs come from a different origin than other LTR and non-LTR retrotransposons. They are non-autonomous, relying on LINEs retro-transcriptase for transposition (Kajikawa and Okada, 2002). LINEs can be autonomous, encoding their own retro-transcriptase and nuclease for transposition. They can reach lengths of several kilobases (Wicker et al., 2007)

Type II TEs, or DNA transposons, move by a DNA intermediate and are found in all eukaryotes and occur in moderate numbers. (Wicker et al., 2007). They include the remaining three classes of transposons: cut and paste, rolling circle and self-synthetizing DNA transposons (Kajikawa and Okada, 2002).

Given the high percentage of repetitive content that can be found in some of the more complex genomes, the annotation of the repetitive elements is considered one of the

major tasks during genome projects and needs to be done before even trying to start with gene prediction (Del Angel et al., 2018). To achieve that, several software packages have been developed to identify and annotate all TE copies in the genome (Flutre et al., 2011). Some of the more frequently used tools include RepeatMasker (Smit et al., 2015b) and the REPET package (Flutre et al., 2011; Quesneville et al., 2005).

After finishing the identification and masking of the repetitive sequences, proper annotation of the genome can take place. First, *ab initio* algorithms can be used to create a baseline prediction of coding genes in the genome. After this initial prediction, other evidence such as protein alignments from closely related species may provide complementary information to the predicted gene models. The best evidence can be obtained from EST or RNA-seq data that will inform the model of the presence of untranslated regions (UTRs), transcription initiation sites and splice sites in addition to information about the coding sequence (CDS) (Ekblom and Wolf, 2014).

Then, all the information collected from the *ab initio* prediction and from the protein and RNA evidence is condensed to produce the final set of annotations. This is a difficult task for which several automated tools, such as MAKER (Cantarel et al., 2008) or PASA (Haas et al., 2003) have been created. Nevertheless, manual inspection and curation of the annotation is important in order to avoid systematic error and provide the best possible annotation (Del Angel et al., 2018).

During the last decades, genome and transcriptome annotation have evolved increasing the amount of information that it is possible to recover from sequencing data, leading to more precise insights of the biological processes that occurs within and between

different organisms (Abril and Castellano, 2019). One biological process for which specific mechanisms still need to be studied is vector competence. Vector competence is the ability of a vector to acquire, maintain and transmit a pathogen (Vogels et al., 2017). It is regulated by environmental, genetic and epigenetic factors (Houé et al., 2019). Moreover, studies in *Aedes aegypti* have found that the endosymbionts can also affect the ability of a vector to successfully transmit a pathogen (Zhang et al., 2013). Therefore, the goal of this portion of the project is to annotate the genomes and transcriptomes of the six Deltocephaline leafhoppers, in order to compare the predicted gene models of vector and non-vector leafhoppers and resulting in the identification of candidate genes that may be related to vector competence.

MATERIAL AND METHODS

All bioinformatics analyses were done using the resources available at the High Performance Computing Center at Oklahoma State University which include the supercomputer cluster Cowboy, comprised of more than 250 individual compute nodes.

Annotation of the genome

To annotate the genome, the first step was to mask the repetitive content that was in the assembly. In order to perform a thorough annotation of the repetitive content of the draft genomes, several software packages were used. The software RepeatModeler, version 1.0.11, (Smit and Hubley, 2015) aided in the *de novo* identification of repeats in each of the genomes to create their own repeat libraries. RepeatMasker, version 4.0.8 (Smit et al., 2015a) was used with the custom libraries as well as with the RepBase, version 24, (Bao et al., 2015) repeat libraries to mask the repetitive content of the draft genomes.

The PASA pipeline (Haas et al., 2003) then produced the initial gene models by aligning the RNA transcripts to the draft genomes. Because of the highly discontinuous nature of the assemblies, other gene structure prediction and annotation software were not able to produce satisfactory results.

Annotation of the transcriptome

To annotate the transcriptomes of the Deltocephaline leafhoppers, the Trinotate pipeline (Bryant et al., 2017) was used. This pipeline is a comprehensive annotation tool that uses a number of well referenced methods to generate functional annotation of *de novo* assembled transcriptomes.

The six leafhoppers transcriptomes, as well as their peptide sequences obtained with TransDecoder, version 5.2.0, (Chapter III) were subjected to a homology search against the UniProtKB/Swiss-Prot dataset (Boutet et al., 2007; The UniProt Consortium, 2018) using Blastp and Blastx searches (Camacho et al., 2009). Protein domains in the peptide sequences were identified by HMMER, version 3.1b2, (Eddy, 2011) with a search against the Pfam database (Finn et al., 2016). Ribosomal RNA (rRNA) genes were identified in the transcriptomes using RNAmmer, version 1.2 (Lagesen et al., 2007).

The presence of signal peptides and their cleavage sites in the peptide dataset of each leafhopper species was identified using SignalP, version 4.1, (Nielsen, 2017; Nielsen et al., 1997). The prediction of the presence of transmembrane helices in the peptide sequences was completed with TMHMM, version 2.0 (Krogh et al., 2001).

Finally, all results were loaded into a SQLite database, and a transcriptome annotation report was generated for each leafhopper species. The transcriptome report

includes the annotation in functional genomics databases such as GO (Ashburner et al., 2000; The Gene Ontology Consortium, 2018), KEGG (Kanehisa and Goto, 2000) and eggNOG (Huerta-Cepas et al., 2015).

In order to search for differences among vector and non-vector leafhoppers, the peptide sequences of each species were clustered together at 85% similarity using CD-HIT (Fu et al., 2012; Li and Godzik, 2006), to determine which proteins were shared among each group or specific to one group (Figure 5).

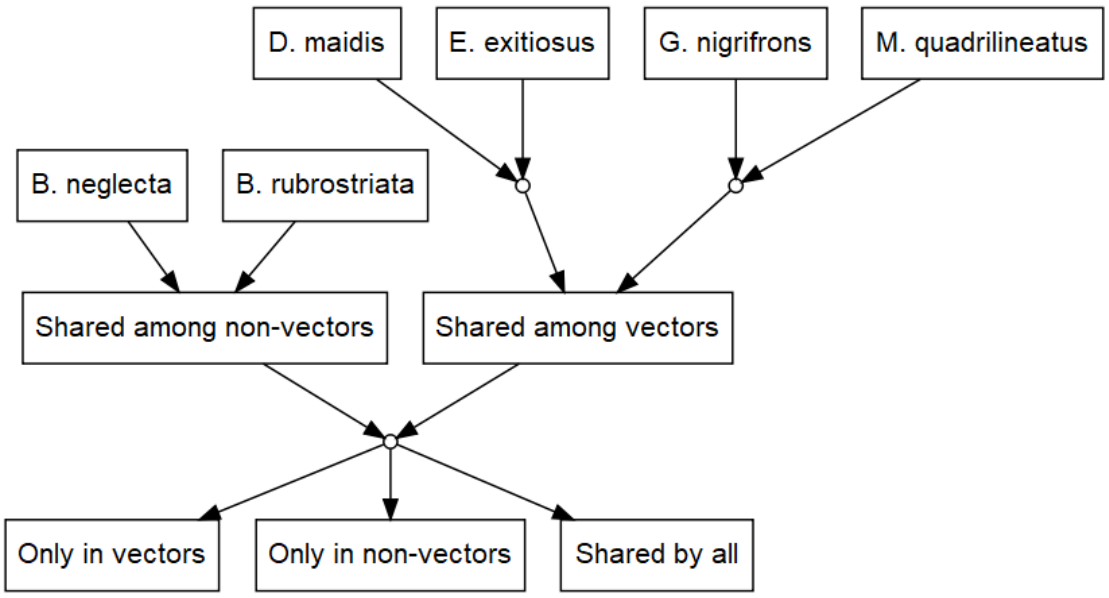

Figure 5. Method outline for the clustering of peptide sequences between vector and non-vector leafhoppers

RESULTS

Repetitive content annotation

The repetitive content in each of the draft genomes was annotated using the Arthropoda repeat libraries from RepBase. These repeat libraries were not able to fully identify the repetitive content of the leafhopper genomes, so custom repeat libraries were constructed using RepeatModeler.

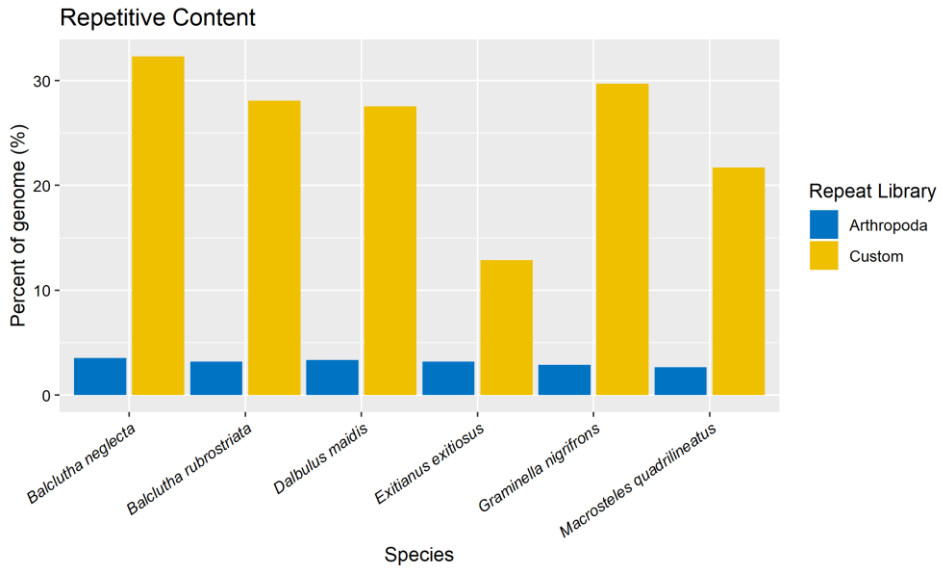

Figure 6. Comparison of the amount of repetitive content discovered using the Arthropoda repeat library and a custom repeat library for each genome

Figure 6 shows the different quantities of repetitive content that are annotated using each of the libraries. *Balclutha neglecta* was the species found to have the highest repetitive content, 32.30%, while *E. exitiosus* was the species with the least amount of repetitive content identified 12.87%.

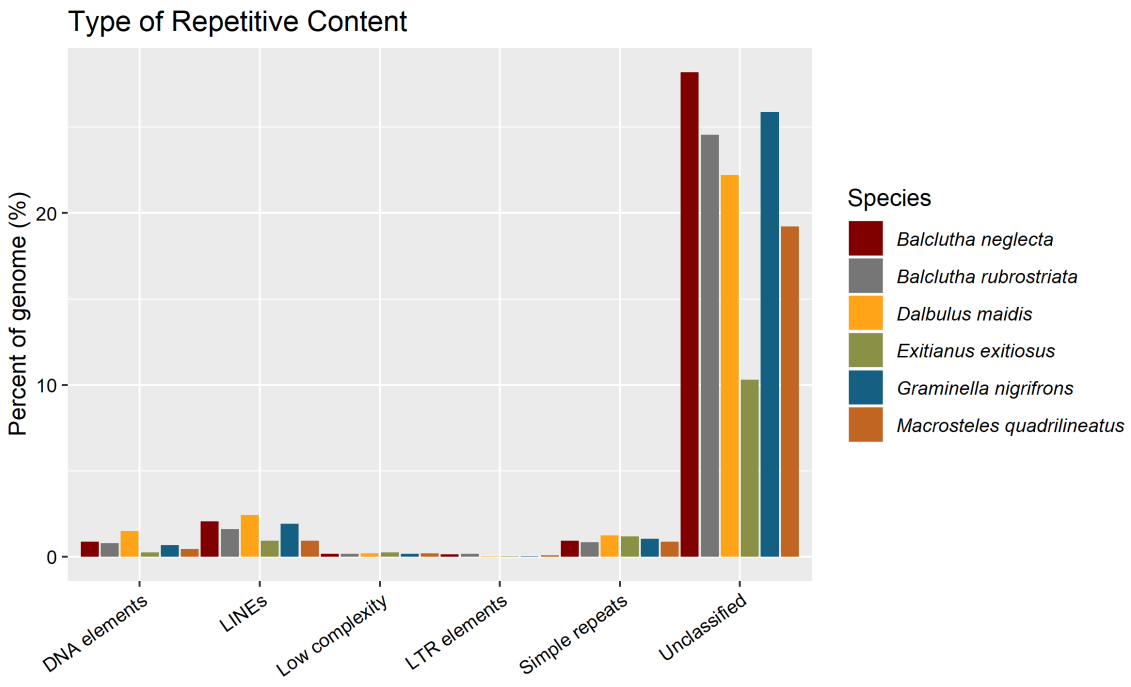

Figure 7. Types of repetitive content in each draft genome

Of all the TEs that were successfully identified, the long interspersed nuclear elements (LINEs) were more abundant in all of the genomes, between 1.6 to 0.9% of the genome (Figure 7). However, because most of the repetitive content is unclassified it was not possible to determine which was the most abundant TE.

The genomes were masked using the information obtained from the identification and annotation of the repetitive content. The masked draft genomes were used together with the transcriptomes to identify gene structures using the PASA pipeline. In Table 9, the possible number of gene structures that were predicted by PASA is presented. The highest number of possible gene structures were predicted for *B. rubrostriata*, while *D. maidis* has the lowest number of predicted gene structures.

Table 9. Number of possible gene structures predicted by PASA

Species	Number of possible gene structures
B. neglecta	47502
B. rubrostriata	116558
D. maidis	48233
E. exitiosus	81875
G. nigrifrons	59785
M. quadrilineatus	96599

Transcriptome annotation

The transcriptomes for the six species of Deltocephaline leafhoppers were annotated using the Trinotate pipeline. This pipeline uses homology searches against a wide array of databases to annotate the genomic features present in each transcriptome. For that reason, the peptide sequences from each transcriptome were obtained with TransDecoder (Figure 8).

The transcriptomes were searched against four different databases (Table 10). Among all of them the higher number of annotations was in the UniprotKB/Swiss-prot database for all the species. In total, 47.78% of *B. neglecta*, 33.45% of *B. rubrostriata*, 45.73% of *D. maidis*, 29.95% of *E. exitiosus*, 42.62% of *G. nigrifrons* and 30.97% of *M. quadrilineatus* transcripts produced a match to at least one of the databases.

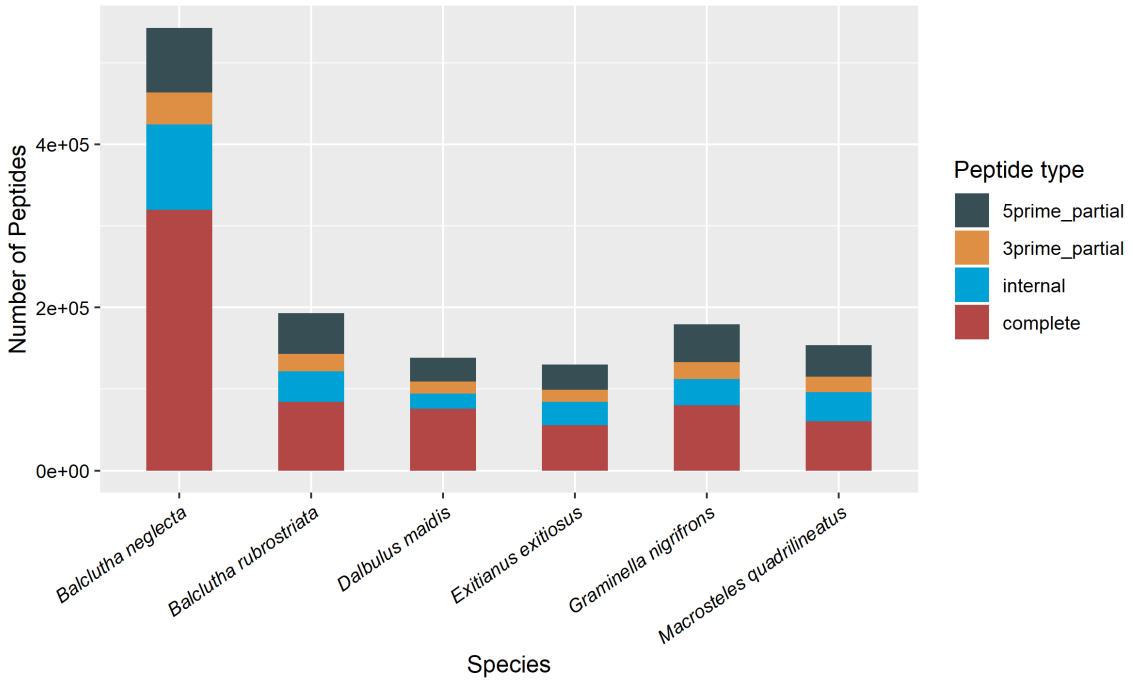

Figure 8. Number and type of peptides obtained from each transcriptome. Type meaning: complete - both the start and stop codon; internal - neither the start nor stop codon; 5prime_partial - only start codon; 3primer_part only stop codon.

Table 10. Annotation results of the transcriptomes in different databases

	UniprotKB/ Swiss-prot	**KEGG**	**GO**	**eggNOG**
B. neglecta	42.53 %	27.81%	31.81%	26.73%
B. rubrostriata	31.45%	17.16%	20.00%	16.51%
D. maidis	42.97%	25.41%	28.65%	24.68%
E. exitiosus	27.32%	16.10%	18.91%	15.38%
G. nigrifrons	39.47%	21.34%	24.61%	20.63%
M. quadrilineatus	28.55%	18.07%	21.20%	17.31%

A small percentage of transcripts were found to belong to ribosomal RNA (rRNA) genes: 0.01% in *B. neglecta*, 0.01% in *B. rubrostriata*, 0.03% in *D. maidis*, 0.05% in *E.*

exitiosus, 0.03% in *G. nigrifrons*, 0.02% in *M. quadrilineatus*. In regard to the peptide sequences, between 5.38% and 8.13% of the plasmids appear to have transmembrane helical domains, and between 2.11% and 3.21% of the transcripts seems to have signal peptides.

A GO analysis of the transcripts was done with the objective to describe their biological functions within the GO classification system. Between 31079 and 134755 transcripts of each species were assigned a GO term. Figure 9 to Figure 14 show the 10 most represented GO terms per each of the three main GO domains: biological process, cellular component and molecular function.

