APPLICATIONS OF POPULATION GENOMICS TO INTEGRATED PEST MANAGEMENT IN COTTON

A Dissertation

by

TYLER JAY RASZICK

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Gregory A. Sword
Committee Members,	Hojun Song
	Aaron M. Tarone
	Giri Athrey
Head of Department,	David W. Ragsdale

December 2018

Major Subject: Entomology

ABSTRACT

In this dissertation, we used a population genomics approach to improve the management of two insect cotton (*Gossypium hirsutum* L.) pests, the boll weevil, *Anthonomus grandis* Boheman (Coleoptera: Curculionidae), and the cotton fleahopper, *Pseudatomoscelis seriatus* (Reuter) (Hemiptera: Miridae). Despite the success of the United States (US) Boll Weevil Eradication Program, the boll weevil remains a threat to cotton production in the southern US, and is arguably the most important cotton pest in Central and South America. The cotton fleahopper is a secondary pest of cotton whose impact in the US has increased in the wake of improved management for other cotton pest species. Management efforts for both species are complicated by a lack of detailed knowledge of population genetic structure, and what little research there is has not taken advantage of high-throughput sequencing technology. Here, we used double digest restriction site-associated DNA sequencing (ddRADseq) to resolve the population genomic structure of the boll weevil in the southern US, northern Mexico, and Argentina, and of the cotton fleahopper in the Brazos Valley of Texas.

Our boll weevil research supported a two-form hypothesis of geographic variants of boll weevil in North America wherein there is a western and eastern form, and suggested that the two variants occur due to geographic isolation rather than host plant association. Boll weevil collections from South America in Argentina were more closely related to the eastern North American boll weevil lineage, but with levels of genetic divergence consistent with isolation-by-distance. We also used a population genomics

approach to identify probable source populations for weevils re-infesting previously eradicated areas in the US.

Our work on the cotton fleahopper revealed high gene flow among populations collected from different host plants in the Brazos Valley. We identified one instance of yearly turnover of local genotypes and one instance of monthly turnover in cotton, indicating that population genomic structure may be labile to time. Our results also identified a probable natural refuge that promotes year-end gene flow between genotypes associated with cotton and those associated with alternative hosts.

DEDICATION

This dissertation is dedicated to my mother, Luanne Betz, and my father George Raszick, who enabled me to continue my education for many years; to my brother, Landen Raszick, my aunt, Ellyn Smith, and my grandmother, Joyce Smith, who encouraged me to pursue a career in science; and to my high school biology teacher, Glenn Pesicek, who helped me discover my passion for biological sciences.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Gregory A. Sword, and my committee members, Dr. Hojun Song, Dr. Aaron M. Tarone, and Dr. Giri Athrey, for their guidance and unflagging support throughout the course of this research.

Computationally intensive components of this research were conducted with the advanced computing resources provided by Texas A&M University High Performance Research Computing. C. Michael Dickens is thanked for his efforts on improving the bioinformatics and utilization of these resources.

I would like to thank Richard Metz of Texas A&M AgriLife Genomics and Bioinformatics Service for his support in planning and carrying out the DNA sequencing. I would also like to thank Jason Wulff, formerly of Texas A&M University, for his assistance with bioinformatics troubleshooting. Isaac Esquivel of Texas A&M University is thanked for contributing map figures. Zoey Kramer of Texas A&M University is thanked for her significant contributions of time and effort to the wet laboratory benchwork.

The boll weevil research in this dissertation would not have been possible without contributions from the Texas Boll Weevil Eradication Foundation and cotton growers in Mexico and Argentina, so I would like to thank them for their efforts. Juan Landivar of Texas A&M AgriLife Research and Extension is also thanked for his efforts helping to organize sample acquisition.

Derrick Hall, Jose Pérez, Kristin Hamons, Yaziri Gonzales, and Amy Berg, all of, or formerly of, the USDA-ARS, are thanked for their efforts collecting and rearing cotton fleahopper for this study. Benjamin Yorke and Douglas Sumerford, both formerly of Monsanto, are thanked for providing support for this project.

Finally, I would like to thank USDA collaborator Xanthe Shirley who not only contributed intellectually and technically with this dissertation, but also supported me personally throughout many years.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Gregory A. Sword (committee chair), Hojun Song, and Aaron M. Tarone, of the Department of Entomology, and Professor Giri Athrey of the Department of Poultry Science.

Samples analyzed in the research were contributed by collaborators with the United States Department of Agriculture Charles P.-C. Suh, Raul Ruiz-Arce, and Theodore N. Boratynski, and a collaborator in Argentina, Marcelo Falco. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

The first year of graduate study was supported by an Excellence fellowship from Texas A&M University.