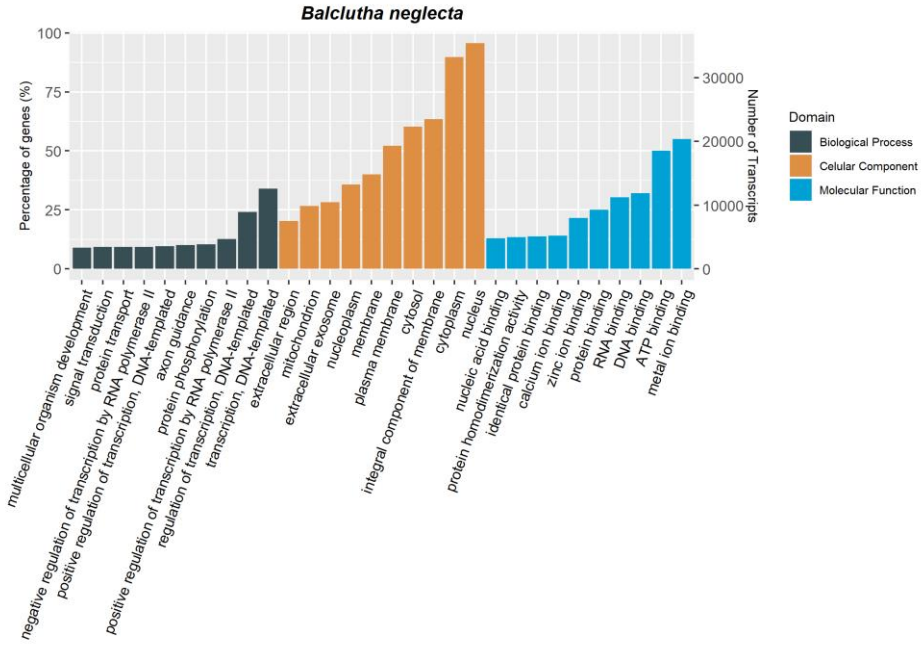

Figure 9. Top ten most represented GO terms of each domain in *B. neglecta*

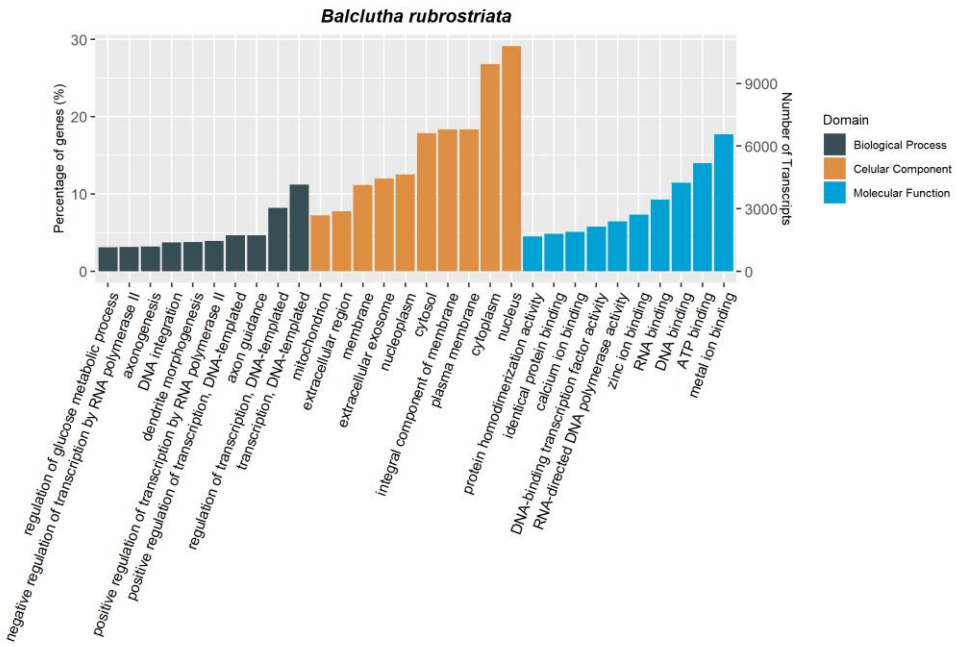

Figure 10. Top ten most represented GO terms of each domain in *B. rubrostriata*

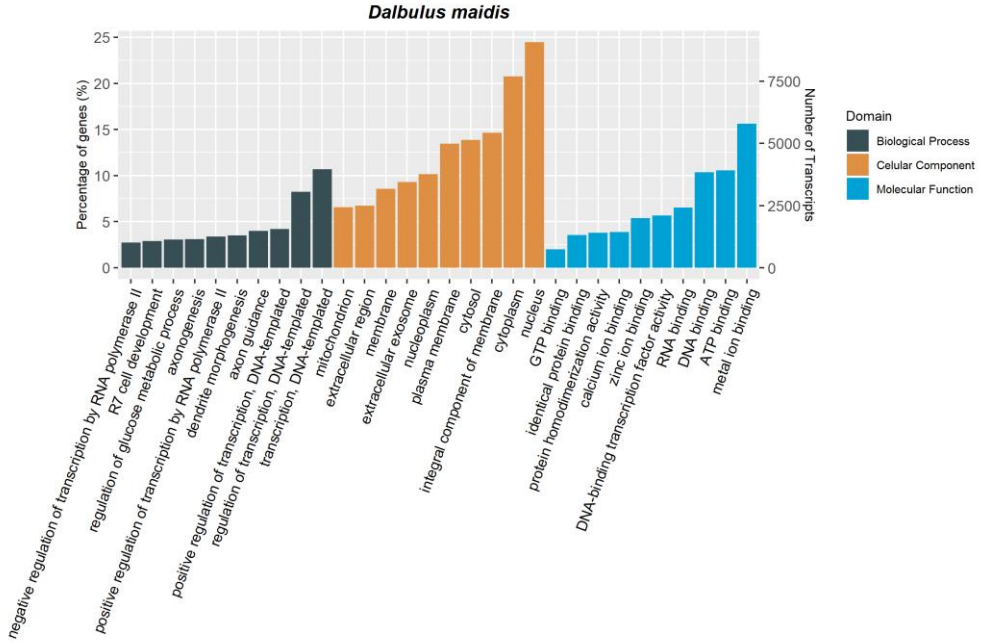

Figure 11. Top ten most represented GO terms of each domain in *D. maidis*

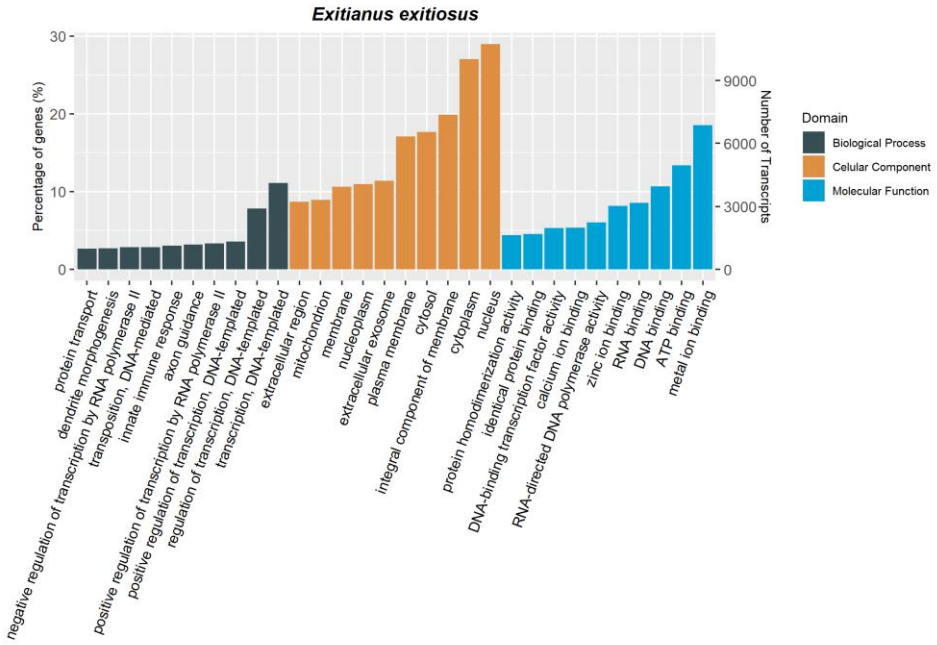

Figure 12. Top ten most represented GO terms of each domain in *E. exitiosus*

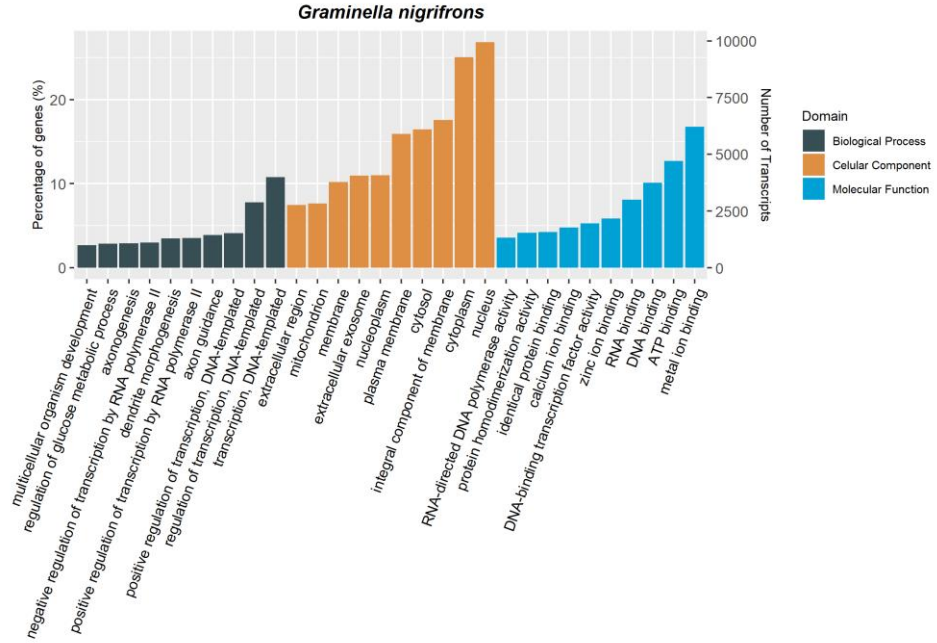

Figure 13. Top ten most represented GO terms of each domain in *G. nigrifrons*

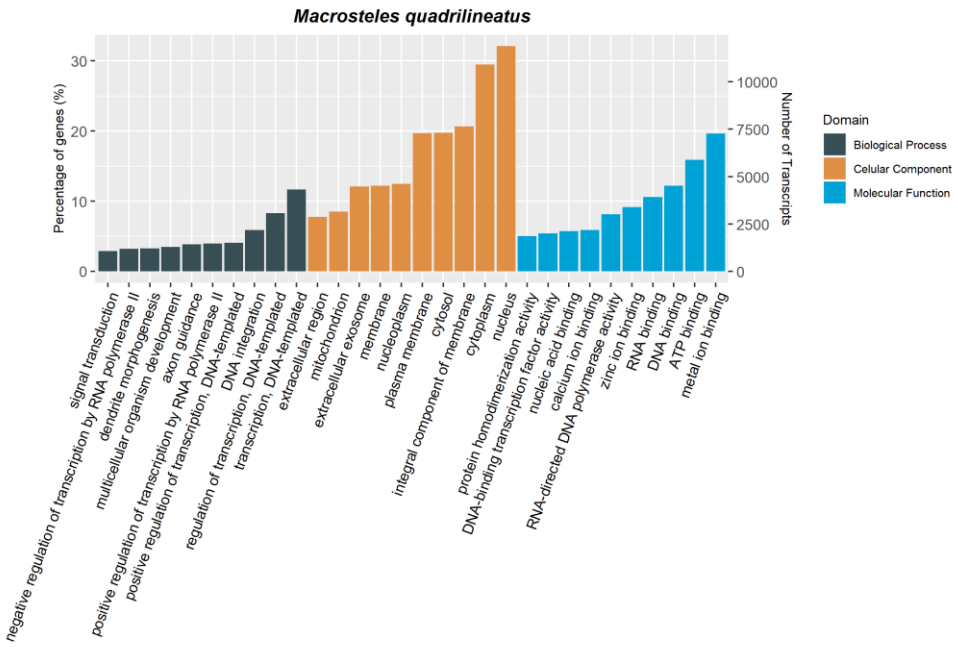

Figure 14. Top ten most represented GO terms of each domain in *M. quadrilineatus*

It can be noted that for all the species, GO terms related with cellular components were among the most represented. There was not a clear difference between the most represented GO terms among the non-vectors and vector leafhoppers.

The function of the transcripts was further classified based in the classification system of clusters of homolog genes (COG) in the eggNOG database. It is important to note that more than 40% of the transcripts were assigned to the “Function unknown” classification group (Figure 15 to Figure 20)

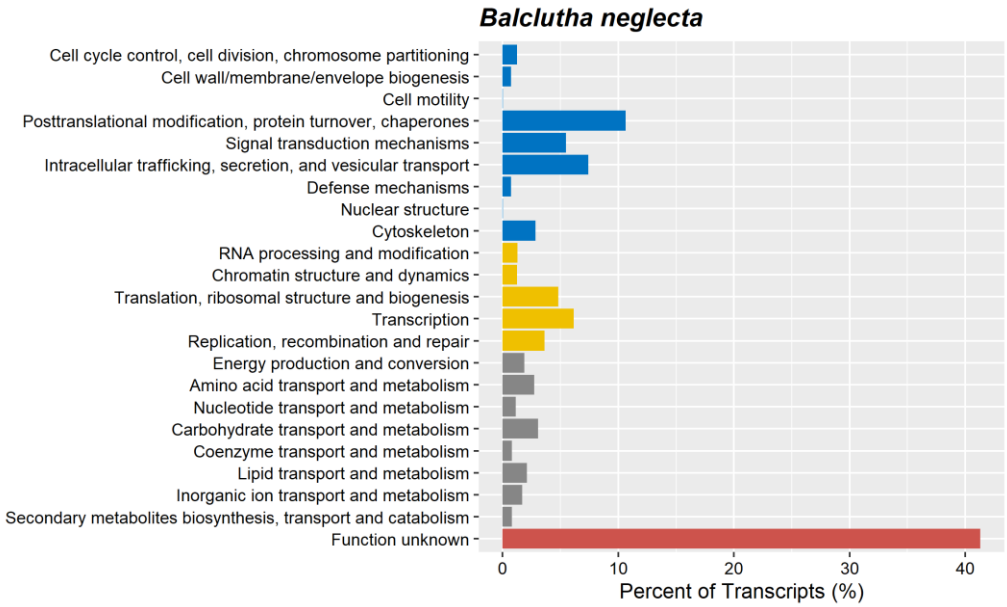

Figure 15. COG classification of the transcripts of *B. neglecta*

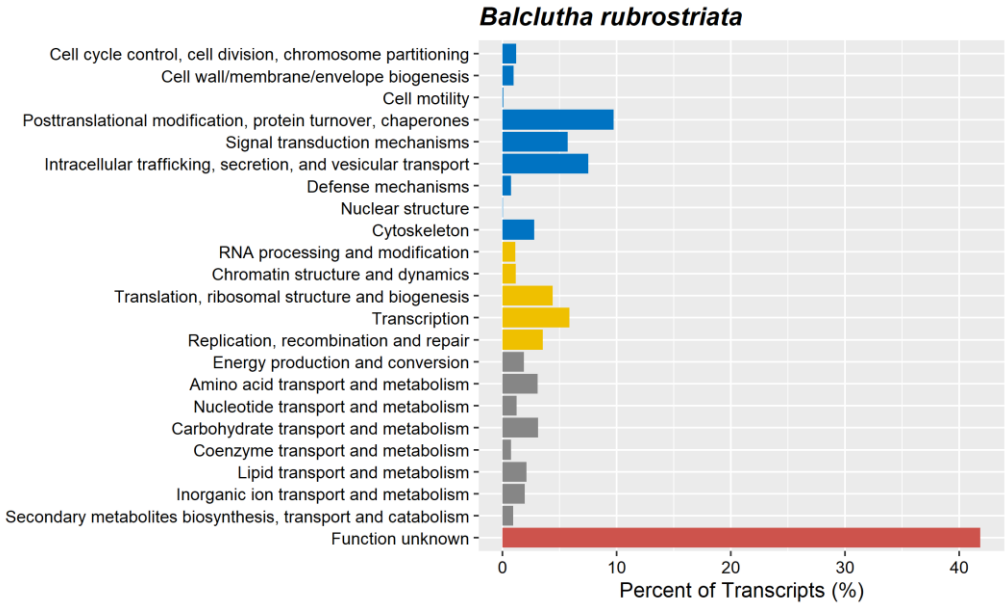

Figure 16. COG classification of the transcripts of *B. rubrostriata*

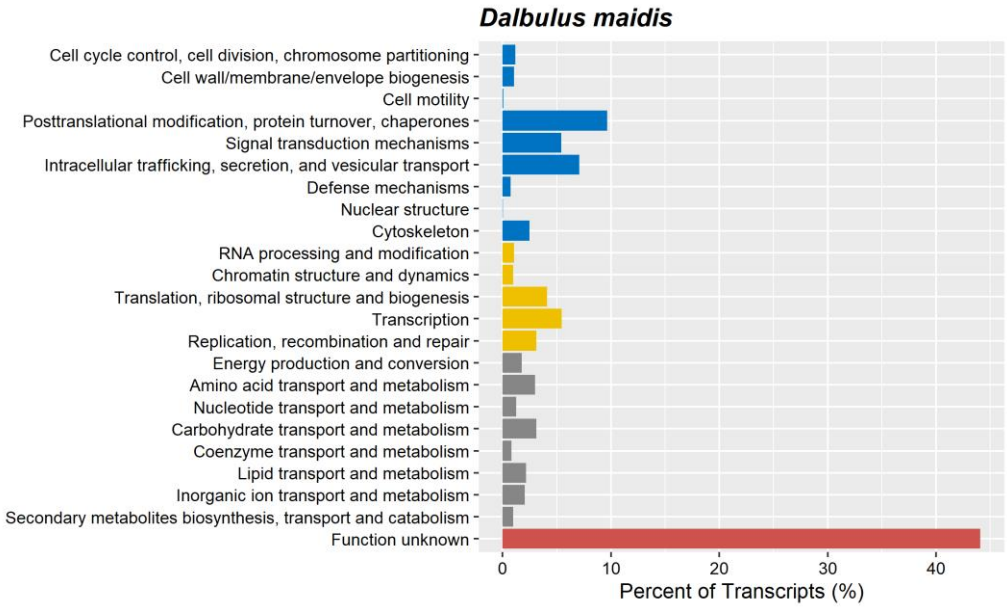

Figure 17. COG classification of the transcripts of *D. maidis*

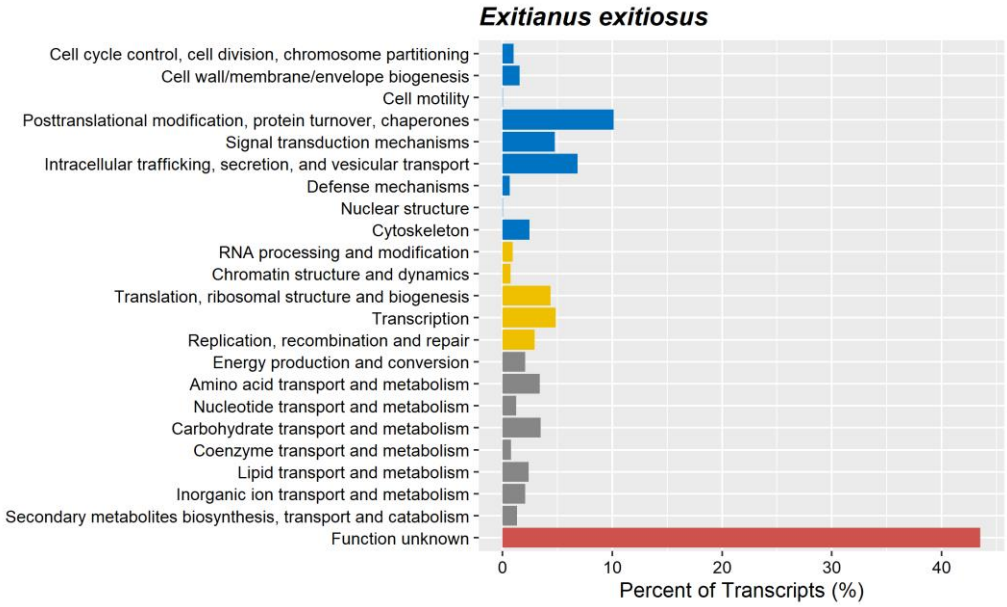

Figure 18. COG classification of the transcripts of *E. exitiosus*

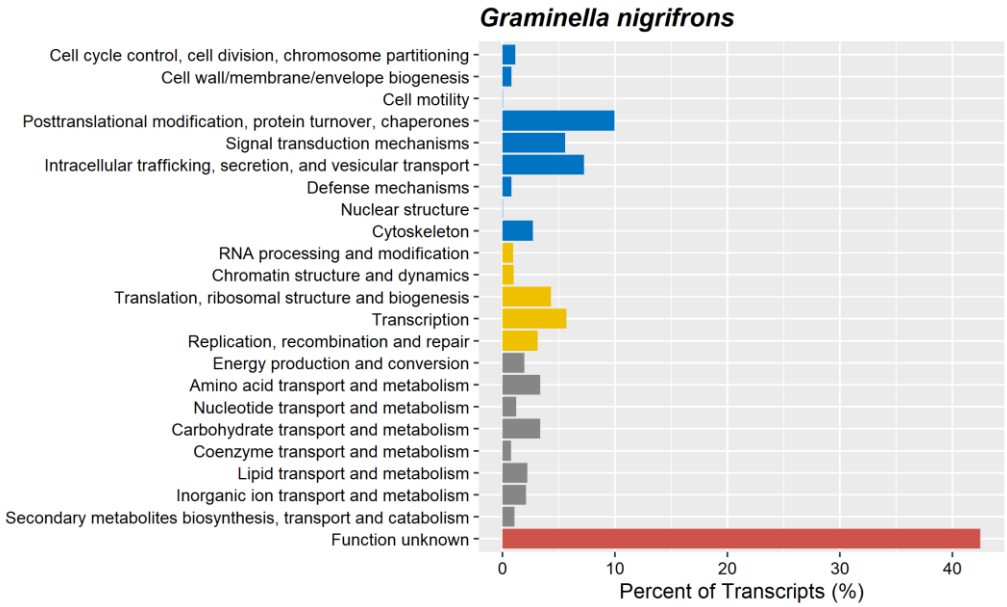

Figure 19. COG classification of the transcripts of *G. nigrifrons*

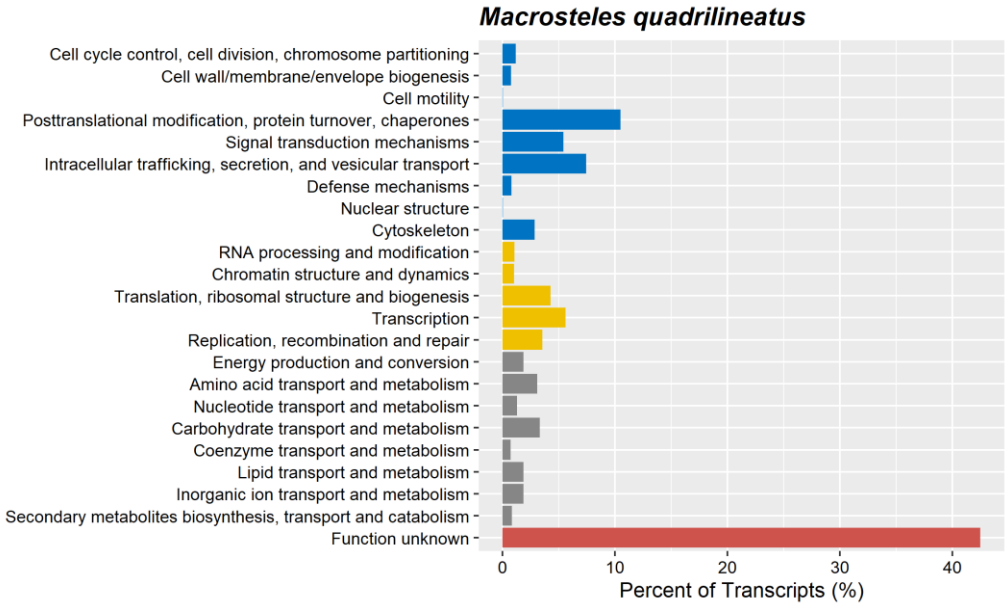

Figure 20. COG classification of the transcripts of *M. quadrilineatus*

Leaving aside transcripts assigned to the “Function Unknown” group, the other most represented categories are related with posttranslational modification of proteins

including signal transduction mechanisms, intercellular trafficking and vesicular transport. Similar to the Gene Ontology analysis, there is not a clear difference between vector and non-vector leafhoppers.

Finally, an analysis of the K terms from the KEGG database assigned to each transcript was done to determine in which metabolic or cellular processes the genes are involved. The 20 most represented KEGG pathways for each species are represented in Figure 21 to Figure 26.

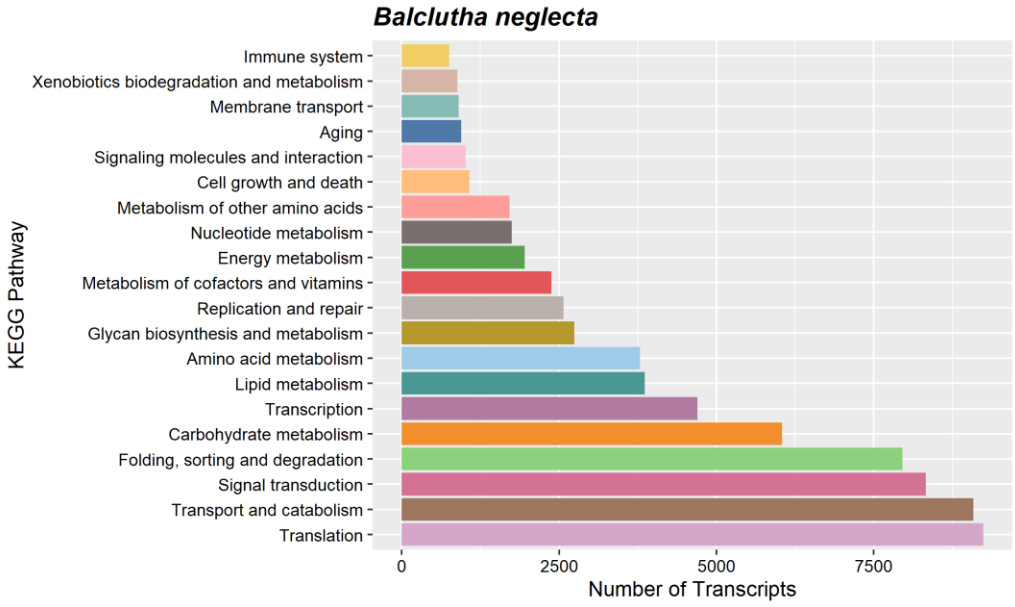

Figure 21. Top 20 most represented KEGG pathways in the *B. neglecta* transcriptome

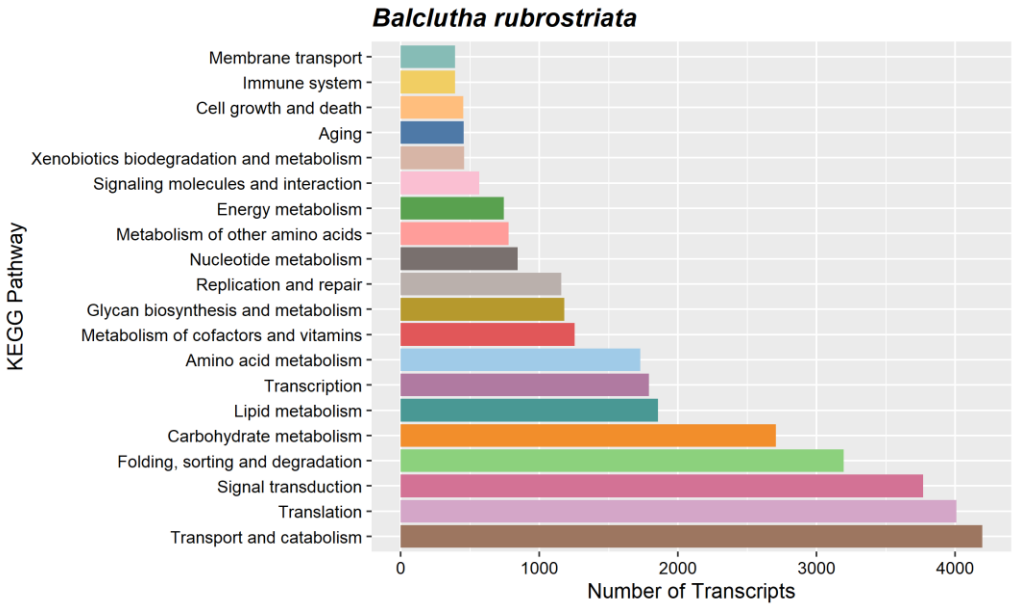

Figure 22. Top 20 most represented KEGG pathways in the *B. rubrostriata* transcriptome

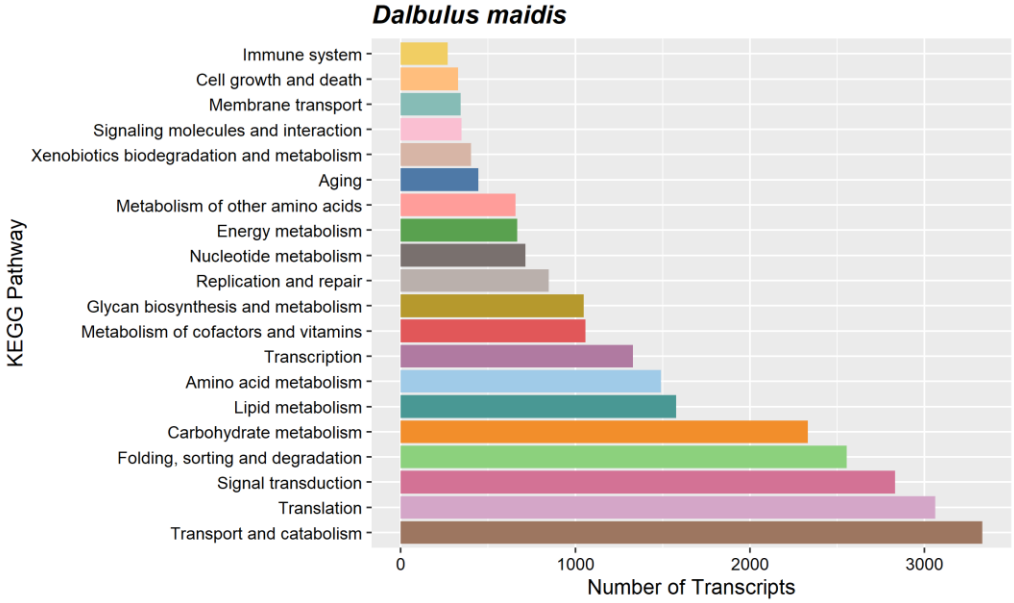

Figure 23. Top 20 most represented KEGG pathways in the *D. maidis* transcriptome

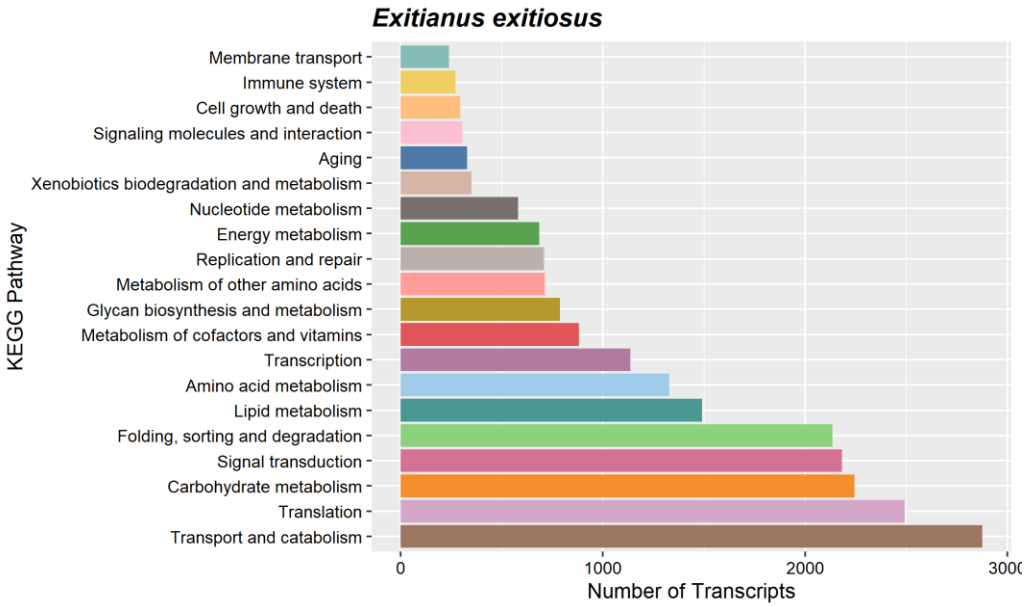

Figure 24. Top 20 most represented KEGG pathways in the *E. exitiosus* transcriptome

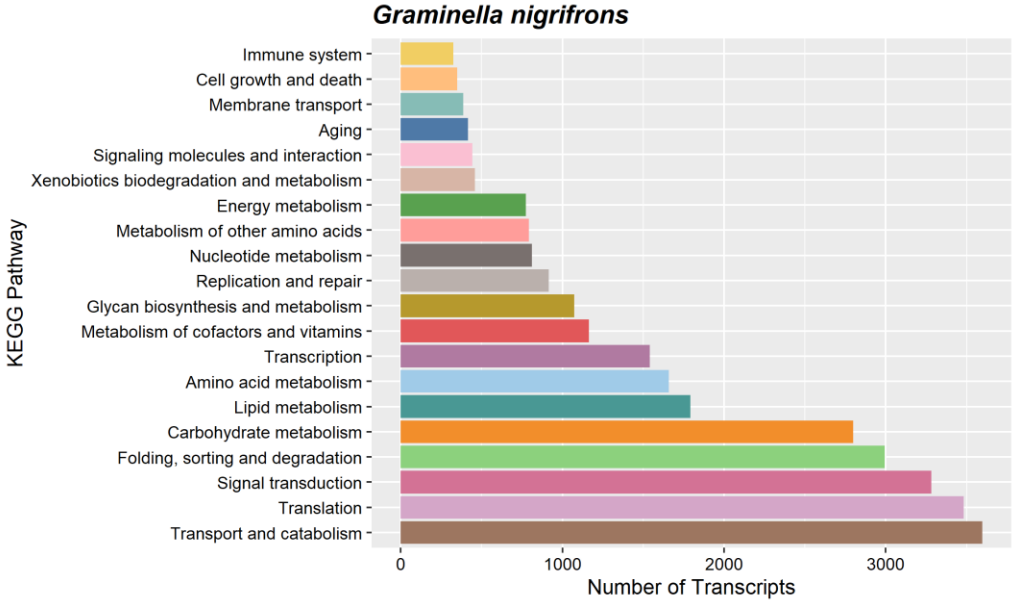

Figure 25. Top 20 most represented KEGG pathways in the *G. nigrifrons* transcriptome

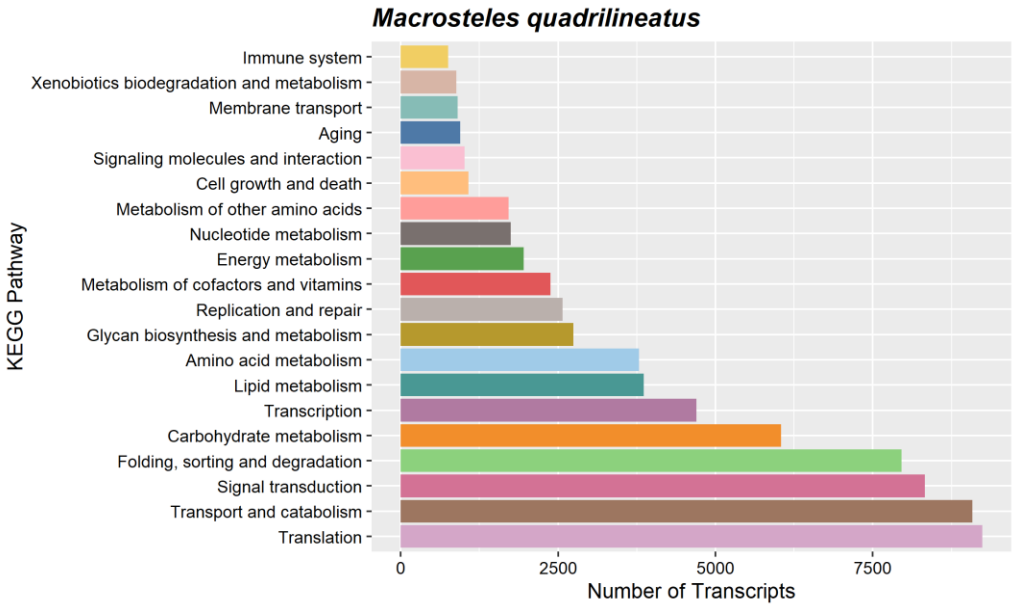

Figure 26. Top 20 most represented KEGG pathways in the *M. quadrilineatus* transcriptome

Similar to the results obtained in the eggNOG database, the analysis if the KEGG pathways also found that most of the transcripts are involved in transport, signal transduction and translation for all of the six species.

Because there was no apparent difference between the GO, KEGG and eggNOG annotations of the vector and non-vector leafhoppers, a clustering of the peptide sequences with more than 85% similarity was performed to clarify any differences between the two groups (Figure 27).

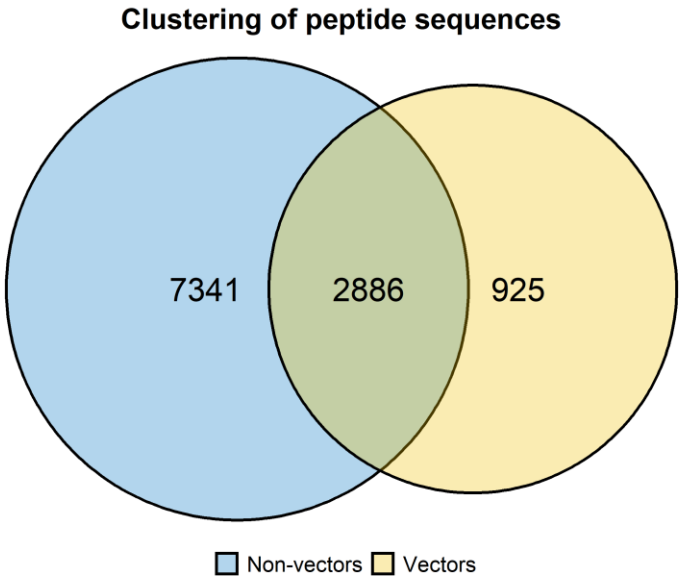

Figure 27. Number of peptides sequences that cluster together between vector and non-vector leafhoppers

A GO analysis of the proteins that did not cluster together in vector and non-vector leafhopper was performed (Figure 28) in order to search for genes or proteins that may be involved with vector competence.

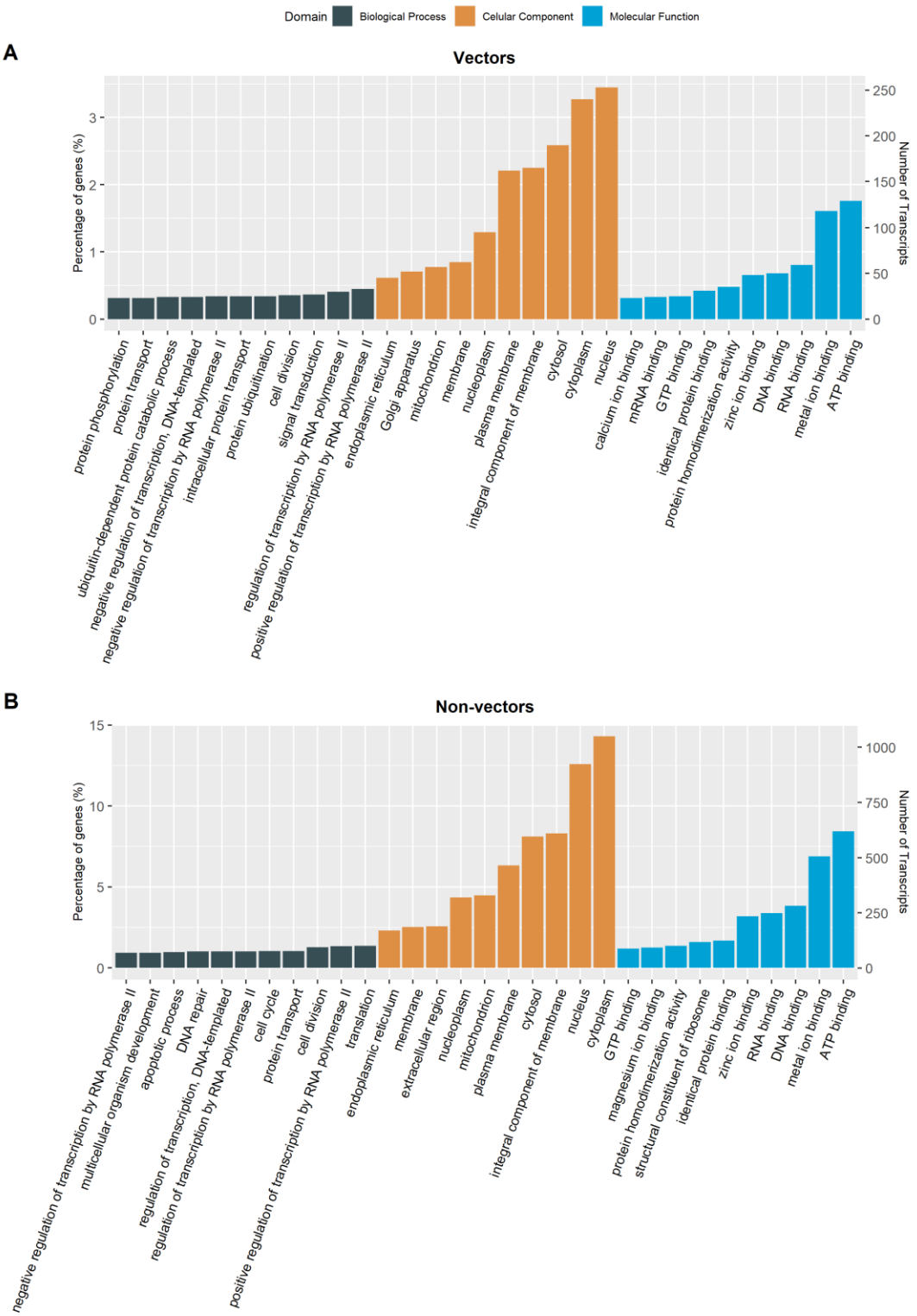

Figure 28. GO analysis of the peptide sequences with less than 85% similarity between vector and non-vector leafhoppers

DISCUSSION

The assembled draft genomes of each leafhopper species (Chapter III) required annotation to be able to fully realize all of the biological information that they could provide. The annotation of the leafhopper genomes began with the use of both the library of repeats from RepBase and a custom library of repeats from each of the leafhopper species. The amount of repetitive content that was identified using the repeat libraries from RepBase, was very low, between 2.67% to 3.52% of the genome. Studies in *Drosophila melanogaster* and *Tribolium castaneum* have found that the repetitive content in those insects represent 20% and 26% of their respective genome (Wang et al., 2008). In that regard, the amount of repetitive content that was identified using the custom libraries constructed with RepeatModeler, between 12.87% to 32.30% of the leafhopper genomes was closer to the expected real repetitive content of these genomes.

These repeat libraries were used to mask the repetitive content in the genome. The masking of these repetitive sequences was computationally challenging due to their variation in abundance, sequence signatures and length. The approaches available for the detection of TEs can be divided in three main categories: those that use repeat libraries, those that use signature-detection methods, and those that use *ab initio* consensus methods (Zeng et al., 2018). In this work, an *ab initio* predictor of TEs was used, RepeatModeler, coupled with a library-based method, RepeatMasker, which ultimately masked the repetitive content.

A genome needs a N50 larger than the median gene length to be produce an acceptable annotation. As an empirical rule, the median gene length for gene can be

estimated based on its genome size (Yandell and Ence, 2012). Thus, for the leafhopper draft genomes with sizes ranging between 0.85 to 1 Gbp, the expected median gene length for our species will be approximately 5000 to 6500 bp. Since all of the N50 of our genomes are at least 3 times smaller than the expected estimated median gene length, it can be assumed that obtaining a good annotation of the leafhopper draft genomes would be challenging.

The *ab initio* gene predictor, GeneMark-ES (Ter-Hovhannisyan et al., 2008), was used to try to annotate the leafhopper genomes without success. GeneMark-ES default parameters are set to ignore contigs with less than 50 kbp, which resulted in more than half of our assemblies not considered for the analysis. For that reason, in order to identify gene structures in the draft genomes, the software PASA was utilized. While, *ab initio* predictors use the genomic sequences together with statistical approaches to identify coding regions and gene signals (Picardi and Pesole, 2010), PASA uses the evidence from the transcriptome sequences to identify possible gene structures in the genome (Haas et al., 2003).

A number of possible genes were predicted using PASA for each of the draft genomes. However, more analysis is needed to fully identify those genes and to assign a possible function to each of them. Moreover, because the BUSCO assessment of the genomes found that they were incomplete it was not possible to obtain a complete structural annotation of the genome (i.e. how the genes look and where their location is in the genome).

Apart from the structural annotation of the genes, the final objective of the annotation process was to produce a functional annotation that revealed biological important information (Del Angel et al., 2018). A functional annotation of the assembled transcriptomes of the six species of leafhopper was carried out using the Trinotate pipeline. To be able to assign functions to the different transcripts, the transcriptomes of all the leafhopper species were subjected to a homology search against several genomic databases. However, less than half of the transcripts produced a match against the UniprotKB-Swiss-Prot. Studies in other organisms, such as plants (Liu et al., 2017) or mammals (Moreno-Santillán et al., 2019) had 60% of their transcripts matching against the same database using similar search parameters.

Studies that performed functional annotation of insects using BLAST against the NCBI nr database have higher number of significant hits than the number of hits obtained with the leafhopper transcriptomes against the UniprotKB-Swiss-Prot database during this study: 41.15% for the Colorado potato beetle, *Leptinoarsa decemlineata* (Kumar et al., 2014) and 55.52% for the cricket, *Grillus bimaculatus* (Zeng et al., 2013). The transcriptomes of other hemipterans also had low number of matches against the UniprotKB/Swiss-prot database, with 24% of matches for the *Adephocoris suturalis* (Tian et al., 2015) and an even lower fraction of genes for the brown marmorated stink bug, *Halyomorpha halys* (exact values cannot be determined because they only present the data of the “gold” and “silver” tier of transcripts) (Sparks et al., 2014).

Since error propagation is a real danger during annotation (Del Angel et al., 2018), the use of minimally or non-curated databases, such as the non-redundant database (nr) of

NCBI, was avoided. Instead only highly curated databases like the UniprotKB-Swiss-prot database were used in this study.