Additional funding was obtained from the Charles R. Parencia Endowment to Gregory A. Sword, a Cotton Incorporated Core Program grant #16-352, “Boll weevil population genomics for monitoring and eradication” to Gregory A. Sword and Charles P.-C. Suh, and a USDA-ARS Cooperative Agreement, “Using population genomics to reveal the level of intermixing among cotton fleahopper populations in cotton and wild weed hosts” to Gregory A. Sword and Charles P.-C. Suh.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. A Brief History of the Boll Weevil

The boll weevil, *Anthonomus grandis* Boheman (Coleoptera: Curculionidae), is arguably the most important pest of the commercially cultivated varieties of its host, *Gossypium hirsutum* L. (Malvales: Malvaceae), commonly known as upland cotton. The boll weevil is well known as one of the worst pests in United States (US) agricultural history (Lange et al., 2009). During the worst periods of infestation in the US, boll weevil damage accounted for over $200 million in losses per year to the cotton industry, and suppression costs totaled an additional $75 million annually (Cross, 1973). The severe economic impact of this pest led to the implementation of the widely successful Boll Weevil Eradication Program, a collaborative effort of the US Department of Agriculture, cotton producers, and academic researchers, which resulted in the near total eradication of *A. grandis* from the US cotton belt. However, this species does remain a severe pest in much of the New World, particularly in Central and South America, and it nonetheless still poses a threat to the domestic cotton industry. Here, we detail the basic biology and ecology of *A. grandis* and then discuss historical and current approaches to management.

1.1.1. Biology and Ecology

During the cotton-growing season, boll weevil females tend to lay their eggs in moderately sized cotton squares (3-6 mm diameter), though they will also attack larger squares and bolls (Ramalho and Wanderley, 1996). Females eat into the square or boll and deposit eggs into the cavity that is formed. Under normal conditions, eggs may hatch

as quickly as 2-3 days after oviposition, and first instar larvae begin feeding immediately. It takes roughly one week for the larvae to progress through three stadia, feeding in the natal square or boll for the duration. The larvae then pupate for 3-4 days and emerge from the square or boll a few days later. Total development time can range from 12-90 days depending on the temperature of the surrounding environment and the population in question (Cross, 1973; Ramalho and Wanderley, 1996). Adult life span has been reported to be around 40 days. In a normal year, there may be as many as 5-6 generations in a single cotton season. Between cotton seasons, some populations of adult weevils enter a facultative diapause and overwinter (Cross, 1973). The mode of diapause also differs between boll weevil variants. For the southeastern boll weevil, the primary pest variant of the species in the US, adults can hibernate in forest litter. However, other variants may diapause as unfed adults in larval cells in bolls. Emergence from diapause is triggered by early summer rains.

There are at least three morphological variants of boll weevil that have been historically described in North America (Warner, 1966; Burke, 1968; Fye, 1968a; Cross, 1973; Burke et al., 1986). The southeastern boll weevil, *A. g. grandis*, is the variant that was a pest throughout the southeastern United States, and it still occurs in northeastern Mexico, Hispaniola, northern Colombia, Venezuela, and Brazil. Though the southeastern variant is known primarily as a pest of cotton, it may also occur on plants of the genera *Cienfuegosia* and *Hibiscus* when *G. hirsutum* is not available (Cross et al., 1975). The Thurberia weevil, *A. g. thuberiae*, is a variant occurring in southern Arizona and northwestern Mexico that has been traditionally regarded as a host-associated variant

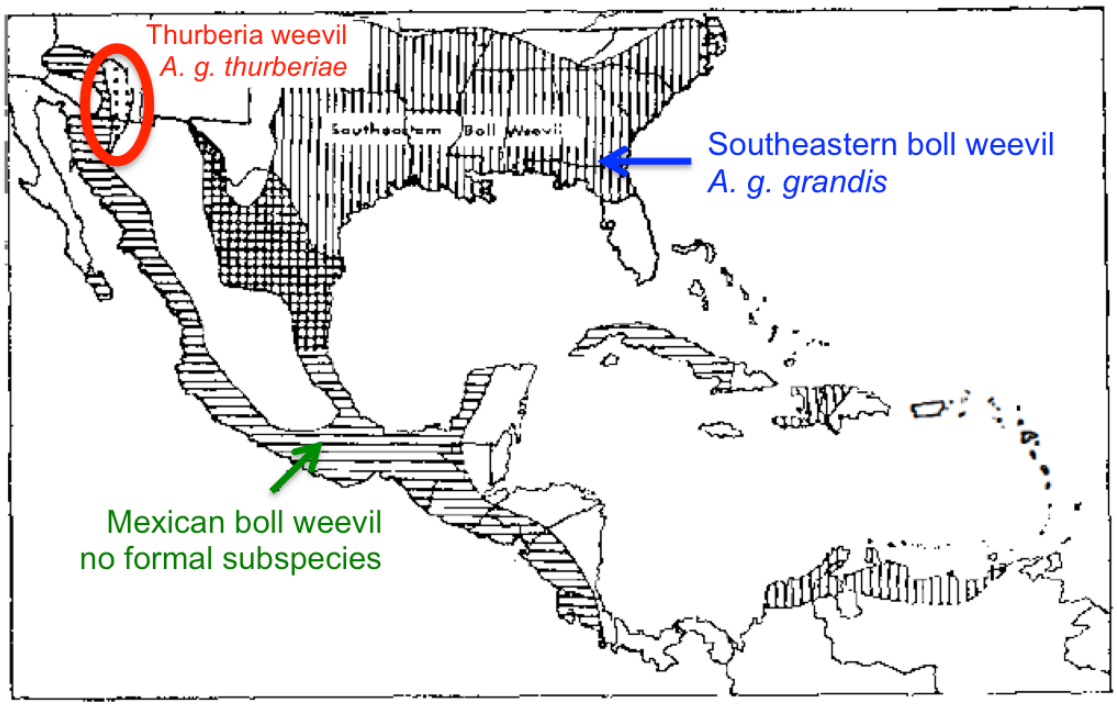

Figure 1. Figure adapted from Cross (1973) illustrating the historical distributions of the three morphological variants of the boll weevil in North America. Cross-hatching in northeastern Mexico indicates range overlap of the southeastern and Mexican boll weevil variants.