Through homology searches, GO terms, KEGG and eggNOG IDs were retrieved for each of the transcriptomes and were used to classify their possible function and the metabolic pathways they might be involved. The GO analysis for all of the species showed that most of the GO terms obtained from the transcripts belong to the cellular component domain. These differ from other insect studies where the most represented domain is biological process (Kumar et al., 2014; Tian et al., 2015).

The classification of the transcripts into the cluster of orthologous genes (COG) families of the eggNOG database, showed that most of the genes of the leafhopper transcriptomes fall under the category of "Function Unknown". This differ from other studies where even though some genes may not have a specific function, they at least are classified into the "General function prediction" category (Liu et al., 2017; Tian et al., 2015; Zhang et al., 2015). Thus, the annotation of these leafhopper genomes is more conservative than those other insects and will require further studies to determine the function of those genes.

Because it was not possible to observe differences between the most represented groups of GO terms, KEGG and eggNOG IDs between the two non-vector and the four vector leafhoppers, a clustering analysis was performed to determine which peptide sequences were different between the two groups. A clustering analysis of the RecA protein from multiple organisms found that apparently there is around 70% of similarity among the peptide sequences of organisms between the same subfamily and around 75% of

similarity for organisms in the same genus (Edwards, 2016). Taking that into account, the clustering of the peptides was set at 85% similarity to focus only in peptide sequences that differ not only because the insects belong to distinct genera, but because their functions may be different.

A total of 925 peptide sequences were found to be shared among the vector species that had less than 85% similarity to peptides in the non-vector species. Conversely, 7,341 peptides were shared by the non-vectors with less than 85% similarity to the peptides in the vectors. Similar to the results of the GO analysis of each species, most of the peptides shared by the vectors and the non-vectors were assigned GO terms that fell under the cellular component domain.

Within the molecular function domain one of the most prevalent terms in both vector and non-vector leafhoppers was the ATP binding category. A protein that was identified by homology to the UniprotKB/Swiss-prot database as Actin-like protein 2-B belongs to three of the most represented GO categories in the vector species: the cellular component “nucleus”, “cytoplasm” and the molecular function “ATP binding”. Studies in *Circulifer haematoceps* have found that *Spiroplasma citri* interact with actin microfilaments during the internalization of the pathogen into the leafhopper cells (Labroussaa et al., 2010; Labroussaa et al., 2011). Similarly, another study in *Candidatus* Phytoplasma asteris, found that the interaction between the host actin with the pathogen AMP that occurs during the phytoplasma co-localization with the actin filaments in the insect gut, may play a role in the species-species recognition (Suzuki et al., 2006).

It has also been demonstrated that there is a vector-specific interaction between the phytoplasma AMP and the insect ATP synthase, supporting the idea that actin is involved in the internalization and mobilization of the pathogen (Galetto et al., 2011). A peptide sequence that codes for the alpha subunit of the ATP synthase, is also one of the sequences shared among vectors with less than 85% similarity to the non-vectors. Analogous to the actin-like protein mentioned before, this ATP synthase alpha subunit was also assigned two of the most represented GO terms among the vectors: "ATP binding" and "plasma membrane".

Another important GO term is cell adhesion, which is the 12th most prevalent term from the "biological process" domain. One of the vector peptides, which has been identified by homology as vinculin, has been associated with this GO term and also seems to be involved in vector transmission. A proteomics study in the transmission of the *Barley yellow dwarf virus* by the Bird cherry-oat aphid, *Rhopalosiphum padi* found that myosin, paramyosin and vinculin, among other proteins, were up-regulated in the aphid after being exposed to the virus (Wang et al., 2015). In other study, vinculin, α-integrin, paxillin and other membrane-cytoeskeleton proteins that are known to be involved in the focal-cell adhesion of pathogenic bacteria were found to be upregulated in the psyllid, *Bactricera trigonica* infected with *Candidatus* Liberibacter soleanacearum (Ghosh et al., 2019). Peptide sequences identified as myosin, paramyosin, α-integrin and paxillin were also included in the peptides found to be shared by the vectors with more than 85% similarity during this study.

Similar to the ones already mentioned, among the 924 peptide sequences shared by the vectors there must be more proteins that are involved in several of the process that

occur during insect transmission. Even more information about the specific proteins that may be involved in the interaction between insect vectors and their pathogens can be discovered with the use of transcriptomics and RNA-seq for differential expression analysis (Chen et al., 2016; Chen et al., 2012; Rotenberg et al., 2015).

However, the use of transcriptomics only allows for the identification *in silico* of gene products and their putative function. Once a gene of interest has been determined functional studies are required to fully characterize a protein and their function within an organism (Bellés, 2009). These functional analyses may include: the cloning and transfection of those proteins into yeast (Zhou et al., 2008) or into eukaryotic cell lines (Grohmann et al., 2003); the use of genome editing techniques such as transcription activator-like effector nucleases (TALENs) (Ma et al., 2012) and the CRISPR/Cas9 system (Zhu et al., 2016); or the use of post-transcriptional silencing methods such as interference RNAi (Yao et al., 2013). Additional functional studies are needed to furher expand on the results of this annotation effort.

CHAPTER V

KNOCK-DOWN OF THE *Dalbulus maidis* LACCASE-1 GENE USING THE INTERFERENCE RNAi MECHANISM

ABSTRACT

Dalbulus maidis is a specialist leafhopper that feeds on maize and can be found in the southern United States, Central and South America. It is a significant pest of corn because it can transmit the three stunting pathogens involved in the "corn stunting complex". Interference RNA is a mechanism that produces post-transcriptional silencing of a target transcript. The silencing is triggered by the presence of dsRNA and will result in the silencing of genes whose sequences are homologous to that dsRNA. It is a powerful tool that allows the study of gene functions and the linkage to a specific phenotype. Several studies have demonstrated that for RNAi there is not a single protocol that is suitable for all species. It is therefore necessary to develop a specific protocol to work with each species.

In consequence, the objective of this study is to demonstrate the use of RNAi in *Dalbulus maidis* by knocking down the expression of the laccase-1 gene. First, a stability assay of the expression of several housekeeping genes was done, using RT-qPCR, to

determine which is the best reference gene for this species. The analysis of the Ct data with four different software packages resulted in the electron transfer flavoprotein (ETF) being selected as the reference gene. For the RNAi assay, third and fourth instar leafhoppers were fed laccase-1 dsRNA over six consecutive days, while the control group was fed GFP dsRNA. Changes in gene expression were measured by RT-qPCR. This analysis found a significant reduction of the transcript levels of the laccase-1 gene after 4 and 6 days compared to the control. Neither group developed into adulthood, and there was not any apparent difference in phenotype between the two groups. However, because reduction in the levels of the laccase-1 gene are expected to cause an impairment in the feeding ability, more detailed assays are required to determine the overall effect of this RNAi silencing.

INTRODUCTION

The corn leafhopper, *Dalbulus maidis*, is a specialist insect whose field hosts are plants of the genus *Zea*, including the wild ancestral teosinte species as well as the domesticated maize, *Zea mays* (Nault, 1998). Some studies have demonstrated that it can also survive in gamagrass, *Tripsacim dactyloides*, and to a lesser degree on Johnsongrass, *Sorghum halepense* (Pitre, 1970; Pitre et al., 1967). The distribution of *D. maidis* ranges from the southern United States to South America wherever maize is grown (de Oliveira et al., 2004). Corn leafhoppers can be collected on maize seedlings as soon as the beginning of the wet season, suggesting that they overwinter as adults. In locations where maize is cultivated with irrigation though the dry winter season, *D. maidis* survives by feeding on those crops. When corn is not available, it is believed that females survive by feeding on perennial grasses and move to corn as soon as it is planted (Moya-Raygoza et al., 2007).

Generally, *D. maidis* nymphs go through five instars before emerging as adults. Based on the temperature of its host plant, the life cycle of the corn leafhopper can show some variation. The developmental time to grow from first to fifth instar diminishes with temperature, at 10 °C it ranges between 12 to 34 days, while at 32 °C it takes them from 2 to 4 days. In regards of adult longevity, at 10 °C males live for 67 days and females for 38 days, at 16 °C males live for 107 days and females for 52 days, at 27 °C longevity is 78 days for males and 30 days for females, and at 32 °C it is 16 days for males and 10 days for females. Unlike nymph developmental times, the number of eggs oviposited raises with temperature, with an average of 4 eggs per female per day at 16 °C to 15 eggs per female per day at 27 °C (Tsai, 2008).

Dalbulus maidis does not cause severe damage to the maize by feeding from its sap (Summers et al., 2004). However, it is considered a significant pest of corn because of its ability to transmit three stunting pathogens involved in the "corn stunting complex": corn stunt spiroplasma (CSS), maize bushy stunt phytoplasma (MBSP) and maize rayado fino marafivirus (MRFM) (Summers et al., 2004). The three pathogens are persistently transmitted by the corn leafhopper. Once acquired, these pathogens require of an incubation period of 14 to 21 days inside the vector before being able to be inoculated in a new plant. The length of the incubation period varies depending on the isolate and titre of the pathogen as well as the biotype and age of the vector (Nault, 1998). CSS has been reported as the most important pathogen affecting corn in Central America, Peru, Brazil and Argentina, where the severity of the disease is high, producing stunting in 100% of the plants (Virla et al., 2004).

Most of the genomics studies related to *Dalbulus maidis* have focused on sequencing and analysis of the genomes of the pathogens harbored by this vector (Hammond and Ramirez, 2001) and on the endosymbionts that occur naturally in this insect (Brentassi et al., 2017; Chang et al., 2015). In this regard, the data presented in herein is an initial attempt to use the assembled genome of *D. maidis* to support a functional study of this species.

Similar to what is happening with sequencing efforts of other organisms, the new set of genomic data available for six leafhoppers species provides the scientific community with a great number of genes for which a function has yet to be be assigned. One very useful technique for functional studies is interference RNA (RNAi) (Bellés, 2009). RNAi is a mechanism that produces post-transcriptional silencing of the target transcript (Bass, 2000). This mechanism is triggered by the presence of endogenous or exogenous dsRNA and will result in the silencing of endogenous genes whose sequences are homologous to that same dsRNA (Tomoyasu et al., 2008). During this process, the dsRNA is cut by the Dicer RNAse III protein into small RNAs, either small interference RNAs (siRNAs) or miRNAs. Then, those siRNAs are loaded into the RNA-induced silencing complexes (RISC) to be used as a guide to find target mRNAs and cleave them (Filipowicz, 2005).

The advantage of using RNAi is that it may allow for comparison of phenotypic variations associated with the loss of a gene function in species for which it is not possible to obtain mutants (Brown et al., 1999).

RNAi was first discovered in the nematode *Caenorhabditis elegans* (Fire et al., 1998) and has since been adapted to function in several other organisms, including insects

of various orders such as: Diptera (Dietzl et al., 2007), Coleoptera (Tomoyasu et al., 2008), Lepidoptera (Terenius et al., 2011) and Hemiptera (Chen et al., 2010; Matsumoto and Hattori, 2016). The ability to study gene functions and to link them to a specific phenotype by using the RNAi mechanism has been fundamental in pushing forward the development of functional genomics (Scott et al., 2013).

A successful RNAi experiment depends upon the formation of a specific RNAi molecule that targets the specific gene of interest. To achieve this the preparation and delivery of the dsRNA are crucial steps (Scott et al., 2013). In most studies, long dsRNA (300-500 bp) obtained from a section of the gene of interest are used to trigger the RNAi cascade (Yao et al., 2013). Delivery of the dsRNA is considered the most limiting factor of the RNAi process. The most common routes for delivery of dsRNA to insects is by feeding or injection. Even though microinjection is a widely used and efficient method to deliver dsRNA directly to the hemolymph, it is time consuming, delicate and requires optimization. On the other hand, ingestion of dsRNA is less time consuming, less invasive and can be used for high-throughput gene screening. However, ingestion of dsRNA may not be suitable for all species and may be less efficient in triggering the RNAi mechanism (Yu et al., 2013). Studies in different species, such as *Rhodnius prolixus* found that larger quantities of dsRNA are required when feeding to induce RNAi (Araujo et al., 2006). This variation in efficiency between feeding and microinjecting dsRNA can be due to differences in the gut environment among different insects (Yu et al., 2013).

RNAi may not work the same in different organisms (Brown et al., 1999). RNAi experiments have been carried out in different insect orders, however not all species show the same degree of sensitivity to this method of silencing. For example, polyneopterans

such as *Blatella germanica*, *Gryllus bimaculatus* and *Oncopeltus fasciatus*, have demonstrated a high sensitivity to RNAi. On the other hand, in dipterans, such as *Drosophila melanogaster*, and lepidopterans, such as *Bombyx mori* and *Manduca sexta*, there are tissues that are more resistant to RNAi (Bellés, 2009)

As more data is accumulated, it has been demonstrated that there is not a single RNAi protocol that is suitable to all species (Scott et al., 2013). Thus, it is necessary to develop a specific protocol for work with each individual species.

This study used RNAi in *D. maidis* to knock-down the laccase-1 gene to demonstrate the effectiveness of this mechanism in this species. This knowledge will allow future researchers to utilize the information available in the draft genome of *Dalbulus maidis* to discover and study the function of any gene of interest.

MATERIAL AND METHODS

Insect rearing

Dalbulus maidis used in this assay were obtained from colonies that were originally collected by Dr. Ismael Badillo-Vargas, Texas A&M University (TAMU) in South Texas. The colony has been maintained at Oklahoma State University for more than two years. All insects were grown on young corn plants, unless noted otherwise. Insects were kept in insect proof cages covered with nylon mesh. Specific information on insect maintenance during each assay will be described below.

Housekeeping genes stability assay

To determine which housekeeping gene was more stable for use as a reference in the gene expression assay, the stability of five different housekeeping genes, actin (*Actin*), electron transfer flavoprotein (*ETF*), ribosomal protein 18 (*18S*), TATA-binding protein (*TATA*) and β-tubulin (TUB) was tested with RT-qPCR. All the primers used in this study where designed using Primer3 (Koressaar et al., 2018; Koressaar and Remm, 2007; Untergasser et al., 2012). Primers for the housekeeping genes and laccase-1 (Table 11) were designed from sequences obtained from the transcriptome of *Dalbulus maidis*.

Table 11. Primers designed to detect the leafhopper housekeeping genes and laccase-1

Target	Primer code	Primer sequence (5'-3')	Amplicon size (bp)	Length (bp)	Tm (°C)	GC%	any	3'
Lac-1	Dm_lac_F2	CCTGGGTTGTTAGCGTGGA	97	19	59.63	57.89	3.00	0.00
	Dm_lac_R2	AACTGGTGAGGAATCTGGAGG		21	59.09	52.38	3.00	0.00
ETF	Dm_ETF-F	GGAAGTGGTCCGTCGTGTT	144	19	59.93	57.89	3.00	0.00
	Dm_ETF-R	GAAGGTTGGAGAAGCGTGTG		20	59.13	55.00	2.00	0.00
18s	Dm_18s_F	TCAGTTGACCTCCAGTTCCC	120	20	58.94	55.00	4.00	0.00
	Dm_18s_R	GTGTGCTGACCTCTCACTCT		20	59.03	55.00	3.00	0.00
TATA	Dm_TATA_F	ACTAGCGAGGACCAATCACAC	96	21	59.23	52.48	4.00	0.00
	Dm_TATA_R	GACTATGCGGAAGTTGGAGAA		21	59.33	47.62	4.00	0.00
actin	Dm_actin_F	AGGCCAACAGGGAGAAGA	95	18	58.29	59.00	4.00	0.00
	Dm_actin_R	AGCGTACAGGGAGAGGACA		19	58.39	58.90	4.00	0.00
TUB	Dm_TUB_F	TACGGACCCATTCGCACTT	107	19	59.02	52.63	3.00	0.00
	Dm_TUB_R	GGCTGTGACGCTTTCTTCTC		20	59.21	55.00	3.00	0.00

Total RNA of *D. maidis* from three different developmental stages, early nymphs (1st to 3rd instar), late nymphs (4th and 5th instar) and adults, was extracted using the E.Z.N.A.® Mollusk RNA Kit (Omega Bio-Tek Inc., Norcross, GA), with the following modification to the manufacturer instructions: the volume of all reagents was scaled down to one quarter of the amount detailed in the product manual. The quantification of the RNA

was done using a Qubit 3.0 spectrophotometer with the Qubit RNA Broad Range Assay Kit (Life Technologies, Carlsbad, CA).

cDNA was synthetized using the Verso cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA) with RT-enhancer, to avoid DNA carryover, using 1 μg of RNA as starting material according to the manufacturer's instructions.

Stability of the housekeeping genes was tested by performing a RT-qPCR from samples of total RNA extracted from *D. maidis* using PowerUP SYBR Green Master Mix (Thermo Fisher Scientific, Waltham, MA) in a RotorGene 6000 (Corbett, Sydney, Australia). Reactions were carried out in a final volume of 10 μL using 1 μL of the synthetized cDNA as template and were done in triplicate. PCR cycle and conditions can be seen in Table 12.

Table 12. Conditions for the RT-qPCR

Cycle	Time	Temperature	No. of cycles
Initial denaturation	4 min	95 °C	1
Denaturation	20 s	95 °C	35
Annealing	20 s	56 °C	
Extension	20 s	72 °C	
Final Extension	3 min	72°C	1

The resulting Ct data was analyzed using BestKeeper (Pfaffl et al., 2004), NormFinder (Andersen et al., 2004), geNorm (Vandesompele et al., 2002) and the web application RefFinder (Xie et al., 2012) to determine which housekeeping gene is the most stable.

dsRNA synthesis

The E-RNAi web application (Horn and Boutros, 2010) was used to design primers conjugated with the 23 bp T7 RNA polymerase promoter for the laccase-1 gene. Two reactions, one using the T7-forward primer and the "normal" reverse primer, and another using the "normal" forward primer and the T7-reverse primer were completed to obtain a pair of PCR products containing the T7 promoter in the forward strand and the T7 promoter in the reverse strand, respectively.

Those PCR products were used as a template in the synthesis of dsRNA using the T7 Ribomax^{TM} Express RNAi System (Promega, Madison, WI) according to the manufacturer's instructions. Quality of the dsRNA was determined by gel electrophoresis.

dsRNA feeding assay

In order to determine the efficiency of the dsRNA in silencing the laccase-1 gene of *D. maidis* a time-course experiment was performed. For the assay, two experimental groups were considered: the first one was given laccase-1 dsRNA (dsLac group) in a liquid diet, and to the second group, the control group, green-fluorescent protein (GFP) dsRNA (dsGFP group) was given in a liquid diet.

Fifteen 3rd to 4th instar nymphs of *Dalbulus maidis* were collected and placed into feeding cages consisting of 2-dram clear plastic measuring cups covered by two layers of stretched Parafilm. Insects were fed 50 µL of D10 feeding solution (10% sucrose, 0.2% fructose, 0.38% potassium phosphate, 0.03% magnesium chloride, 1% fetal bovine serum, pH 7.0; filter sterilized) (Alivizatos, 1982), mixed with 500 ng/µL of dsRNA specific to

each treatment group. The 50 µL of liquid diet were placed between the Parafilm layers and replaced every day. Three biological replicates were done for each experimental group.

Two live insects per cage were collected and pooled together at 2, 4 and 6 days after exposure. Total RNA was extracted and cDNA was synthetized immediately after collecting the insects using the same protocols previously mentioned. Samples were stored at -80 °C until used for RT-qPCR analysis.

RT-qPCR and data analysis

Reverse transcriptase quantitative PCR (RT-qPCR) was performed to analyze the relative expression of the target gene, *laccase-1*, against the housekeeping gene previously determined to be the most stable, *ETF*. The reaction was performed using the same 3-step amplification cycle used to assay the housekeeping genes (Table 12). RT-qPCR reactions were done in a final volume of 10 µL using 1 µL of the synthetized cDNA as template with PowerUP SYBR Green Master Mix (Thermo Fisher Scientific, Waltham, MA) in a RotorGene 6000 (Corbett, Sydney, Australia).

The primers used to amplify the laccase-1 transcripts (Table 11) were designed so they do not overlap with the region used for the synthesis of dsRNA in order to avoid an overestimation of the transcript. The relative expression of the target against the negative control (GFP dsRNA group) was analyzed by the 2-ΔΔCt method (Livak and Schmittgen, 2001), using the Bioconductor package *ddCt* (Zhang et al., 2015). the difference in relative expression was tested using the Student's *t*-test.

RESULTS

Housekeeping genes stability assay

The expression of the selected housekeeping genes was verified by endpoint PCR during the different developmental stages of *Dalbulus maidis*. In Figure 29, it is possible to observe, that primers designed for β-tubulin (TUB) did not work and thus that gene was withdrawn from this study.

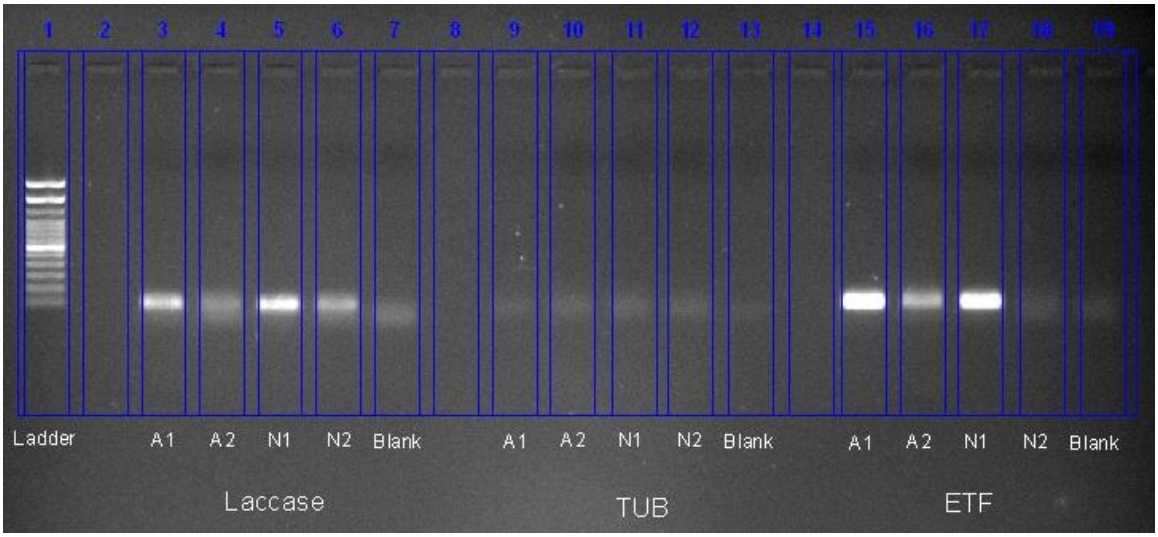

Figure 29. Expression verification of some of the selected housekeeping genes, β-tubulin (TUB), electron transfer flavoprotein (ETF), and the laccase-1 gene

RT-qPCR reactions to determine the stability of the four housekeeping genes during three developmental stages of *Dalbulus maidis* were carried out as described in the materials and methods section. The Ct values of each sample were obtained using the Rotor-Gene 6000 Series Software 1.7 and analyzed using four different software packages.