utilizing Arizona wild cotton, *Gossypium thurberi* Todaro, but it will opportunistically utilize *G. hirsutum* as a host. The Thurberia weevil is described as having an extremely restricted geographic range (Fig. 1), in correlation with the distribution of its host. The Mexican boll weevil, which has no formal subspecies designation and is sometimes referred to as an intermediate form, is found throughout Mexico, Central America, and Cuba (Fig. 1). Though *A. g. grandis* and *A. g. thurberiae* are widely regarded as distinct subspecies, a number of studies have shown that the morphological characteristics that define the groups are labile to rearing condition (Warner, 1966; Fye, 1968a; Burke et al., 1986). This has led to some confusion in the literature and necessitated that population

genetic studies are carried out in order to determine if there are genetic differences between the two variants. Furthermore, none of these studies adequately included the Mexican boll weevil variant in their analyses.

The common ancestor of the modern variants of boll weevil likely originated in Meso-America in what is now Mexico and other parts of Central America. This ancestral anthonomine was probably associated with malvaceous plants in the genus *Hampea*. This hypothesis has been extensively reviewed in Burke et al. (1986). The rise of these variants can be attributed to range expansions of the ancestral boll weevil mediated by host plant shifts. A northwestern expansion of the ancestral boll weevil range was likely facilitated by air currents strong enough to facilitate long dispersal, and likely gave rise to the Thurberia weevil variant (Fye, 1968b; Burke et al., 1986). A similar northeastern expansion of the ancestral boll weevil's range into northern Mexico likely gave rise to the clade that is now considered to be the southeastern boll weevil (Burke et al., 1986). The southeastern boll weevil was first reported in the United States in Texas in 1894, and the invasion was likely enabled by increased cotton production in southern Texas and northern Mexico in the 1860s (Howard, 1894; Burke et al., 1986). Once established in the US cotton belt, it rapidly spread throughout the area up until it reached the Atlantic coast. Establishment of the southeastern boll weevil in the West Indies and in South America (in Venezuela) is likely due to accidental introductions in the early 1900s and 1940s, respectively (Whitcomb and Britton, 1953; Burke et al., 1986). The range of the boll weevil in South America then expanded into Colombia in the 1950s and to Brazil in the 1980s (Marin, 1981; Sobrinho and Lukefahr, 1983). Though the boll

weevils in South America are considered to be morphologically similar to the southeastern boll weevil, only a single genetic study to date has yet attempted to determine the origins of these weevils using molecular methods (Scataglini et al., 2006).

1.1.2. Impact on US Agriculture and the Boll Weevil Eradication Program

The introduction of *A. g. grandis* to Texas in 1894 marked the beginning of a long and costly war. Lange *et al.* (2009) refer to the boll weevil as America's most celebrated pest, and some attribute the diversification of the economy of the American South to the invasion of the boll weevil. However, despite the fact that the cotton industry in the US was expanding during the late 1800s and early 1900s, the establishment of the pest in the US cotton belt devastated local economies. Cotton yields were reduced across the board by about 50 percent, and boll weevil damage has accounted for over $200 million in losses per year to the cotton industry (Cross, 1973; Lange et al., 2009). Suppression costs have totaled an additional $75 million annually, and the total cost to US cotton producers is estimated at more than $15 billion (Cross, 1973; National Cotton Council, 2010). Yield loss occurs because squares and bolls that have been oviposited in tend to abscise and fall to the ground, ultimately reducing the amount of lint that the plant produces. (Ramalho and Wanderley, 1996).

Due to the large impact on US agriculture in the 1900s, there was a great need for effective methods of suppression of boll weevil populations infesting cotton. When the pest was still a relatively new invader to the US, chemical control was the primary approach to suppression. Methyl parathion was a popular chemical insecticide used by cotton growers, but it was most effective in supplementation with other chemicals,

particularly DDT, and likely had detrimental effects on beneficial insects as well as pests (McGarr and Chapman, 1966; McGarr and Wolfenbarger, 1969; McGarr and Wolfenbarger, 1970). Other pesticides showed varying levels of effectiveness, and often had to be specifically tweaked towards boll weevil suppression (Cross, 1973). Research showed that at least one insecticide, aldicarb, was effective against emerging overwintered weevils, granting effective early season control (Coppedge et al., 1969). Finally, as the weevil developed resistance to these numerous pesticides, it became clear that other methods of control would need to be implemented in order to successfully control populations of the boll weevil (Eden, 1968; Cross, 1973).