BestKeeper is an Excel-based tool that searches for the most stable reference gene based on the standard deviation (Std. Dev) of the raw Ct values of each candidate reference gene. The smaller the std. deviation, the higher the stability (Pfaffl et al., 2004).

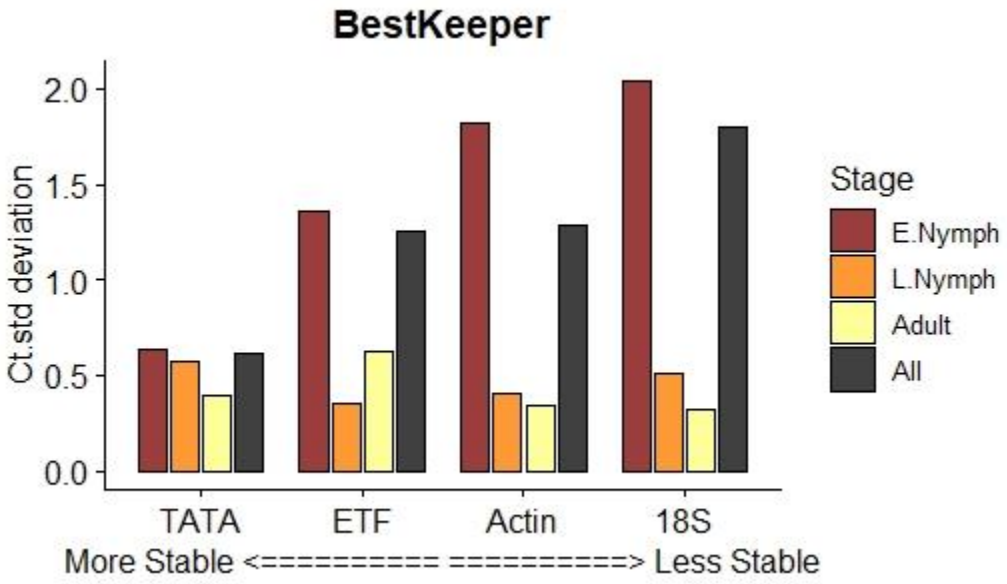

Figure 30. Analysis of the standard deviation of the Ct values of the four candidate genes during each developmental stage using BestKeeper.

Figure 30 shows that, the most stable gene is *TATA* with a Ct std. deviation of 0.62, for the combined dataset, according with BestKeeper. This same gene also seems to be the most stably expressed in all developmental stages. Nonetheless, the other three genes, *ETF*, *Actin* and *18S* also seem to have a high stability in late nymphs and adults, however the high instability of these genes during the early nymph stage contributes to their overall higher instability compared with *TATA*.

NormFinder is a Visual Basic Application (VBA) for Excel, that calculates a stability value for all candidate genes. It uses a strategy that allows the estimation of

variation not only between the candidate genes but also also between the sample subgroups of the sample dataset. The lower the stability value, the more stably expressed the gene (Andersen et al., 2004).

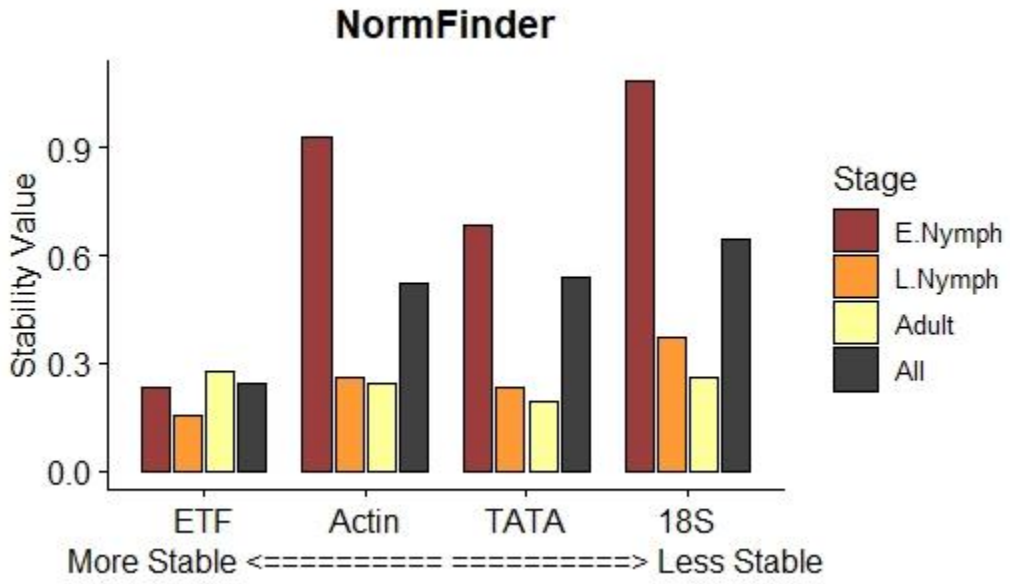

Figure 31. Stability values of the candidate reference genes estimated using NormFinder.

According to the software package NormFinder (Figure 31), the most stable gene is *ETF*, with a stability value of 0.241 for the combined dataset. *ETF* is also the most stably expressed gene during every developmental stage tested, except for Adults, for which the most stable gene is *TATA*. Similar to Bestkeeper, this analysis revealed that *Actin, TATA* and *18S* are not stable on the early nymphs.

The software geNorm is another VBA for Excel that measures gene stability by assigning a stability index M, defined as the average pairwise variation of a gene compared to all the other genes in the study. The lower the index M, the higher the stability (Vandesompele et al., 2002).

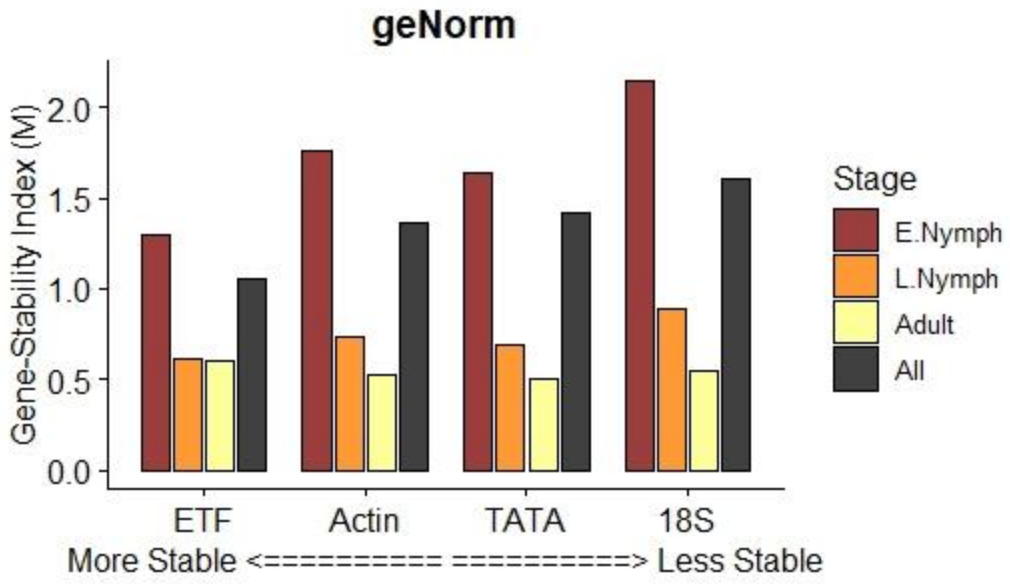

Figure 32. Stability Index (M) obtained with geNorm of all the candidate reference genes obtained.

Similar to NormFinder, geNorm determine that *ETF* is the gene with the higher stability (Figure 32), with a stability index of 1.05 for the combined dataset. Like with the other two software packages, the gene stability of the four genes is similar during the late nymph and adult stage, with the early nymph stage having the lowest gene stability for each gene.

RefFinder was used to integrate the data of all three software analyses and select the reference gene with the most stable expression. This comprehensive web-based tool uses all of the mentioned computational programs to rank the stability of the genes. It assigns a weight to each gene based on the rankings of the previous software and calculates a geometric mean of the weights to produce a final ranking (Xie et al., 2012). Based on this value, *ETF* is determined to be the reference gene with the most stable expression among all the candidate genes (Figure 33).

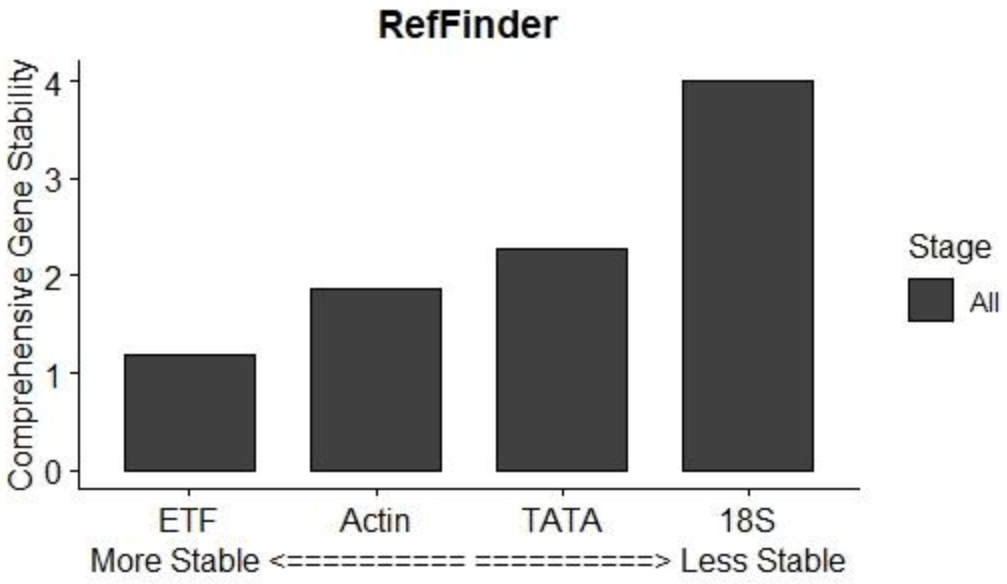

Figure 33. Comprehensive gene stability determined by RefFinder

RNAi assay

Third and fourth instar nymphs of *D. maidis* were feed dsRNA of *laccase-1* continuously for 6 days. Every two days, two insects were removed from each of the cages for RNA extraction, followed by the immediate synthesis of cDNA.

Relative expression analysis of the RT-qPCR data was done using the ΔΔCt method (Figure 34). The analysis revealed that at day 2, there was not any apparent reduction of the levels of *laccase-1* transcripts in the dsLac group compared to the dsGFP group (p=0.94) At day 4 however, there was a significant 0.7-fold reduction of *laccase-1* transcripts in the dsLac group compared to the dsGFP group (p=0.014). Finally, on day 6, there was also a significant 0.84-fold reduction of *laccase-1* levels between the dsGFP group and the dsLac group (p=0.046).

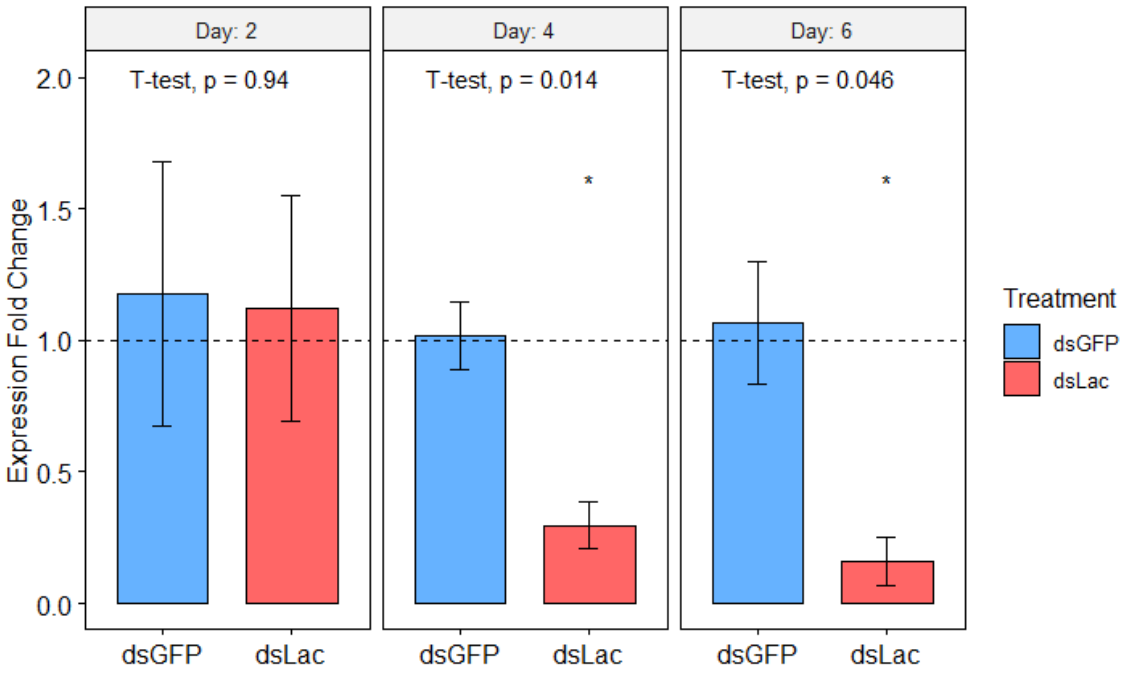

Figure 34. Relative expression levels of *laccase-1* transcripts normalized with *ETF* as reference gene. The bars represent the standard deviation (n=3). non-significant (ns): $p > 0.05$, *: $p <= 0.05$, **: $p < 0.01$, ***: $p < 0.001$

These results demonstrate that delivering dsRNA to *D. maidis* using a liquid diet is an effective method to reduce the transcript levels of a gene of interest.

DISCUSSION

Over the last decade, the use of RT-qPCR in genome and transcriptome expression studies in insect species has been increasing (Lü et al., 2018). However, in order to reliable quantify mRNA levels with RT-qPCR data, normalization of that data is a prerequisite. Of the several strategies available to normalize qPCR data, the use of reference genes is widely used because it provides sufficient control in every stage of the RT-qPCR and it is easy to use. The biggest problem with using reference genes to normalize RT-qPCR data is that

the stability of these genes can vary between different tissues, developmental stages, species or experimental conditions (Shakeel et al., 2018). Using an unsuitable reference gene can cause an inaccurate assessment of the expression of the target gene level of expression (Ma et al., 2016).

The four candidate genes, actin, electron transfer flavoprotein (ETF), ribosomal protein 18 (18S) and TATA-binding protein (TATA), tested in this study have been previously demonstrated to have a stable expression in other insect species (Ma et al., 2016; Maroniche et al., 2011; Singh et al., 2018; Xu et al., 2017). During this experiment, ETF demonstrated having the overall higher expression stability among the tested candidate genes. The electron transfer flavoprotein is an electron acceptor, part of the β-oxidation pathway of long fatty acids (Watmough and Frerman, 2010).

Studies in other insect species have found that ribosomal protein genes are commonly among the most stable genes to serve as candidate reference genes that have been tested (Shakeel et al., 2018). For example, RPS11, a gene member of the ribosomal protein family has been reported to have a stable expression among different organs, starvation and pesticide treatments in *Nilaparvata lugens* (Yuan et al., 2014). Regarding ETF, previous reports have found this gene to have a relatively good expression stability, but not better than RP13, another member of the ribosomal protein family (Singh et al., 2018).

The ribosomal protein 18 (18S) is a gene that has been previously used as a reference gene for gene expression analysis in the cotton aphid, *Aphis gossypii,* (Gong et al., 2014) and was found to produce accurate results when used as a reference gene for

feeding assays in the same aphid (Ma et al., 2016). However, our data ranked 18S as the least stable gene of the four candidates tested, further demonstrating the importance of the validation of gene expression stability prior to conducting a RT-qPCR experiment (Shakeel et al., 2018).

Once a suitable reference gene was established the RNAi assay was performed by feeding dsRNA of a portion of the laccase-1 gene to nymphs of *D. maidis*. RNAi is a versatile and highly useful tool to determine the function of a gene of interest. However, the response to RNAi may vary between different organisms (Scott et al., 2013), thus it is necessary to develop a protocol that works well in *D. maidis*.

For this assay, the dsRNA was administered through feeding as it allows an easy and quick delivery of the dsRNA without damaging the insect (Yu et al., 2013). Feeding dsRNA is done with the intention of triggering a systemic RNAi response, where the levels of the transcripts are decreased throughout all the insect body (Gu and Knipple, 2013). For this systemic response to occur, it is believed that the presence of the *sid-1* (systemic RNA interference deficient-1) gene is necessary and sufficient. This is a transmembrane protein that is able to import RNAi signals, such as dsRNA or siRNA, into different cells (Winston et al., 2002).

The presence of *sid-1* has been reported in several insect species such as *Schistocerca americana* (Dong and Friedrich, 2005), *Aphis gossypii* and *Sitobion avenae* however it is absent in Diptera species such as *Drosophila melanogaster* (Xu and Han, 2008). Phylogenetic analysis of *sid-1* suggest that this protein was present in the last common ancestor of insects, and was then lost in Antliophora (i.e. Mercoptera, Diptera,

Siphonaptera) (Dowling et al., 2016). Because the RNAi mechanism seems to be conserved, and previous studies were successful in triggering a systemic RNAi response by feeding dsRNA to *Peregrinus maidis* (Yao et al., 2013) and some other Hemiptera (Li et al., 2013), it was correctly assumed that this method of delivery was going to work in *D. maidis*.

During this assay, nymphs of *D. maidis* were continuously fed 0.5 μg/μL of laccase-1 dsRNA over 6 days. RT-qPCR results showed that there was not any significant difference of the transcript levels at day 2. However, on day 4 and on day 6 a 0.7-fold reduction and a 0.84-fold reduction, respectively, of lacasse-1 levels compared to the negative control group was observed. In mammalian cells, RNAi effect on gene expression can be seen within hours (Portnoy et al., 2011) and experiments in other insects have found effects as soon as 24 hours (Jaubert-Possamai et al., 2007). This could mean that *D. maidis* is not efficiently amplifying and spreading the RNAi signal among its cells. One explanation may be, that an insufficient amount of dsRNA is being ingested by the insect, as oral delivery usually requires higher doses of dsRNA (Scott et al., 2013). The difficulty in determining the exact amount of dsRNA that is ingested by the insect during feeding is one of the limitations of this mode of delivery (Yu et al., 2013)

Despite the reduction of transcripts levels of laccase-1 compared to the control (dsGFP group), it was not possible to observe any apparent phenotypical effect in the insects that was different between treatment groups. Even though the insects survived throughout all the feeding assay, they never molted into adults. Because this happened in both the dsLac group and the dsGFP group (control group), it is difficult to determine if

they never made it to adults as a result of feeding on laccase-1 dsRNA or because other environmental conditions.

Through studies carried out in *Nephotettix cincticeps*, which is also a Deltocephaline leafhopper, it was hypothesized that laccase-1 may be involved in the detoxification of plant-derived monolignols during insect feeding (Hattori et al., 2005). It was also found that there is an isoform of laccase-1 that is expressed exclusively in the salivary glands (Hattori et al., 2010). Therefore, it was expected that the silencing of the laccase-1 gene resulted in the impairment of the ability of *D. maidis* to feed.

In order to assess if the silencing of the laccase-1 gene had had an effect in the feeding ability of the corn leafhopper, other assays should be performed. For example, electrical penetration graph (EPG) technology can be used to observe changes in the probing behavior (Carpane et al., 2011). Even though RT-qPCR is a powerful tool to measure gene expression and determine whether or not there is a reduction in the transcript levels of the gene of interest, other assays are required to test the expected phenotypical effects and take full advantage of the RNAi mechanism as a tool to study gene functions and to link them to a specific phenotype.

REFERENCES

Abril, J. F., and Castellano, S. 2019. Genome Annotation. Pages 195-209 in: Encyclopedia of Bioinformatics and Computational Biology. S. Ranganathan, M. Gribskov, K. Nakai and C. Schönbach, eds. Academic Press, Oxford.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., and Galle, R. F. 2000. The genome sequence of Drosophila melanogaster. Science 287:2185-2195.

Alhakami, H., Mirebrahim, H., and Lonardi, S. 2017. A comparative evaluation of genome assembly reconciliation tools. Genome biology 18:93-93.

Alivizatos, A., and Markham, P. 1986. Acquisition and transmission of corn stunt spiroplasma by its leafhopper vector Dalbulus maidis. Annals of applied biology 108:535-544.

Alivizatos, S. 1982. Feeding behavior of the spiroplasma vectors Dalbulus maidis and Euscelidius variegatus in vivo and in vitro. Chronika Benaki Phytopathological Institutu (Greece).

Almeida, R. P., Blua, M. J., Lopes, J. R., and Purcell, A. H. 2005. Vector transmission of Xylella fastidiosa: applying fundamental knowledge to generate disease management strategies. Annals of the Entomological Society of America 98:775-786.