One of the primary concerns with the continued use of broad-spectrum insecticides for boll weevil control was the non-target impact on beneficial insects such as pollinators and the natural enemies of the boll weevil. In the 1920s and 1930s, natural enemies were one of the primary components of biological control upon which farmers relied; however, this was only somewhat effective (Lincoln, 1969; Cross, 1973). There are at least 42 known species of arthropods that attack the boll weevil; however, none of these provide complete suppression of local populations at natural levels, though releases of parasitoids have been employed with limited success (Cross and McGovern, 1969; Cross et al., 1969a; Cross and Chesnut, 1971). Ironically the best source of biological control has come from the invasion of another pest species; the red imported fire ant, *Solenopsis invicta* Buren, has been shown to significantly suppress weevil populations in some areas (Sterling, 1978).

In lieu of widespread pesticide use and marginally effective biological controls, the cotton community was forced to develop alternative methods of control that exploit the biology and ecology of *A. grandis* in order to suppress its populations. One of the primary ways US growers attempt to thwart boll weevil persistence in between seasons is through mechanical destruction of cotton stalks at the end of the growing season (Cross, 1973). This drastically reduces the overwintering habitat for weevils, and is particularly effective in areas where there is not a high availability of alternative hosts. During the growing season, collection and destruction of infected bolls and squares can significantly reduce the number of weevils without further impacting yield, and a variety of methods have been developed to carry this out (Burt et al., 1968; Mistric and Covington, 1968; Parencia, 1968; Burt et al., 1969). Another commonly used strategy has been the diapause method (Cross, 1973). The diapause method is actually a combination of chemical control and cultural practices that involves precise timing of pesticide application (usually methyl parathion prior to 1950s, usually malathion more recently) in order to reduce the number of reproductive weevils that will produce the overwintering generation. The timing of this application varies from region to region depending on when weevils in that area tend to enter diapause. This practice is supplemented by early plant destruction in order to reduce the weevils' food source. The diapause method has been extremely successful in suppression of boll weevil populations in many areas (Lloyd et al., 1966; Lloyd et al., 1967; Bottrell and Almand, 1968; Fye et al., 1968; Rummel and Adkisson, 1971; Lloyd et al., 1972).

In the 1960s, using specialized traps for the capture of boll weevil began for the purposes of monitoring and control. Though not an effective measure of control on their own, baited traps were extremely effective for monitoring populations of boll weevil (Cross and Hardee, 1968; Cross et al., 1969b). Early trapping efforts used live male weevils as bait, but isolation of the boll weevil pheromone, grandlure, eventually made it more practical to use the pheromone as the trapping bait (Tumlinson et al., 1969; Hardee et al., 1971). Though the design of the traps has changed over the years, pheromone-baited traps remain one of the most important staples of the US monitoring strategy for *A. grandis* (Cross, 1973).

Despite the numerous developments of different control strategies over the years, the boll weevil remained a significant US agricultural pest throughout much of the 1900s, and total eradication of the pest was a desirable goal. In 1958, the National Cotton Council officially recognized the economic impact of the boll weevil and began working with Congress to develop an eradication plan (National Cotton Council, 2010). The Boll Weevil Eradiation Program was an effort designed to coordinate management efforts using combinations of the previously described management tools in order to eradicate the boll weevil from the entirety of the US cotton belt. The Eradication Program is widely regarded as one of the earliest and most successful implementations of an integrated pest management (IPM) approach. The main components of the approach are use of pheromone-baited traps for monitoring and detection, cultural practices to reduce boll weevil habitat, and timely, minimal use of malathion (an organophosphate insecticide to which the weevil has not developed resistance). In 1978,

USDA-APHIS launched the first phase of the program in Virginia and North Carolina. The program later expanded into the rest of cotton belt, reaching as far west as California. As of 2009, eradication of the boll weevil was completed in the US, except for a small persistent region in south Texas.

Despite the success of the Eradication Program, the boll weevil remains a significant threat to domestic cotton production due to continuing infestations along the US-Mexico border. This situation can be attributed to control discrepancies across the border, insecurity and safety issues that affect management, and the fact that this area is a part of the species' natural range. This continued insecurity has led to at least one recent re-infestation of domestic cotton in an area where eradication had previously been successful (Texas Boll Weevil Eradication Foundation, 2018). This re-infestation has required large-scale, area-wide spraying of malathion, cost over $5 million over 3 years to treat. Though that outbreak was restricted to central Texas, such events are of broader concern at the national level because Texas is the single largest corridor of entry for potential re-infestations of US cotton by populations of boll weevil migrating or otherwise being moved from Mexico. This highlights the current need for preventative management along the US-Mexico border. Currently, pheromone-baited traps and strict regulations on cotton growers are employed to prevent re-infestations of areas where the boll weevil has been eradicated. In cases where weevils are detected in traps, malathion is used to suppress emerging populations. This approach has been mostly successful at maintaining eradication. In Central and South America, where the weevil is still a major

concern, implementation of an IPM approach similar to the Eradication Program is underway (Ramalho and Wanderley, 1996).

1.1.3. Population Genomics to Improve Boll Weevil Eradication in the US

If the ultimate goal of the Eradication Program is to fully eradicate the weevil from the US, then the movement patterns of Mexican and Texas populations will need to be fully resolved. Control of the boll weevil in border areas is complicated by the existence of the aforementioned morphologically similar variants and a lack of knowledge about the movement of populations between eradicated areas and those where there are still infestations. Understanding the movement of populations is critical to management because inappropriately applied control strategies can create source-sink dynamics that nullify the effects of local suppression (Hanski and Gilpin, 1991; Harrison, 1991; Zaller et al., 2008; Sword et al., 2010; Carrière et al., 2012). In this dissertation, we demonstrate the application of a population genomics approach to enable the inference of movement of boll weevils across the species' range. Furthermore, we also genetically identify source populations of re-infestations of previously eradicated areas.

1.2. Population Genomics for Insect Resistance Management in Cotton Fleahopper

Another pest of cotton in North America is the cotton fleahopper, *Pseudatomoscelis seriatus* (Reuter) (Hemiptera: Miridae). Prior to the eradication of the boll weevil and the development of *Bacillus thuringiensis* (Bt) toxin approaches to management for lepidopteran pests, the cotton fleahopper was generally regarded to be a secondary pest of commercial cotton. However, since the boll weevil is nearly eradicated and Bt transgenic cotton varieties have become widespread, *P. seriatus* has been identified as an

important target for management because the species attacks cotton squares early in the growing season, which causes abscission of the squares and subsequently leads to yield losses (Ewing, 1929; Powell, 1979). In 2017, Monsanto Company developed a new Bt transgenic cotton cultivar that is resilient to a variety of mirid species including *P. seriatus* (Bachman et al., 2017). Since previous research has described a pattern of host-associated genetic differentiation (HAD) in the cotton fleahopper, there is concern from an insect resistance management (IRM) perspective that there could be a cotton-associated host-race (Barman et al., 2012; Antwi et al., 2015). That would have important implications for management, particularly with regards to the evolution of resistance to the transgenic control over time (Georghiou and Taylor, 1977; Guse et al., 2002; Onstad et al., 2002). If there are cotton-associated genotypes that do not exchange gene flow with genotypes associated with other host plants, then the cotton-associated genotypes could, in principle, quickly develop resistance to the transgenic cotton (Gould, 1998; Caprio, 2001). This necessitates the use of non-toxic refuge plants to maintain insect susceptibility. A refuge strategy has been applied to maintaining insect susceptibility to Bt transgenic crops (Caprio, 2001; Caprio et al., 2004; Carrière et al., 2004). In the case of the cotton fleahopper, there may be a natural refuge in wild host plants that occur near cotton-growing areas. However, if there is not, a refuge strategy will need to be deliberately deployed as part of an IRM plan to maintain insect susceptibility. In this dissertation, we apply a population genomics approach to determine if there is a natural refuge for cotton fleahopper that acts as a site of admixture for populations feeding on cotton and those feeding on non-toxic wild host plants.

2. POPULATION GENOMIC DIVERSITY OF THE BOLL WEEVIL, *ANTHONOMUS GRANDIS* BOHEMAN (COLEOPTERA: CURCULIONIDAE), AND ITS VARIANTS IN THE UNITED STATES, NORTHERN MEXICO, AND ARGENTINA

2.1. Introduction

Anthonomus grandis Boheman (Coleoptera: Curculionidae), commonly known as the boll weevil, is a major pest of commercially cultivated upland cotton, *Gossypium hirsutum* L. (Malvaceae), across the Americas. Despite the success of the United States Boll Weevil Eradication Program, the boll weevil remains a continued threat to cotton production in the southern United States (US), and is arguably the most important cotton pest in Central and South America. Management of this species can be complicated by the existence of morphologically similar variants (Warner, 1966; Burke, 1968; Fye, 1968a; Burke et al., 1986) that can confound identification efforts. It is also important for management and eradication programs to understand broad-scale patterns of gene flow because inappropriately applied control strategies can create source-sink dynamics that nullify the effects of local suppression (Hanski and Gilpin, 1991; Harrison, 1991; Zaller et al., 2008; Sword et al., 2010; Carrière et al., 2012). Studies of population genetics can provide solutions to some of these management problems by enabling more reliable diagnoses of boll weevil variants and by identifying which populations act as sources for re-infestations of areas where eradication has been successful. Population genetic markers have been shown to be effective for determining the sources of re-infestations (Kim et al., 2006; Kim et al., 2008), and haplotype analysis of the

cytochrome *c* oxidase subunit I gene (COI) is currently utilized by the US Department of Agriculture to diagnose pest and non-pest boll weevil variants collected in pheromone-baited monitoring traps (Barr et al., 2013). While a number of studies have investigated the population genetic structure of the boll weevil in North America (Bartlett, 1981; Roehrdanz, 2001; Kim and Sappington, 2004; Kim and Sappington, 2006; Barr et al., 2013; Alvarado et al., 2017) or South America (Scataglini et al., 2000; Martins et al., 2007), only one has investigated the genetic relationship of populations on the broader geographic scale of both continents (Scataglini et al., 2006). No study to date has taken advantage of high throughput sequencing (HTS) technology to generate a powerful genome-wide multilocus dataset that can provide substantially more resolution than classic population genetic markers.