Ammar, E. D. 1994. Propagative transmission of plant and animal viruses by insects: factors affecting vector specificity and competence. Advances in Disease Vector Research:289-331.

Ammar, E.-D. 1985. Internal morphology and ultrastructure of leafhoppers and planthoppers. The leafhoppers and planthoppers:127-162.

Andersen, C., Ledet-Jensen, J., and Orntoft, T. 2004. Normalization of real-time quantitative RT-PCR data: a mode-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64:5245-5250.

Andreason, S. 2016. Discrimination and trancriptional response analysis of hemipteran phytopathogen vectors. Oklahoma State University.

Andreason, S., Gard, M., and Wayadande, A. 2015. Detection of the red-streaked leafhopper, Balclutha rubrostriata (Melichor)(Hemiptera: Cicadellidae), on an invasive grass species in Oklahoma. Southwestern Entomologist 40:479-485.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Araujo, R., Santos, A., Pinto, F., Gontijo, N., Lehane, M., and Pereira, M. 2006. RNA interference of the salivary gland nitrophorin 2 in the triatomine bug Rhodnius prolixus (Hemiptera: Reduviidae) by dsRNA ingestion or injection. Insect biochemistry and molecular biology 36:683-693.

Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., and Hartmann, E. M. 2014. Non-model organisms, a species endangered by proteogenomics. Journal of proteomics 105:5-18.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25-29.

Ashford, D., Smith, W., and Douglas, A. 2000. Living on a high sugar diet: the fate of sucrose ingested by a phloem-feeding insect, the pea aphid Acyrthosiphon pisum. Journal of Insect Physiology 46:335-341.

Backus, E. A. 1988. Sensory systems and behaviours which mediate hemipteran plant-feeding: a taxonomic overview. Journal of Insect Physiology 34:151159-157165.

Bao, W., Kojima, K. K., and Kohany, O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile Dna 6:11.

Bass, B. L. 2000. Double-stranded RNA as a template for gene silencing. Cell 101:235-238.

Baumann, P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. 59:155-189.

Beanland, L., Madden, L., Hoy, C., Miller, S., and Nault, L. 2005. Temporal distribution of aster leafhopper sex ratios and spatial pattern of aster yellows phytoplasma disease in lettuce. Annals of the Entomological Society of America 98:756-762.

Bellés, X. 2009. Beyond Drosophila: RNAi In Vivo and Functional Genomics in Insects. Annual Review of Entomology 55:111-128.

Bennett, G. M., and Moran, N. A. 2013. Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect. Genome Biology and Evolution 5:1675-1688.

Biedermann, R., Achtziger, R., Nickel, H., and Stewart, A. J. A. 2005. Conservation of Grassland Leafhoppers: A Brief Review. Journal of Insect Conservation 9:229-243.

Blanc, S., Drucker, M., and Uzest, M. 2014. Localizing viruses in their insect vectors. Annual review of phytopathology 52:403-425.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Bosco, D., Galetto, L., Leoncini, P., Saracco, P., Raccah, B., and Marzachi, C. 2007. Pattern of chrysanthemum yellows phytoplasma multiplication in three leafhopper vector species (Cicadellidae Deltocephalinae). Bulletin of Insectology 60:227.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., and Bairoch, A. 2007. UniProtKB/Swiss-Prot. Methods in molecular biology (Clifton, N.J.) 406:89-112.

Bradshaw, C. J., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.-M., Simard, F., and Courchamp, F. 2016. Massive yet grossly underestimated global costs of invasive insects. Nature communications 7:12986.

Bragard, C., Caciagli, P., Lemaire, O., Lopez-Moya, J. J., MacFarlane, S., Peters, D., Susi, P., and Torrance, L. 2013. Status and Prospects of Plant Virus Control Through Interference with Vector Transmission. Annual Review of Phytopathology 51:177-201.

Brentassi, M. E., Franco, E., Balatti, P., Medina, R., Bernabei, F., and Marino de Remes Lenicov, A. M. 2017. Bacteriomes of the corn leafhopper, Dalbulus maidis (DeLong & Wolcott, 1923) (Insecta, Hemiptera, Cicadellidae: Deltocephalinae) harbor Sulcia symbiont: molecular characterization, ultrastructure, and transovarial transmission. Protoplasma 254:1421-1429.

Brown, S. J., Mahaffey, J. P., Lorenzen, M. D., Denell, R. E., and Mahaffey, J. W. 1999. Using RNAi to investigate orthologous homeotic gene function during development of distantly related insects. Evolution & development 1:11-15.

Brown, S., Been, B., and McLaughlin, W. 2007. The lethal yellowing (16SrIV) group of phytoplasmas. Pest Technology 1:61-69.

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., and Davis, F. G. 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell reports 18:762-776.

Buchner, P. 1965. Endosymbiosis of animals with plant microorganisms.

Buntin, D. G. 1988. Species composition and phenology of leafhoppers (Homoptera: Cicadellidae) and planthoppers (Homoptera: Delphacidae) in bermudagrass forage. Journal of economic entomology 81:201-207.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421-421.

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18:188-196.

Cao, X., and Jiang, H. 2019. Integrated Modeling of Structural Genes Using MCuNovo. in: Insect Genomics: Methods and Protocols. S. Brown and M. Pfrender, eds. Humana Press, New York.

Carpane, P., Wayadande, A., Backus, E., Dolezal, W., and Fletcher, J. 2011. Characterization and correlation of new electrical penetration graph waveforms for the corn leafhopper (Hemiptera: Cicadellidae). Annals of the Entomological Society of America 104:515-525.

Cassone, B., Carter, F. M. C., Michel, A. P., Stewart, L. R., and Redinbaugh, M. G. 2014a. Genetic Insights into Graminella nigrifrons Competence for Maize fine streak virus Infection and Transmission. PloS one 9:e113529.

Cassone, B., Wijeratne, S., P Michel, A., R Stewart, L., Chen, Y., Yan, P., and Redinbaugh, M. 2014b. Virus-independent and common transcriptome responses of leafhopper vectors feeding on maize infected with semi-persistently and persistent propagatively transmitted viruses.

Chain, P., Grafham, D., Fulton, R., Fitzgerald, M., Hostetler, J., Muzny, D., Ali, J., Birren, B., Bruce, D., and Buhay, C. 2009. Genome project standards in a new era of sequencing. Science 326:236-237.

Chakravarthy, A. K. 2015. New Horizons in Insect Science: Towards Sustainable Pest Management. Springer.

Chang, H.-H., Cho, S.-T., Canale, M. C., Mugford, S. T., Lopes, J. R., Hogenhout, S. A., and Kuo, C.-H. 2015. Complete genome sequence of "Candidatus Sulcia muelleri" ML, an obligate nutritional symbiont of maize leafhopper (Dalbulus maidis). Genome Announc. 3:e01483-01414.

Chen, B., and Francki, R. I. B. 1990. Cucumovirus transmission by the aphid Myzus persicae is determined solely by the viral coat protein. Journal of General Virology 71:939-944.

Chen, B., Wang, X. H., Tu, Z., and Huang, Y. P. 2016. Leap forward with insect genomics. Insect science 23:332-334.

Chen, Q., Wang, H., Ren, T., Xie, L., and Wei, T. 2015. Interaction between non-structural protein Pns10 of rice dwarf virus and cytoplasmic actin of leafhoppers is correlated with insect vector specificity. Journal of General Virology 96:933-938.

Chen, T., Hou, J., and Lin, Y. 2016. Transcriptome datasets supply basic gene information for RNAi pest management and gene functional studies in Nephotettix cincticeps (Uhler). Journal of Integrative Agriculture 15:840-847.

Chen, Y. 2013. Molecular interactions between Maize fine streak virus and insect vector, Graminella nigrifrons. The Ohio State University.

Chen, Y., Cassone, B. J., Bai, X., Redinbaugh, M. G., and Michel, A. P. 2012. Transcriptome of the Plant Virus Vector Graminella nigrifrons, and the Molecular Interactions of Maize fine streak rhabdovirus Transmission. PLOS ONE 7:e40613.

Chen, Y., Xie, D., Li, W. Y., Cheung, C. M., Yao, H., Chan, C. Y., Chan, C.-y., Xu, F.-P., Liu, Y.-H., and Sung, J. J. 2010. RNAi targeting EZH2 inhibits tumor growth and liver metastasis of pancreatic cancer in vivo. Cancer letters 297:109-116.

Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y., and Hwang, C.-C. 2013. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. PLOS ONE 8:e62856.

Childers, A. 2019. Sequenced Arthropod Genomes. i5K.

Chou, H.-H., and Holmes, M. H. 2001. DNA sequence quality trimming and vector removal. Bioinformatics 17:1093-1104.

Chovatia, M., and Sharma, A. 2014. Genomic DNA Sample QC. Joint Genome Institute.

Christensen, N. M., Axelsen, K. B., Nicolaisen, M., and Schulz, A. 2005. Phytoplasmas and their interactions with hosts. Trends in plant science 10:526-535.

Cisak, E., Wójcik-Fatla, A., Zając, V., Sawczyn, A., Sroka, J., and Dutkiewicz, J. 2015. Spiroplasma – an emerging arthropod-borne pathogen? Annals of Agricultural and Environmental Medicine 22:589-593.

da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L., Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., and Pereira, R. J. 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Marine Genomics 30:3-13.

Dakhil, H. A., Hammad, E. A.-F., El-Mohtar, C., and Abou-Jawdah, Y. 2011. Survey of Leafhopper Species in Almond Orchards Infected with Almond Witches'-Broom Phytoplasma in Lebanon. Journal of Insect Science 11:60.

de Oliveira, C. M., Lopes, J. R. S., Camargo, L. E. A., Fungaro, M. H. P., and Nault, L. R. 2007. Genetic diversity in populations of Dalbulus maidis (DeLong and Wolcott)(Hemiptera: Cicadellidae) from distant localities in Brazil assessed by RAPD-PCR markers. Environmental entomology 36:204-212.

de Oliveira, C. M., Lopes, J. R. S., Dias, C. T. D. S., and Nault, L. R. 2004. Influence of latitude and elevation on polymorphism among populations of the corn leafhopper, Dalbulus maidis (DeLong and Wolcott)(Hemiptera: Cicadellidae), in Brazil. Environmental entomology 33:1192-1199.

Del Angel, V. D., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Pettersson, O. V., Amselem, J., Bouri, L., Bocs, S., and Klopp, C. 2018. Ten steps to get started in Genome Assembly and Annotation. F1000Research 7.

DeLong, D. M., and Hershberger, R. V. 1947. The genus Exitianus in North America including Mexico (Homoptera: Cicadellidae).

Denton, J. F., Lugo-Martinez, J., Tucker, A. E., Schrider, D. R., Warren, W. C., and Hahn, M. W. 2014. Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. PLOS Computational Biology 10:e1003998.

Desai, A., Marwah, V. S., Yadav, A., Jha, V., Dhaygude, K., Bangar, U., Kulkarni, V., and Jere, A. 2013. Identification of optimum sequencing depth especially for de novo genome assembly of small genomes using next generation sequencing data. PLoS One 8:e60204.

Dietrich, C. H. 1999. The role of grasslands in the diversification of leafhoppers (Homoptera: Cicadellidae): a phylogenetic perspective. Pages 44-49 in: Proceedings of the Fifteenth North American Prairie Conference Natural Areas Assoc Bend, OR.

Dietrich, C. H. 2005. Keys to the families of Cicadomorpha and subfamilies and tribes of Cicadellidae (Hemiptera: Auchenorrhyncha). Florida Entomologist:502-517.

Dietrich, C. H. 2006. Guide to the Subfamilies of Leafhoppers (Cicadellidae).

Dietrich, C. H., and Dmitriev, D. A. 2003. Reassessment of the Leafhopper Tribes Koebeliini Baker and Grypotini Haupt (Hemiptera: Cicadellidae). Annals of the Entomological Society of America 96:766-775.

Dietrich, C. H., and Rakitov, R. A. 2002. Some Remarkable New Deltocephaline Leafhoppers (Hemiptera: Cicadellidae: Deltocephalinae) from the Amazonian Rainforest Canopy. Journal of the New York Entomological Society 110:1-48.

Dietrich, C. H., Rakitov, R. A., Holmes, J. L., and Black, W. C. 2001. Phylogeny of the Major Lineages of Membracoidea (Insecta: Hemiptera: Cicadomorpha) Based on 28S rDNA Sequences. Molecular Phylogenetics and Evolution 18:293-305.

Dietrich, C. H., Whitcomb, R. F., and Black, W. C. 1997. Phylogeny of the Grassland Leafhopper GenusFlexamia(Homoptera: Cicadellidae) Based on Mitochondrial DNA Sequences. Molecular Phylogenetics and Evolution 8:139-149.

Dietzgen, R. G., Mann, K. S., and Johnson, K. N. 2016. Plant virus–insect vector interactions: Current and potential future research directions. Viruses 8:303.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., and Dickson, B. J. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448:151-156.

Dittmer, N. T., Suderman, R. J., Jiang, H., Zhu, Y.-C., Gorman, M. J., Kramer, K. J., and Kanost, M. R. 2004. Characterization of cDNAs encoding putative laccase-like multicopper oxidases and developmental expression in the tobacco hornworm, Manduca sexta, and the malaria mosquito, Anopheles gambiae. Insect biochemistry and molecular biology 34:29-41.

Dominguez del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J.-F., Vlasova, A., Leskosek, B. L., Soler, L., Binzer-Panchal, M., and Lantz, H. 2018. Ten steps to get started in Genome Assembly and Annotation. F1000Research 7:ELIXIR-148.

Dong, Y., and Friedrich, M. 2005. Nymphal RNAi: systemic RNAi mediated gene knockdown in juvenile grasshopper. BMC biotechnology 5:25.

Donoghue, M. J., and Cantino, P. D. 1988. Paraphyly, ancestors, and the goals of taxonomy: a botanical defense of cladism. The Botanical Review 54:107-128.

Douglas, A. E. 2006. Phloem-sap feeding by animals: problems and solutions. Journal of Experimental Botany 57:747-754.

Dowling, D., Pauli, T., Donath, A., Meusemann, K., Podsiadlowski, L., Petersen, M., Peters, R. S., Mayer, C., Liu, S., Zhou, X., Misof, B., and Niehuis, O. 2016. Phylogenetic Origin and Diversification of RNAi Pathway Genes in Insects. Genome Biology and Evolution 8:3784-3793.

Du, Y., Dai, W., and Dietrich, C. H. 2017a. Mitochondrial genomic variation and phylogenetic relationships of three groups in the genus Scaphoideus (Hemiptera: Cicadellidae: Deltocephalinae). Scientific reports 7:16908.

Du, Y., Zhang, C., Dietrich, C. H., Zhang, Y., and Dai, W. 2017b. Characterization of the complete mitochondrial genomes of Maiestas dorsalis and Japananus hyalinus (Hemiptera: Cicadellidae) and comparison with other Membracoidea. Scientific Reports 7:14197.

Eddy, S. R. 2011. Accelerated profile HMM searches. PLoS computational biology 7:e1002195.

Edwards, R. 2016. Percent similarity at different taxonomic levels.

Ekblom, R., and Wolf, J. B. W. 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl 7:1026-1042.

Ellegren, H. 2014. Genome sequencing and population genomics in non-model organisms. Trends in ecology & evolution 29:51-63.

Fijen, T. P., Scheper, J. A., Boom, T. M., Janssen, N., Raemakers, I., and Kleijn, D. 2018. Insect pollination is at least as important for marketable crop yield as plant quality in a seed crop. Ecology letters 21:1704-1713.

Filipowicz, W. 2005. RNAi: the nuts and bolts of the RISC machine. Cell 122:17-20.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research 44:D279-D285.

Finnegan, D. J. 1989. Eukaryotic transposable elements and genome evolution. Trends in genetics 5:103-107.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. nature 391:806-811.

Fletcher, J., Wayadande, A., Melcher, U., and Ye, F. 1998. The phytopathogenic mollicute-insect vector interface: a closer look. Phytopathology 88:1351-1358.

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. 2011. Considering Transposable Element Diversification in De Novo Annotation Approaches. PLOS ONE 6:e16526.

Frost, K. E., Willis, D. K., and Groves, R. L. 2011. Detection and Variability of Aster Yellows Phytoplasma Titer in Its Insect Vector, Macrosteles quadrilineatus (Hemiptera: Cicadellidae). Journal of Economic Entomology 104:1800-1815.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-3152.

Gahm, N. 2017. Ecology of Grassland Leafhoppers (Hemiptera: Cicadellidae) with Emphasis on the Genus Balclutha.

Galetto, L., Bosco, D., Balestrini, R., Genre, A., Fletcher, J., and Marzachì, C. 2011. The major antigenic membrane protein of "Candidatus Phytoplasma asteris" selectively interacts with ATP synthase and actin of leafhopper vectors. PLoS One 6:e22571.

Gálvez, G. E. 1968. Purification and characterization of rice tungro virus by analytical density-gradient centrifugation. Virology 35:418-426.

Gámez-Lobo, R. 1969. A new leafhopper-borne virus of corn in Central America. Un nuevo virus del maíz transmitido por saltahojas en Centroamérica. Plant Disease Reporter. 53:929-932.

Ghosh, S., Jassar, O., Kontsedalov, S., Lebedev, G., Wang, C., Turner, D., Levy, A., and Ghanim, M. 2019. A Transcriptomics Approach Reveals Putative Interaction of Candidatus Liberibacter Solanacearum with the Endoplasmic Reticulum of Its Psyllid Vector. Insects 10:279.

Gillham, M. C., and Claridge, M. F. 1994. A multivariate approach to host plant associated morphological variation in the polyphagous leafhopper, Alnetoidia alneti (Dahlbom). Biological Journal of the Linnean Society 53:127-151.

Gong, Y.-H., Yu, X.-R., Shang, Q.-L., Shi, X.-y., and Gao, X.-W. 2014. Oral Delivery Mediated RNA Interference of a Carboxylesterase Gene Results in Reduced Resistance to Organophosphorus Insecticides in the Cotton Aphid, Aphis gossypii Glover. PLOS ONE 9:e102823.

González, V. R., and Gámez-Lobo, R. 1974. Algunos factores que afectan la transmisión del virus del rayado fino del maiz por Dalbulus maidis DeLong & Wolcott. Some factors that affect the transmission of the rayado fino virus of maize by Dalbulus maidis DeLong & Wolcott. Turrialba. 24:51-57.

Goodwin, S., McPherson, J. D., and McCombie, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17:333.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology 29:644-652.

Gray, S. M., and Banerjee, N. 1999. Mechanisms of arthropod transmission of plant and animal viruses. Microbiology and Molecular Biology Reviews 63:128-148.

Gray, S. M., Caillaud, M. C., Burrows, M., and Smith, D. M. 2007. Transmission of two viruses that cause Barley Yellow Dwarf is controlled by different loci in the aphid, Schizaphis graminum. Journal of Insect Science 7.

Grohmann, L., Blenau, W., Erber, J., Ebert, P. R., Strünker, T., and Baumann, A. 2003. Molecular and functional characterization of an octopamine receptor from honeybee (Apis mellifera) brain. Journal of Neurochemistry 86:725-735.

Gu, L., and Knipple, D. C. 2013. Recent advances in RNA interference research in insects: Implications for future insect pest management strategies. Crop Protection 45:36-40.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

Gustin, R. D., and Stoner, W. N. 1973. Life History of Exitianus exitiosus (Homoptera: Cicadellidae) in the Laboratory1. Annals of the Entomological Society of America 66:388-389.

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., and Town, C. D. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic acids research 31:5654-5666.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., and Lieber, M. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8:1494.

Haas, B. J., Zeng, Q., Pearson, M. D., Cuomo, C. A., and Wortman, J. R. 2011. Approaches to fungal genome annotation. Mycology 2:118-141.

Hamilton, K. A. 1995. Evaluation of leafhoppers and their relatives (Insecta: Homoptera: Auchenorrhyncha) as indicators of prairie preserve quality. Pages 211-226 in: Proceedings of the 14th North American Prairie Conference. Preney Print and Litho Inc. Windsor.

Hamilton, K. A., and Whitcomb, R. F. 2010. Leafhoppers (Homoptera: Cicadellidae): a major family adapted to grassland habitats. Arthropods of Canadian grasslands 1:169-197.

Hammond, R., and Ramirez, P. 2001. Molecular characterization of the genome of Maize rayado fino virus, the type member of the genus Marafivirus. Virology 282:338-347.

Hammond, S. M. 2005. Dicing and slicing. FEBS letters 579:5822-5829.

Hanboonsong, Y., Ritthison, W., Choosai, C., and Sirithorn, P. 2006. Transmission of sugarcane white leaf phytoplasma by Yamatotettix flavovittatus, a new leafhopper vector. Journal of Economic Entomology 99:1531-1537.

Hanrahan, S. J., and Johnston, J. S. 2011. New genome size estimates of 134 species of arthropods. Chromosome Research 19:809.

Hardison, R. C. 2003. Comparative genomics. PLoS biology 1:e58.

Hardy, J. L., Houk, E. J., Kramer, L. D., and Reeves, W. C. 1983. Intrinsic factors affecting vector competence of mosquitoes for arboviruses. Annual review of entomology 28:229-262.

Harvey, B., and Walker, J. 1999. Studies with plant laccases: I. Comparison of plant and fungal laccases. Biochem. Mol. Biol. Biophys 3:45-51.

Hattori, M., Konishi, H., Tamura, Y., Konno, K., and Sogawa, K. 2005. Laccase-type phenoloxidase in salivary glands and watery saliva of the green rice leafhopper, Nephotettix cincticeps. Journal of Insect Physiology 51:1359-1365.

Hattori, M., Tsuchihara, K., Noda, H., Konishi, H., Tamura, Y., Shinoda, T., Nakamura, M., and Hasegawa, T. 2010. Molecular characterization and expression of laccase genes in the salivary glands of the green rice leafhopper, Nephotettix cincticeps (Hemiptera: Cicadellidae). Insect Biochemistry and Molecular Biology 40:331-338.

Hibino, H., Saleh, N., and Roechan, M. 1979. Transmission of two kinds of rice tungro-associated viruses by insect vectors. Phytopathology 69:1266-1268.

Hoff, K. J., and Stanke, M. 2015. Current methods for automated annotation of protein-coding genes. Current Opinion in insect science 7:8-14.

Hogenhout, S. A., Ammar, E.-D., Whitfield, A. E., and Redinbaugh, M. G. 2008a. Insect vector interactions with persistently transmitted viruses. Annu. Rev. Phytopathol. 46:327-359.

Hogenhout, S. A., Oshima, K., AMMAR, E. D., Kakizawa, S., KINGDOM, H. N., and Namba, S. 2008b. Phytoplasmas: bacteria that manipulate plants and insects. Molecular Plant Pathology 9:403-423.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. C., and Wides, R. 2002. The genome sequence of the malaria mosquito Anopheles gambiae. science 298:129-149.

Horn, T., and Boutros, M. 2010. E-RNAi: a web application for the multi-species design of RNAi reagents—2010 update. Nucleic acids research 38:W332-W339.

Houé, V., Bonizzoni, M., and Failloux, A.-B. 2019. Endogenous non-retroviral elements in genomes of Aedes mosquitoes and vector competence. Emerging microbes & infections 8:542-555.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C., and Bork, P. 2015. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research 44:D286-D293.

Hull, R. 2013. Plant virology. Academic press.

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., and Otto, T. D. 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biology 14:R47.

i5K Consortium. 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. Journal of Heredity 104:595-600.

Illumina. 2010. Illumina Sequencing Technology.

Jaubert-Possamai, S., Le Trionnaire, G., Bonhomme, J., Christophides, G. K., Rispe, C., and Tagu, D. 2007. Gene knockdown by RNAi in the pea aphid Acyrthosiphon pisum. BMC Biotechnology 7:63.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and genome research 110:462-467.

Kajikawa, M., and Okada, N. 2002. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 111:433-444.

Kanakala, S., and Ghanim, M. 2016. RNA Interference in Insect Vectors for Plant Viruses. Viruses 8:329.

Kanehisa, M., and Goto, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28:27-30.

Kapitonov, V. V., and Jurka, J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. Nature Reviews Genetics 9:411.

Kaur, N., Hasegawa, D. K., Ling, K.-S., and Wintermantel, W. M. 2016. Application of genomics for understanding plant virus-insect vector interactions and insect vector control. Phytopathology 106:1213-1222.

Kennerdell, J. R., and Carthew, R. W. 1998. Use of dsRNA-Mediated Genetic Interference to Demonstrate that frizzled and frizzled 2 Act in the Wingless Pathway. Cell 95:1017-1026.

Kirkpatrick, B. C., and Smart, C. D. 1995. Phytoplasmas: Can Phylogeny Provide the Means to Understand Pathogenicity? Pages 187-212 in: Advances in Botanical Research, vol. 21. J. A. Callow, J. H. Andrews and I. C. Tommerup, eds. Academic Press.

Knight, W. 1987. Leafhoppers of the grass-feeding genus Balclutha (Homoptera, Cicadellidae) in the Pacific region. Journal of Natural History 21:1173-1224.

Kobiałka, M., Michalik, A., Szwedo, J., and Szklarzewicz, T. 2018a. Diversity of symbiotic microbiota in Deltocephalinae leafhoppers (Insecta, Hemiptera, Cicadellidae). Arthropod structure & development 47:268-278.

Kobiałka, M., Michalik, A., Walczak, M., and Szklarzewicz, T. 2018b. Dual “bacterial-fungal” symbiosis in deltocephalinae leafhoppers (Insecta, Hemiptera, Cicadomorpha: Cicadellidae). Microbial ecology 75:771-782.

Kobiałka, M., Michalik, A., Walczak, M., Junkiert, Ł., and Szklarzewicz, T. 2016. Sulcia symbiont of the leafhopper Macrosteles laevis (Ribaut, 1927) (Insecta, Hemiptera, Cicadellidae: Deltocephalinae) harbors Arsenophonus bacteria. Protoplasma 253:903-912.

Koressaar, T., and Remm, M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289-1291.

Koressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R., and Remm, M. 2018. Primer3_masker: integrating masking of template sequence with primer design software. Bioinformatics 34:1937-1938.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567-580.

Kumar, A., Congiu, L., Lindström, L., Piiroinen, S., Vidotto, M., and Grapputo, A. 2014. Sequencing, De Novo Assembly and Annotation of the Colorado Potato Beetle, Leptinotarsa decemlineata, Transcriptome. PLOS ONE 9:e86012.

Labroussaa, F., Arricau-Bouvery, N., Dubrana, M.-P., and Saillard, C. 2010. Entry of Spiroplasma citri into Circulifer haematoceps cells involves interaction between spiroplasma phosphoglycerate kinase and leafhopper actin. Appl. Environ. Microbiol. 76:1879-1886.

Labroussaa, F., Dubrana, M.-P., Arricau-Bouvery, N., Béven, L., and Saillard, C. 2011. Involvement of a Minimal Actin-Binding Region of Spiroplasma citri Phosphoglycerate Kinase in Spiroplasma Transmission by Its Leafhopper Vector. PLOS ONE 6:e17357.

Labroussaa, F., Ionescu, M., Zeilinger, A. R., Lindow, S. E., and Almeida, R. P. 2017. A chitinase is required for Xylella fastidiosa colonization of its insect and plant hosts.

Laetsch, D., and Blaxter, M. 2017. BlobTools: Interrogation of genome assemblies [version 1; referees: 2 approved with reservations]. F1000Research 6.

Lagesen, K., Hallin, P., Rødland, E., Stærfeldt, H., Rognes, T., and Ussery, D. 2007. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res 35:3100-3108.

Lawson, D., Arensburger, P., Atkinson, P., Besansky, N. J., Bruggner, R. V., Butler, R., Campbell, K. S., Christophides, G. K., Christley, S., and Dialynas, E. 2008. VectorBase: a data resource for invertebrate vector genomics. Nucleic acids research 37:D583-D587.

Lee, I.-M., Davis, R. E., and Gundersen-Rindal, D. E. 2000. Phytoplasma: phytopathogenic mollicutes. Annual Reviews in Microbiology 54:221-255.

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., and Gilbert, M. T. P. 2018. Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences 115:4325-4333.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674-1676.

Li, J., Wang, X. P., Wang, M. Q., Ma, W. H., and Hua, H. X. 2013. Advances in the use of the RNA interference technique in Hemiptera. Insect science 20:31-39.

Li, W., and Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.

Liu, H.-Y., Gumpf, D., Oldfield, G., and Calavan, E. 1983. The relationship of Spiroplasma citri and Circulifer tenellus. Phytopathology 73:585-590.

Liu, J., Xu, Y., Zhang, L., Li, W., Cai, Z., Li, F., Peng, M., Li, F., and Hu, B. 2017. De novo assembly and analysis of the transcriptome of Rumex patientia L. during cold stress. PloS one 12:e0186470.

Liu, S., He, X., Park, G., Josefsson, C., and Perry, K. L. 2002. A Conserved Capsid Protein Surface Domain of Cucumber Mosaic Virus Is Essential for Efficient Aphid Vector Transmission. Journal of Virology 76:9756-9762.

Livak, K. J., and Schmittgen, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. methods 25:402-408.

Love, R. R., Weisenfeld, N. I., Jaffe, D. B., Besansky, N. J., and Neafsey, D. E. 2016. Evaluation of DISCOVAR de novo using a mosquito sample for cost-effective short-read genome assembly. BMC genomics 17:187.

Lü, J., Yang, C., Zhang, Y., and Pan, H. 2018. Selection of Reference Genes for the Normalization of RT-qPCR Data in Gene Expression Studies in Insects: A Systematic Review. Frontiers in Physiology 9.

Łukasik, P., van Asch, M., Guo, H., Ferrari, J., and Charles J. Godfray, H. 2013. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. Ecology letters 16:214-218.

Ma, K.-S., Li, F., Liang, P.-Z., Chen, X.-W., Liu, Y., and Gao, X.-W. 2016. Identification and Validation of Reference Genes for the Normalization of Gene Expression Data in qRT-PCR Analysis in Aphis gossypii (Hemiptera: Aphididae). Journal of Insect Science 16.

Ma, S., Zhang, S., Wang, F., Liu, Y., Liu, Y., Xu, H., Liu, C., Lin, Y., Zhao, P., and Xia, Q. 2012. Highly Efficient and Specific Genome Editing in Silkworm Using Custom TALENs. PLOS ONE 7:e45035.

Maggi, F., Galetto, L., Marzachì, C., and Bosco, D. 2014. Temperature-dependent transmission of Candidatus phytoplasma asteris by the vector leafhopper Macrosteles quadripunctulatus Kirschbaum. 2014 2.

Mäkinen, V., Salmela, L., and Ylinen, J. 2012. Normalized N50 assembly metric using gap-restricted co-linear chaining. BMC Bioinformatics 13:255.

Manoj, S. 2007. Nucleotide distribution patterns in insect genomes. arXiv preprint q-bio/0702036.

Mao, M., Yang, X., and Bennett, G. 2017. The complete mitochondrial genome of Macrosteles quadrilineatus (Hemiptera: Cicadellidae). Mitochondrial DNA Part B 2:173-175.

Maramorosch, K. 2014. Leafhopper and Planthopper rearing. in: Rearing Animal and Plant Pathogen Vectors. K. Maramorosch and F. Mahmood, eds. CRC Press.

Marçais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764-770.

Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. Trends in genetics 24:133-141.

Markham, P. 1983. Spiroplasmas in leafhoppers: a review. The Yale journal of biology and medicine 56:745.

Markham, P. G. 1982. Insect Vectors. Pages 307-326 in: Plant and Insect Mycoplasma Techniques. M. J. Daniels and P. G. Markham, eds. Springer Netherlands, Dordrecht.

Maroniche, G. A., Sagadín, M., Mongelli, V. C., Truol, G. A., and del Vas, M. 2011. Reference gene selection for gene expression studies using RT-qPCR in virus-infected planthoppers. Virology Journal 8:308.

Matsukawa, M., Tasaki, M., Doi, K., Ito, K., Kawakita, K., and Tanaka, T. 2018. Regional population differences of the brown planthopper (Nilaparvata lugens Stål) in Cambodia using genotyping-by-sequencing. Bulletin of entomological research 108:471-478.

Matsumoto, Y., and Hattori, M. 2016. Gene silencing by parental RNA interference in the Green Rice Leafhopper, Nephotettix cincticeps (Hemiptera: Cicadellidae). Archives of insect biochemistry and physiology 91:152-164.

Mayer, A. M., and Staples, R. C. 2002. Laccase: new functions for an old enzyme. Phytochemistry 60:551-565.

Megy, K., Emrich, S. J., Lawson, D., Campbell, D., Dialynas, E., Hughes, D. S., Koscielny, G., Louis, C., MacCallum, R. M., and Redmond, S. N. 2011. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic acids research 40:D729-D734.

Miller, J. R., Koren, S., and Sutton, G. 2010. Assembly algorithms for next-generation sequencing data. Genomics 95:315-327.

Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., and Prochnik, S. E. 2002. Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. Genome biology 3:research0083. 0081.

Mitchell, P. L. 2004. Heteroptera as vectors of plant pathogens. Neotropical Entomology 33:519-545.

Montllor, C. B., Maxmen, A., and Purcell, A. H. 2002. Facultative bacterial endosymbionts benefit pea aphids Acyrthosiphon pisum under heat stress. Ecological Entomology 27:189-195.

Moran, N. A., Tran, P., and Gerardo, N. M. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. Applied and Environmental Microbiology 71:8802-8810.

Moreno-Santillán, D. D., Machain-Williams, C., Hernández-Montes, G., and Ortega, J. 2019. De Novo transcriptome Assembly and Functional Annotation in Five species of Bats. Scientific reports 9:6222.

Morgan, A. R., Smith-Herron, A. J., and Cook, J. L. 2013. Rapid spread of Balclutha rubrostriata (Hemiptera: Cicadellidae) in Texas and southwestern Louisiana, USA with notes on its associated host plants. Florida Entomologist 96:477-482.

Moya-Raygoza, G., Palomera-Avalos, V., and Galaviz-Mejia, C. 2007. Field overwintering biology of Spiroplasma kunkelii (Mycoplasmatales: Spiroplasmataceae) and its vector Dalbulus maidis (Hemiptera: Cicadellidae). Annals of applied biology 151:373-379.

Munoz-Torres, M. C., Reese, J. T., Childers, C. P., Bennett, A. K., Sundaram, J. P., Childs, K. L., Anzola, J. M., Milshina, N., and Elsik, C. G. 2010. Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. Nucleic acids research 39:D658-D662.

Munyaneza, J. E., Crosslin, J. M., and Lee, I. 2007. Phytoplasma diseases and insect vectors in potatoes of the Pacific Northwest of the United States. Bulletin of Insectology 60:181.

Munyaneza, J. E., Crosslin, J. M., Upton, J. E., and Buchman, J. L. 2010. Incidence of the beet leafhopper-transmitted virescence agent phytoplasma in local populations of the beet leaf hopper, Circulifer tenellus, in Washington State. Journal of insect science 10:18.

Murral, D. J., Nault, L. R., Hoy, C. W., Madden, L. V., and Miller, S. A. 1996. Effects of temperature and vector age on transmission of two Ohio strains of aster yellows phytoplasma by the aster leafhopper (Homoptera: Cicadellidae). Journal of Economic Entomology 89:1223-1232.

Nagarajan, N., and Pop, M. 2009. Parametric complexity of sequence assembly: theory and applications to next generation sequencing. Journal of computational biology 16:897-908.

Nagarajan, N., and Pop, M. 2013. Sequence assembly demystified. Nature Reviews Genetics 14:157.

Nagata, T., Inoue-Nagata, A. K., Van Lent, J., Goldbach, R., and Peters, D. 2002. Factors determining vector competence and specificity for transmission of Tomato spotted wilt virus. Journal of General Virology 83:663-671.

Nakasuji, F., and Kiritani, K. 1970. Ill-effects of rice dwarf virus upon its vector, Nephotettix cincticeps Uhler (Hemiptera: Deltocephalidae), and its significance for changes in relative abundance of infected individuals among vector populations. Applied entomology and zoology 5:1-12.

Nandety, R. S., Kuo, Y.-W., Nouri, S., and Falk, B. W. 2015. Emerging strategies for RNA interference (RNAi) applications in insects. Bioengineered 6:8-19.

Nasu, S. 1963. Studies on some leafhoppers and planthoppers which transmit virus diseases of rice plant in Japan. Bull. Kyushu Agric. Exp. Stn 8:153-349.

Nault, L. 1997. Arthropod transmission of plant viruses: a new synthesis. Annals of the entomological Society of America 90:521-541.

Nault, L. R. 1980. Maize bushy stunt and corn stunt: a comparison of disease symptoms, pathogen host ranges, and vectors. Phytopathology 70:659-662.

Nault, L. R. 1990. Evolution of an insect pest: maize and the corn leafhopper, a case study. Maydica 35:165-175.

Nault, L. R. 1998. Dalbulus maidis identification, biology, ecology and pest status. Diagnosing Maize Diseases in Latin America 9.

Nault, L. R., and Ammar, E.-D. 1989. Leafhopper and planthopper transmission of plant viruses. Annual review of entomology 34:503-529.

Nault, L. R., and Madden, L. 1988. Phylogenetic relatedness of maize chlorotic dwarf virus leafhopper vectors. Phytopathology 78:1683-1687.

Nault, L. R., and Rodriguez, J. G. 1985. The leafhoppers and planthoppers.

Nault, L. R., Styer, W., Knoke, J., and Pitre, H. 1973b. Semipersistent transmission of leafhopper-borne maize chlorotic dwarf virus. Journal of Economic Entomology 66:1271-1273.

Nault, L., Styer, W., Knoke, J., and Pitre, H. 1973a. Semipersistent transmission of leafhopper-borne maize chlorotic dwarf virus. Journal of Economic Entomology 66:1271-1273.

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z., Loftus, B., Xi, Z., Megy, K., Grabherr, M., Ren, Q., Zdobnov, E. M., Lobo, N. F., Campbell, K. S., Brown, S. E., Bonaldo, M. F., Zhu, J., Sinkins, S. P., Hogenkamp, D. G., Amedeo, P., Arensburger, P., Atkinson, P. W., Bidwell, S., Biedler, J., Birney, E., Bruggner, R. V., Costas, J., Coy, M. R., Crabtree, J., Crawford, M., deBruyn, B., DeCaprio, D., Eiglmeier, K., Eisenstadt, E., El-Dorry, H., Gelbart, W. M., Gomes, S. L., Hammond, M., Hannick, L. I., Hogan, J. R., Holmes, M. H., Jaffe, D., Johnston, J. S., Kennedy, R. C., Koo, H., Kravitz, S., Kriventseva, E. V., Kulp, D., LaButti, K., Lee, E., Li, S., Lovin, D. D., Mao, C., Mauceli, E., Menck, C. F. M., Miller, J. R., Montgomery, P., Mori, A., Nascimento, A. L., Naveira, H. F., Nusbaum, C., O'Leary, S., Orvis, J., Pertea, M., Quesneville, H., Reidenbach, K. R., Rogers, Y.-H., Roth, C. W., Schneider, J. R., Schatz, M., Shumway, M., Stanke, M., Stinson, E. O., Tubio, J. M. C., VanZee, J. P., Verjovski-Almeida, S., Werner, D., White, O., Wyder, S., Zeng, Q., Zhao, Q., Zhao, Y., Hill, C. A., Raikhel, A. S., Soares, M. B., Knudson, D. L., Lee, N. H., Galagan, J., Salzberg, S. L., Paulsen, I. T., Dimopoulos, G., Collins, F. H., Birren, B., Fraser-Liggett, C. M., and Severson, D. W. 2007. Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. Science 316:1718-1723.

Ng, J. C., and Falk, B. W. 2006. Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. Annu. Rev. Phytopathol. 44:183-212.

Nielsen, H. 2017. Predicting Secretory Proteins with SignalP. Pages 59-73 in: Protein Function Prediction: Methods and Protocols. D. Kihara, ed. Springer New York, New York, NY.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein engineering 10:1-6.

Nielson, M. W. 1968. The leafhopper vectors of phytopathogenic viruses (Homoptera, Cicadellidae): taxonomy, biology, and virus transmission. US Department of Agriculture.

Nielson, M. W. 1979. Taxonomic Relationships of Leafhopper Vectors of Plant Pathogens. in: Leafhopper vectors and plant disease agents. K. Maramorosch and K. Harris, eds. Academic Press, New York.

Oliver, K. M., Russell, J. A., Moran, N. A., and Hunter, M. S. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proceedings of the National Academy of Sciences 100:1803-1807.

Olivier, C., Séguin-Swartz, G., Galka, B., and Olfert, O. 2011. Aster yellows in leafhoppers and field crops in Saskatchewan, Canada, 2001–2008. Am J Plant Sci Biotechnol 141:425-462.

Papanicolaou, A., Schetelig, M. F., Arensburger, P., Atkinson, P. W., Benoit, J. B., Bourtzis, K., Castañera, P., Cavanaugh, J. P., Chao, H., Childers, C., Curril, I., Dinh, H., Doddapaneni, H., Dolan, A., Dugan, S., Friedrich, M., Gasperi, G., Geib, S., Georgakilas, G., Gibbs, R. A., Giers, S. D., Gomulski, L. M., González-Guzmán, M., Guillem-Amat, A., Han, Y., Hatzigeorgiou, A. G., Hernández-Crespo, P., Hughes, D. S. T., Jones, J. W., Karagkouni, D., Koskinioti, P., Lee, S. L., Malacrida, A. R., Manni, M., Mathiopoulos, K., Meccariello, A., Murali, S. C., Murphy, T. D., Muzny, D. M., Oberhofer, G., Ortego, F., Paraskevopoulou, M. D., Poelchau, M., Qu, J., Reczko, M., Robertson, H. M., Rosendale, A. J., Rosselot, A. E., Saccone, G., Salvemini, M., Savini, G., Schreiner, P., Scolari, F., Siciliano, P., Sim, S. B., Tsiamis, G., Ureña, E., Vlachos, I. S., Werren, J. H., Wimmer, E. A., Worley, K. C., Zacharopoulou, A., Richards, S., and Handler, A. M. 2016. The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. Genome Biology 17:192.

Pereira-Marques, J., Hout, A., Ferreira, R. M., Weber, M., Pinto-Ribeiro, I., van Doorn, L.-J., Knetsch, C. W., and Figueiredo, C. 2019. Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Frontiers in Microbiology 10.

Perilla-Henao, L. M., and Casteel, C. L. 2016. Vector-borne bacterial plant pathogens: interactions with hemipteran insects and plants. Frontiers in Plant Science 7:1163.

Pfaffl, M. W., Tichopad, A., Prgomet, C., and Neuvians, T. P. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotechnology letters 26:509-515.

Picardi, E., and Pesole, G. 2010. Computational methods for ab initio and comparative gene finding. Pages 269-284 in: Data mining techniques for the life sciences. Springer.

Pitre, H. 1970. Observations on the life cycle of Dalbulus maidis on three plant species. Florida Entomologist:33-37.

Pitre, H., Douglas, W., Combs Jr, R., and Hepner, L. 1967. Annual movement of Dalbulus maidis into the Southeastern United States and its role as vector of the corn stunt virus. Journal of Economic Entomology 60:616-617.

Portnoy, V., Huang, V., Place, R. F., and Li, L.-C. 2011. Small RNA and transcriptional upregulation. Wiley Interdiscip Rev RNA 2:748-760.

Pucker, B. 2019. Mapping-based genome size estimation. bioRxiv:607390.

Purcell, A. H. 1982. Insect vector relationships with procaryotic plant pathogens. Annual Review of Phytopathology 20:397-417.