It is generally accepted that the most recent common ancestor of the boll weevil and its closely related variants originated in southern Mexico and Central America and diverged from the sister species, *Anthonomus hunteri* Burke and Cate, during the Pliocene (Burke et al., 1986; Alvarado et al., 2017). The original host plant for this weevil species was probably *Hampea* Schltdl. spp. (Malvaceae), and the weevil underwent at least one host shift to endemic *Gossypium* L. species, and later shifted to *G. hirsutum* after its cultivation began in the Americas (Howard, 1894; Burke et al., 1986). In the late 1800s, the boll weevil greatly expanded its geographic range northward through Mexico and eventually across the entire cotton growing region of the southern US where it became an infamous agricultural foe (Burke et al., 1986; Lange et al., 2009). Classic descriptions of boll weevil variants since this range expansion have

generally referred to three forms: the southeastern boll weevil (*A. g. grandis*), the Thurberia weevil (*A. g. thurberiae*), which has traditionally been regarded as associated with Arizona wild cotton, *Gossypium thurberi* Todaro (Malvaceae), and the Mexican boll weevil, an intermediate form with no formal subspecies designation (Warner, 1966; Cross et al., 1975; Burke et al., 1986). However, subspecies status has been inconsistently applied to these variants, and recent research has suggested that the genetic lineage designated as *A. g. thurberiae* may be divergent due to the vicariant effect of the Sierra Madre Occidental mountain range, rather than host plant association (Kuester et al., 2012; Alvarado et al., 2017). These studies have opposed the three-form hypothesis altogether, instead suggesting a two-form hypothesis wherein the Thurberia weevil is treated as the western form and other weevils are treated as the eastern form. A second boll weevil range expansion has more recently occurred in South America. Similar to the boll weevil advance throughout North America over the past 150 years, the range expansion in South America has been attributed to the expanding cultivation of commercial cotton. The boll weevil was first recorded in Venezuela in 1949, Colombia in 1951, Brazil in 1983, Paraguay in 1991, Argentina in 1993, and Bolivia in 1997 (Scataglini et al., 2006). By 2016, the boll weevil had spread as far south as the Argentine province of Santiago del Estero and as far west as the province of Salta. Scataglini et al. (2006) found that the introduction of boll weevil to South America was associated with a range expansion of the southeastern form.

In this study, we used double digest restriction site-associated DNA sequencing (ddRADseq, Peterson et al., 2012) to generate a population genomic dataset of single

nucleotide polymorphism (SNP) markers as a means to better understand spatial and temporal patterns of variation in boll weevil genetic population structure. We sampled boll weevil populations on a broad geographic scale and generated a SNP dataset with thousands of loci to better resolve the population genetic structure of the species. Using this dataset, we formally test the two-form and three-form hypotheses using a phylogeographic approach and identify geographic populations within those lineages. We hypothesized that there are two major geographic lineages of boll weevil but that there is also significant sub-structure within those lineages. Finally, we discuss the implications of our findings for boll weevil management in northern Mexico and the southern US.

2.2. Materials and Methods

2.2.1. Specimen Sampling

A total of 292 weevil specimens were collected and processed from Arizona and cotton-producing areas of south Texas, Mexico, and Argentina (Table 2.1, Fig. 2). Weevil specimens were mainly collected using boll weevil pheromone-baited traps (Cross and Hardee, 1968; Cross et al., 1969b; Tumlinson et al., 1969; Hardee et al., 1971), whereas those from Arizona were collected directly from Arizona wild cotton (*G. thurberi*) using a beat bucket technique. Insects from all localities were collected alive as adults and preserved in 95-100% ethanol. Other than during shipping or transportation, all specimens were stored at -80°C until they were prepared for DNA isolation. Since at any one collection locality there may be multiple pheromone-baited traps, for those collections, the midpoint GPS coordinates were determined from the GPS coordinates of

Table 2.1. Collection information and within-population genetic summary statistics for all boll weevil collections in the study. Locality is used to denote different collections from the same state and year. N is the number of individuals analyzed from each collection. $1\text{-}Q_{intra}$ is the average genetic diversity within individuals and $1\text{-}Q_{inter}$ is the average genetic diversity among individuals within a collection. F_{IS} is the inbreeding coefficient.

Date	Country	State/Prov.	Locality	Latitude	Longitude	N	$1\text{-}Q_{intra}$	$1\text{-}Q_{inter}$	F_{IS}
22-Sep-14	Mexico	Sonora	-	27.4209	-109.9758	5	0.2006	0.1682	-0.1929
22-Sep-14	Mexico	Chihuahua	-	28.3431	-105.572	4	0.1252	0.0909	-0.3763
23-Sep-14	Mexico	Durango	-	26.1229	-103.4147	5	0.1411	0.112	-0.2602
12-Sep-14	Mexico	Tamaulipas	-	25.8247	-98.0672	16	0.1408	0.1311	-0.0741
Aug/Sep-14	USA	Texas	-	26.0713	-97.4655	18	0.1402	0.1309	-0.0712
28-Aug-16	USA	Arizona	Mt. Lemmon	32.3262	-110.7004	12	0.1662	0.14	-0.1875
29-Aug-16	USA	Arizona	Sahuarita	31.9633	-110.8075	12	0.162	0.1377	-0.1766
29-Aug-16	USA	Arizona	Highway 83	31.947	-110.664	12	0.1532	0.1337	-0.1459
29-Aug-16	USA	Arizona	Agua Caliente	31.6845	-110.9585	12	0.1748	0.1454	-0.2021
30-Aug-16	USA	Arizona	Bisbee (West)	31.4877	-109.9873	12	0.1521	0.1337	-0.138
30-Aug-16	USA	Arizona	Bisbee (East)	31.4421	-109.8268	12	0.1481	0.1295	-0.1438
Jul/Aug-16	Mexico	Tamaulipas	-	25.8283	-98.0561	34	0.1511	0.1414	-0.069
Jul/Aug-16	USA	Texas	-	26.1594	-97.8234	30	0.1309	0.1083	-0.2084
Aug/Sep-17	Mexico	Sonora	-	27.3086	-109.9939	30	0.1848	0.1646	-0.1232
7-Aug-17	Mexico	Coahuila	-	25.8134	-102.991	30	0.131	0.111	-0.1805
Jun/Jul-17	Argentina	Chaco	Gral. Pinedo	-27.2533	-61.4942	12	0.1084	0.077	-0.4085
Jun/Jul-17	Argentina	Chaco	Saenz Peña	-26.8553	-60.4378	8	0.1012	0.072	-0.4053
Jun/Jul-17	Argentina	Salta	-	-25.4256	-63.8483	8	0.1009	0.0695	-0.4507
Jun/Jul-17	Argentina	S. del Estero	-	-29.2397	-62.9083	8	0.1029	0.0758	-0.3576
Jun/Jul-17	Argentina	Formosa	-	-24.6978	-59.4717	12	0.1059	0.0739	-0.4325

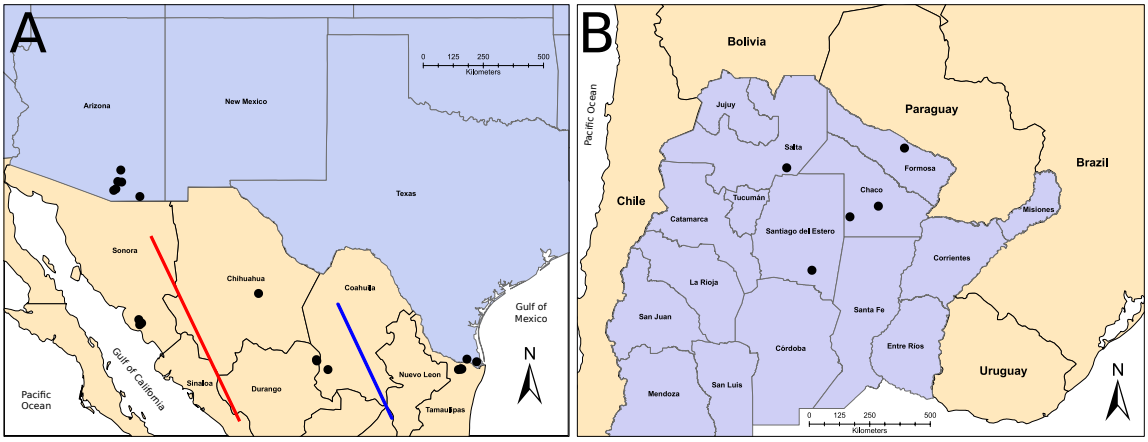

Figure 2. Geographic distribution of sampled populations (black dots) in North America (A) and Argentina (B). In A, the red bar shows the approximate location of the Sierra Madre Occidental mountain range and the blue bar shows the approximate location of the Sierra Madre Oriental mountain range.

traps using the geographic midpoint calculator available from www.geomidpoint.com using the center of gravity method.

Weevils were first collected in 2014 in northern Mexico from the cotton-producing states of Sonora, Chihuahua, Durango, and Tamaulipas, as well as the Lower Rio Grande Valley (LRGV) cotton production area in Texas, just north of Tamaulipas along the US-Mexico border (Fig. 2A). In 2016, the LRGV and Tamaulipas localities were resampled in Texas and northern Mexico, respectively. Weevil specimens were also collected in 2016 from six localities in southeastern Arizona. In 2017, the Sonora locality was resampled, but weevils from Coahuila, Mexico were obtained instead of resampling the Durango and Chihuahua localities due to variation in weevil presence from year to year. The Coahuila and Durango sampling localities were only 54.58 km apart and in close proximity to Torreon, a city near the border of the two states. Specimens from Argentina were also collected in 2017 from pheromone traps established in the four cotton-producing provinces of Chaco, Salta, Santiago del Estero, and Formosa (Fig. 2B).