Purcell, A. H., and Almeida, R. P. 2005. Insects as vectors of disease agents. Encyclopedia of plant and crop science 10.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., and Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341.

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., and Anxolabehere, D. 2005. Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLOS Computational Biology 1:e22.

Redak, R. A., Purcell, A. H., Lopes, J. R. S., Blua, M. J., Mizell Iii, R. F., and Andersen, P. C. 2003. THE BIOLOGY OF XYLEM FLUID–FEEDING INSECT VECTORS OF XYLELLA FASTIDIOSA AND THEIR RELATION TO DISEASE EPIDEMIOLOGY. Annual Review of Entomology 49:243-270.

Redinbaugh, M., Seifers, D., Meulia, T., Abt, J., Anderson, R., Styer, W., Ackerman, J., Salomon, R., Houghton, W., and Creamer, R. 2002. Maize fine streak virus, a new leafhopper-transmitted rhabdovirus. Phytopathology 92:1167-1174.

Rhoads, A., and Au, K. F. 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics 13:278-289.

Richards, S., and Murali, S. C. 2015. Best practices in insect genome sequencing: what works and what doesn’t. Current opinion in insect science 7:1-7.

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., Gibbs, R., Beeman, R. W., Brown, S. J., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P., Klingler, M., Lorenzen, M., Richards, S., Roth, S., Schröder, R., Tautz, D., Zdobnov, E. M., Muzny, D., Gibbs, R. A., Weinstock, G. M., Attaway, T., Bell, S., Buhay, C. J., Chandrabose, M. N., Chavez, D., Clerk-Blankenburg, K. P., Cree, A., Dao, M., Davis, C., Chacko, J., Dinh, H., Dugan-Rocha, S., Fowler, G., Garner, T. T., Garnes, J., Gnirke, A., Hawes, A., Hernandez, J., Hines, S., Holder, M., Hume, J., Jhangiani, S. N., Joshi, V., Khan, Z. M., Jackson, L., Kovar, C., Kowis, A., Lee, S., Lewis, L. R., Margolis, J., Morgan, M., Nazareth, L. V., Nguyen, N., Okwuonu, G., Parker, D., Richards, S., Ruiz, S.-J., Santibanez, J., Savard, J., Scherer, S. E., Schneider, B., Sodergren, E., Tautz, D., Vattahil, S., Villasana, D., White, C. S., Wright, R., Park, Y., Beeman, R. W., Lord, J., Oppert, B., Lorenzen, M., Brown, S., Wang, L., Savard, J., Tautz, D., Richards, S., Weinstock, G., Gibbs, R. A., Liu, Y., Worley, K., Weinstock, G., Elsik, C. G., Reese, J. T., Elhaik, E., Landan, G., Graur, D., Arensburger, P., Atkinson, P., Beeman, R. W., Beidler, J., Brown, S. J., Demuth, J. P., Drury, D. W., Du, Y.-Z., Fujiwara, H., Lorenzen, M., Maselli, V., Osanai, M., Park, Y., Robertson, H. M., Tu, Z., Wang, J.-j., Wang, S., Richards, S., Song, H., Zhang, L., Sodergren, E., Werner, D., Stanke, M., Morgenstern, B., Solovyev, V., Kosarev, P., Brown, G., Chen, H.-C., Ermolaeva, O., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Maglott, D., Pruitt, K., Sapojnikov, V., Souvorov, A., Mackey, A. J., Waterhouse, R. M., Wyder, S., Zdobnov, E. M., Zdobnov, E. M., Wyder, S., Kriventseva, E. V., Kadowaki, T., Bork, P., Aranda, M., Bao, R., Beermann, A., Berns, N., Bolognesi, R., Bonneton, F., Bopp, D., Brown, S. J., Bucher, G., Butts, T., Chaumot, A., Denell, R. E., Ferrier, D. E. K., Friedrich, M.,

Gordon, C. M., Jindra, M., Klingler, M., Lan, Q., Lattorff, H. M. G., Laudet, V., von Levetsow, C., Liu, Z., Lutz, R., Lynch, J. A., da Fonseca, R. N., Posnien, N., Reuter, R., Roth, S., Savard, J., Schinko, J. B., Schmitt, C., Schoppmeier, M., Schröder, R., Shippy, T. D., Simonnet, F., Marques-Souza, H., Tautz, D., Tomoyasu, Y., Trauner, J., Van der Zee, M., Vervoort, M., Wittkopp, N., Wimmer, E. A., Yang, X., Jones, A. K., Sattelle, D. B., Ebert, P. R., Nelson, D., Scott, J. G., Beeman, R. W., Muthukrishnan, S., Kramer, K. J., Arakane, Y., Beeman, R. W., Zhu, Q., Hogenkamp, D., Dixit, R., Oppert, B., Jiang, H., Zou, Z., Marshall, J., Elpidina, E., Vinokurov, K., Oppert, C., Zou, Z., Evans, J., Lu, Z., Zhao, P., Sumathipala, N., Altincicek, B., Vilcinskas, A., Williams, M., Hultmark, D., Hetru, C., Jiang, H., Grimmelikhuijzen, C. J. P., Hauser, F., Cazzamali, G., Williamson, M., Park, Y., Li, B., Tanaka, Y., Predel, R., Neupert, S., Schachtner, J., Verleyen, P., Raible, F., Bork, P., Friedrich, M., Walden, K. K. O., Robertson, H. M., Angeli, S., Forêt, S., Bucher, G., Schuetz, S., Maleszka, R., Wimmer, E. A., Beeman, R. W., Lorenzen, M., Tomoyasu, Y., Miller, S. C., Grossmann, D., Bucher, G., Tribolium Genome Sequencing, C., Project, l., Principal, i., White, p., Analysis, l., sequence, D. N. A., global analysis, D. N. A. s., sequencing, E. S. T., genome, a., content, G. C., repetitive Dna, t., telomeres, gene, p., consensus gene, s., global gene content, a., Developmental, p., signalling, p., Pest biology, s. M., channels, R. N. l. g. i., oxidative, p., genes, P., chitin, cuticular, p., digestive, p., immunity, neurohormones, Gpcrs, neuropeptide processing, e., opsins, odorant, r., gustatory, r., odorant, b., chemosensory, p., Medea, and systemic, R. 2008. The genome of the model beetle and pest Tribolium castaneum. Nature 452:949-955.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., and Meisel, R. P. 2005. Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome research 15:1-18.

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., Lawson, D., Okamuro, J., Robertson, H. M., and Schneider, D. J. 2011. Creating a buzz about insect genomes. Science 331:1386-1386.

Rotenberg, D., Jacobson, A. L., Schneweis, D. J., and Whitfield, A. E. 2015. Thrips transmission of tospoviruses. Current Opinion in Virology 15:80-89.

Schuster, S. C. 2008. Next-generation sequencing transforms today's biology. Nature methods 5:16-18.

Scott, J. G., Michel, K., Bartholomay, L., Siegfried, B. D., Hunter, W. B., Smagghe, G., Zhu, K. Y., and Douglas, A. E. 2013. Towards the elements of successful insect RNAi. Journal of insect physiology 59:10.1016/j.jinsphys.2013.1008.1014.

Shakeel, M., Rodriguez, A., Tahir, U. B., and Jin, F. 2018. Gene expression studies of reference genes for quantitative real-time PCR: an overview in insects. Biotechnology Letters 40:227-236.

Shih, J. D., and Hunter, C. P. 2011. SID-1 is a dsRNA-selective dsRNA-gated channel. Rna 17:1057-1065.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212.

Simpson, J. T., and Pop, M. 2015. The theory and practice of genome sequence assembly. Annual review of genomics and human genetics 16:153-172.

Singh, S., Gupta, M., Pandher, S., Kaur, G., Rathore, P., and Palli, S. R. 2018. Selection of housekeeping genes and demonstration of RNAi in cotton leafhopper, Amrasca biguttula biguttula (Ishida). PloS one 13:e0191116.

Smit, A., and Hubley, R. 2015. RepeatModeler Open-1.0.

Smit, A., Hubley, R., and Green, P. 2015a. RepeatMasker Open-4.0.

Smit, A., Hubley, R., and Green, P. 2015b. RepeatMasker Open-4.0.

Sohn, J.-i., and Nam, J.-W. 2016. The present and future of de novo whole-genome assembly. Briefings in bioinformatics 19:23-40.

Sparks, M. E., Shelby, K. S., Kuhar, D., and Gundersen-Rindal, D. E. 2014. Transcriptome of the Invasive Brown Marmorated Stink Bug, Halyomorpha halys (Stål) (Heteroptera: Pentatomidae). PLOS ONE 9:e111646.

Sugumaran, M., Giglio, L. B., Kundzicz, H., Saul, S., and Semensi, V. 1992. Studies on the enzymes involved in puparial cuticle sclerotization in Drosophila melanogaster. Archives of insect biochemistry and physiology 19:271-283.

Summers, C., Newton Jr, A., and Opgenorth, D. 2004. Overwintering of corn leafhopper, Dalbulus maidis (Homoptera: Cicadellidae), and Spiroplasma kunkelii (Mycoplasmatales: Spiroplasmataceae) in California's San Joaquin Valley. Environmental Entomology 33:1644-1651.

Suzuki, S., Oshima, K., Kakizawa, S., Arashida, R., Jung, H.-Y., Yamaji, Y., Nishigawa, H., Ugaki, M., and Namba, S. 2006. Interaction between the membrane protein of a pathogen and insect microfilament complex determines insect-vector specificity. Proceedings of the National Academy of Sciences 103:4252-4257.

Tabachnick, W. J. 1994. Genetics of insect vector competence for arboviruses. Pages 93-108 in: Advances in disease vector research. Springer.

Takiya, D. M., Tran, P. L., Dietrich, C. H., and Moran, N. A. 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. Molecular Ecology 15:4175-4191.

Terenius, O., Papanicolaou, A., Garbutt, J. S., Eleftherianos, I., Huvenne, H., Kanginakudru, S., Albrechtsen, M., An, C., Aymeric, J.-L., and Barthel, A. 2011. RNA interference in Lepidoptera: an overview of successful and unsuccessful studies and implications for experimental design. Journal of insect physiology 57:231-245.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18:1979-1990.

The Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic acids research 32:D258-D261.

The Gene Ontology Consortium. 2018. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research 47:D330-D338.

The UniProt Consortium. 2018. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47:D506-D515.

Tian, C., Tay, W. T., Feng, H., Wang, Y., Hu, Y., and Li, G. 2015. Characterization of Adelphocoris suturalis (Hemiptera: Miridae) transcriptome from different developmental stages. Scientific reports 5:11042.

Tomoyasu, Y., Miller, S. C., Tomita, S., Schoppmeier, M., Grossmann, D., and Bucher, G. 2008. Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in Tribolium. Genome biology 9:R10.

Trivellone, V., Mitrović, M., Dietrich, C. H., and Toševski, I. 2017. Osbornellus auronitens (Hemiptera: Cicadellidae: Deltocephalinae), an introduced species new for the Palaearctic region. The Canadian Entomologist:1-9.

Tsai, J. H. 2008. Corn Leafhopper, Dalbulus maidis (Delong and Wolcott) (Hemiptera: Cicadellidae). Pages 1072-1074 in: Encyclopedia of Entomology. J. L. Capinera, ed. Springer Netherlands, Dordrecht.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. 2012. Primer3--new capabilities and interfaces. Nucleic acids research 40:e115-e115.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology 3:research0034. 0031.

Virla, E. G., Díaz, C., Carpane, P., Laguna, I. G., Ramallo, J., Geronimo Gomez, L., and Giménez Pecci, M. 2004. Estimación preliminar de la disminución en la producción de maíz causada por el Corn Stunt Spiroplasma (CSS) en Tucumán, Argentina.

Vogels, C. B., Göertz, G. P., Pijlman, G. P., and Koenraadt, C. J. 2017. Vector competence of European mosquitoes for West Nile virus. Emerging microbes & infections 6:1-13.

Wang, H., Wu, K., Liu, Y., Wu, Y., and Wang, X. 2015. Integrative proteomics to understand the transmission mechanism of Barley yellow dwarf virus-GPV by its insect vector Rhopalosiphum padi. Scientific reports 5:10971.

Wang, J.-J., and Xing, J.-C. 2019. Complete mitochondrial genome of Abrus expansivus (Hemiptera: Cicadellidae: Deltocephalinae) from China. Mitochondrial DNA Part B 4:197-198.

Wang, S., Lorenzen, M. D., Beeman, R. W., and Brown, S. J. 2008. Analysis of repetitive DNA distribution patterns in the Tribolium castaneum genome. Genome biology 9:R61.

Waterhouse, R. M. 2015. A maturing understanding of the composition of the insect gene repertoire. Current Opinion in Insect Science 7:15-23.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., and Zdobnov, E. M. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular biology and evolution 35:543-548.

Watmough, N. J., and Frerman, F. E. 2010. The electron transfer flavoprotein: Ubiquinone oxidoreductases. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1797:1910-1916.

Wayadande, A. C. 1994. Electronic monitoring of leafhoppers and planthoppers: feeding behavior and application in host-plant resistance studies. Pages 86-105 in: Principles and Applications of Electronic Monitoring and Other Techniques in the Study of Homopteran Feeding Behavio, G. P. Walker and E. Backus, eds. Thomas Say.

Wayadande, A. C., and Fletcher, J. 1995. Transmission of Spiroplasma citri lines and their ability to cross gut and salivary gland barriers within the leafhopper vector Circulifer tenellus. Phytopathology 85:1256-1259.

Wayadande, A. C., and Nault, L. R. 1993. Leafhopper probing behavior associated with maize chlorotic dwarf virus transmission to maize. Phytopathology 83:522-526.

Wayadande, A. C., Baker, G. R., and Fletcher, J. 1997. Comparative ultrastructure of the salivary glands of two phytopathogen vectors, the beet leafhopper, Circulifer tenellus (Baker), and the corn leafhopper, Dalbulus maidis DeLong and Wolcott (Homoptera : Cicadellidae). International Journal of Insect Morphology and Embryology 26:113-120.

Weinstock, G. M., Robinson, G. E., Gibbs, R. A., Worley, K. C., Evans, J. D., Maleszka, R., Robertson, H. M., Weaver, D. B., Beye, M., and Bork, P. 2006. Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443:931-949.

Weintraub, P. G. 2007. Insect vectors of phytoplasmas and their control-an update. Bulletin of Insectology 60:169.

Weintraub, P. G., and Beanland, L. 2006. Insect vectors of phytoplasmas. Annu. Rev. Entomol. 51:91-111.

Welch, E., Hunter, W., Shelby, K., Mizell, R., Tipping, C., Katsar, C., and Bextine, B. 2011. Leafhopper Comparative Genomics-Identifying Similarities and Differences across Leafhopper Vectors of Xylella fastidiosa. Southwestern Entomologist 36:305-321.

Whitfield, A. E., Falk, B. W., and Rotenberg, D. 2015. Insect vector-mediated transmission of plant viruses. Virology 479-480:278-289.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., and Panaud, O. 2007. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8:973.

Wilson, E. O. 1992. The Diversity of Life. New York, W. NY: WW Norton & Company 48.

Wilson, M. R., and Weintraub, P. G. 2007. An introduction to Auchenorrhyncha phytoplasma vectors. Bulletin of Insectology 60:177.

Winston, W. M., Molodowitch, C., and Hunter, C. P. 2002. Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. Science 295:2456-2459.

Wu, G., Wu, Z., Dai, Z., Yang, Y., Wang, W., Liu, C., Wang, B., Wang, J., and Yin, Y. 2013. Dietary requirements of "nutritionally non-essential amino acids" by animals and humans. Amino acids 44:1107-1113.

Wu, Y., Dai, R., Zhan, H., and Qu, L. 2016. Complete mitochondrial genome of Drabescoides nuchalis (Hemiptera: Cicadellidae). Mitochondrial DNA Part A 27:3626-3627.

Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., and Chai, C. 2004. A draft sequence for the genome of the domesticated silkworm (Bombyx mori). Science 306:1937-1940.

Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant molecular biology 80:75-84.

Xu, J., Lu, M.-X., Cui, Y.-D., and Du, Y.-Z. 2017. Selection and Evaluation of Reference Genes for Expression Analysis Using qRT-PCR in Chilo suppressalis (Lepidoptera: Pyralidae). Journal of Economic Entomology 110:683-691.

Xu, W., and Han, Z. 2008. Cloning and phylogenetic analysis of sid-1-like genes from aphids. Journal of Insect Science 8:1-6.

Yandell, M., and Ence, D. 2012. A beginner's guide to eukaryotic genome annotation. Nature Reviews Genetics 13:329.

Yao, J., Rotenberg, D., Afsharifar, A., Barandoc-Alviar, K., and Whitfield, A. E. 2013. Development of RNAi methods for Peregrinus maidis, the corn planthopper. PloS one 8:e70243.

Yin, C., Shen, G., Guo, D., Wang, S., Ma, X., Xiao, H., Liu, J., Zhang, Z., Liu, Y., and Zhang, Y. 2015. InsectBase: a resource for insect genomes and transcriptomes. Nucleic acids research 44:D801-D807.

Yu, N., Christiaens, O., Liu, J., Niu, J., Cappelle, K., Caccia, S., Huvenne, H., and Smagghe, G. 2013. Delivery of dsRNA for RNAi in insects: an overview and future directions. Insect science 20:4-14.

Yuan, M., Lu, Y., Zhu, X., Wan, H., Shakeel, M., Zhan, S., Jin, B.-R., and Li, J. 2014. Selection and Evaluation of Potential Reference Genes for Gene Expression Analysis in the Brown Planthopper, Nilaparvata lugens (Hemiptera: Delphacidae) Using Reverse-Transcription Quantitative PCR. PLOS ONE 9:e86503.

Zahniser, J. N. 2008. Systematics of the leafhopper subfamily Deltocephalinae (Hemiptera: Cicadellidae) and the tribe Chiasmini: phylogeny, classification, and biogeography. University of Illinois at Urbana-Champaign.

Zahniser, J. N. 2017. Deltocephalinae.

Zahniser, J. N., and Dietrich, C. 2013. A review of the tribes of Deltocephalinae (Hemiptera: Auchenorrhyncha: Cicadellidae). European Journal of Taxonomy 45.

Zahniser, J. N., and Dietrich, C. H. 2008. Phylogeny of the leafhopper subfamily Deltocephalinae (Insecta: Auchenorrhyncha: Cicadellidae) and related subfamilies based on morphology. Systematics and Biodiversity 6:1-24.

Zahniser, J. N., and Dietrich, C. H. 2010. Phylogeny of the leafhopper subfamily Deltocephalinae (Hemiptera: Cicadellidae) based on molecular and morphological data with a revised family-group classification. Systematic Entomology 35:489-511.

Zahniser, J. N., Taylor, S. J., and Krejca, J. K. 2010. First reports of the invasive grass-feeding leafhopper Balclutha rubrostriata (Melichar)(Hemiptera: Cicadellidae) in the United States. Entomological News 121:132-139.

Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. 2000. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell 101:25-33.

Zeng, L., Kortschak, R. D., Raison, J. M., Bertozzi, T., and Adelson, D. L. 2018. Superior ab initio identification, annotation and characterisation of TEs and segmental duplications from genome assemblies. PloS one 13:e0193588-e0193588.

Zeng, V., Ewen-Campen, B., Horch, H. W., Roth, S., Mito, T., and Extavour, C. G. 2013. Developmental Gene Discovery in a Hemimetabolous Insect: De Novo Assembly and Annotation of a Transcriptome for the Cricket Gryllus bimaculatus. PLOS ONE 8:e61479.

Zhang, G., Hussain, M., O'Neill, S. L., and Asgari, S. 2013. Wolbachia uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue virus inhibition in Aedes aegypti. Proceedings of the National Academy of Sciences 110:10276-10281.

Zhang, J., Biczok, R., and Ruschhaupt, M. 2015. ddCt: The ddCt Algorithm for the Analysis of Quantitative Real-Time PCR (qRT-PCR). URL http://bioconductor. org. R package version 1:128.

Zhang, W., Chen, J., Keyhani, N. O., Zhang, Z., Li, S., and Xia, Y. 2015. Comparative transcriptomic analysis of immune responses of the migratory locust, Locusta migratoria, to challenge by the fungal insect pathogen, Metarhizium acridum. BMC genomics 16:867.

Zhou, X. R., Horne, I., Damcevski, K., Haritos, V., Green, A., and Singh, S. 2008. Isolation and functional characterization of two independently-evolved fatty acid Δ12-desaturase genes from insects. Insect molecular biology 17:667-676.

Zhu, B.-H., Xiao, J., Xue, W., Xu, G.-C., Sun, M.-Y., and Li, J.-T. 2018. P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. BMC genomics 19:175-175.

Zhu, G.-H., Xu, J., Cui, Z., Dong, X.-T., Ye, Z.-F., Niu, D.-J., Huang, Y.-P., and Dong, S.-L. 2016. Functional characterization of SlitPBP3 in Spodoptera litura by CRISPR/Cas9 mediated genome editing. Insect biochemistry and molecular biology 75:1-9.

VITA

Christian Orlando Ayala Ortiz

Candidate for the Degree of

Master of Science

Thesis: COMPARATIVE GENOMICS OF SIX SPECIES OF DELTOCEPHALINE LEAFHOPPERS TO DISCOVER VECTOR-COMPETENCE-RELATED GENES

Major Field: Entomology and Plant Pathology

Biographical:

Education:

Completed the requirements for the Master of Science in Entomology and Plant Pathology at Oklahoma State University, Stillwater, Oklahoma in December, 2019.

Completed the requirements for the Bachelor of Science in Biotechnology Engineering at University of the Armed Forces – ESPE, Sangolqui, Ecuador in 2016.

Experience:

Graduate Research Assistant at Oklahoma State University, Stillwater, Oklahoma from August, 2017 to December, 2019.

Quality Control Chief at Arenas Bakery, Quito, Ecuador from June, 2016 to February, 2017.

Intern at Oklahoma State University, Stillwater, Oklahoma from August, 2015 to November, 2016.

Professional Memberships:

Entomological Society of America