2.2.2. DNA Isolation, Library Preparation, and Sequencing

The Gentra Puregene Cell and Tissue Kit (Qiagen, Hilden, Germany) was used to isolate genomic DNA from whole weevil specimens. Individuals collected in 2014 were processed and sequenced in 2015 and individuals collect in 2016 and 2017 were processed and sequenced in 2017. DNA was isolated from individuals in both years using the same protocol (Appendix A), but with a slight modification in that 2015, specimens were mechanically disrupted prior to tissue lysing using dissecting scissors,

whereas in 2017, specimens were mechanically disrupted by freezing in liquid nitrogen and crushed with disposable pestles. Isolated DNA from all 292 specimens was verified for high molecular weight via electrophoresis on a 1.5% agarose gel.

Genomic DNA isolated from weevils was delivered to the Texas A&M AgriLife Genomics and Bioinformatics Service (TxGen) for purification, library preparation and sequencing. DNA was purified using the Agencourt AMPure XP purification system (Beckman Coulter, Brea, CA, USA) prior to library preparation. Double digest restriction site-associated DNA sequencing (ddRADseq) library preparation was nearly identical in 2015 and 2017, but 2015 libraries were prepared for a HiSeq 2500 (Illumina, San Diego, CA, USA), and 2017 libraries were prepared for a NovaSeq (Illumina, San Diego, CA, USA). To prepare the ddRADseq libraries, purified genomic DNA was digested using the NlaIII and HindIII restriction enzymes, and the resulting digested DNA was selected for fragment sizes ranging from 250 to 500 base pairs (bp) using a Pippin Prep (Sage Science, Beverly, MA, USA). Size-selected fragments were then ligated with standard Illumina adapters, multiplexing indexes, and sequencing primers, albeit with a single notable exception; the R1 reads (forward reads; those sequenced in the 5' direction) were ligated with a custom sequencing primer that contained the 5' restriction site remnant. 2015 libraries were sequenced on a HiSeq 2500 using 125x125 sequencing cycles, and 2017 libraries were sequenced on a NovaSeq using 150x150 sequencing cycles.

2.2.3. Raw Sequence Processing and SNP Calling

TxGen provided 1.422779 TB of demultiplexed raw reads and FastQC version 0.11.3 (Andrews, 2010) reports for each specimen to the authors. Potential bacterial contamination was filtered using Kraken version 1.1 (Wood and Salzberg, 2014) to match sequences to the non-redundant bacterial database hosted by the National Center for Biotechnology Information (NCBI). Homologous SNP loci were identified using the software pipeline dDocent version 2.5.5 (Puritz et al., 2014a; Puritz et al., 2014b) on the entire dataset of 2015 and 2017 sequences. dDocent was run using default parameters, except for that after initial trimming, dDocent was paused, and Trimmomatic version 0.38 (Bolger et al., 2014) was used to further trim sequences to a uniform length of 90 bp. dDocent was then restarted using the 90 bp sequences as the input and a percent clustering similarity parameter of 90%. VCFtools version 0.1.15 (Danecek et al., 2011) was used to filter the dDocent output in variant call format (vcf). SNP loci that were not present in 100% of individuals were also removed in order to create a SNP dataset with 0% missing data. All file conversions needed for downstream analyses were carried out using PGDSpider version 2.1.1.3 (Lischer and Excoffier, 2012).

2.2.4. Population Genetic Analyses

RStudio version 1.1.456 (R Core Team, 2018) and some associated packages were used to generate population genetic summary statistics for each previously described collection. R/vcfR version 1.8.0 (Knaus and Grünwald, 2017) was used to read the filtered vcf file and prepare objects for use with other packages. R/adegenet version 2.1.1 (Jombart, 2008; Jombart and Ahmed, 2011) was used to create a biallelic genlight

object that was used carry out many of the downstream analyses. First, to verify that any observed population genetic structure was not due to sequencing batch effects, we carried out an analysis of molecular variance (AMOVA, Excoffier et al., 1992) using the year of sequencing as a hierarchical subdivision of the data above collection. The AMOVA was carried out using R/poppr version 2.8.1 (Kamvar et al., 2014; Kamvar et al., 2015). Significance testing was carried out using Monte Carlo resampling permutation with R/ade4 version 1.7-13 (Dray and Dufour, 2007; Bougeard and Dray, 2018).

R/genepop version 1.0.5 (Raymond and Rousset, 1998; Rousset, 2008) was used to estimate gene flow between pairs of collections by calculating pairwise F_{ST} values (Weir and Cockerham, 1984) for all pairs of collections in the study. We also calculated the global F_{ST} for the entire dataset. Pairwise exact conditional contingency-table tests for genotypic differentiation (dememorization = 1000, batches = 10, iterations = 500) were also implemented to determine if genetic differences between collections were statistically significant. R/adegenet was used to carry out a principal components analysis (PCA) using the biallelic genlight object, and R/ggplot2 version 3.0.0 (Wickham, 2016) was used to visualize the spatial clustering of individual genotypes by plotting principal component 2 as a function of principal component 1.

The software fastSTRUCTURE version 1.0 (Raj et al., 2014) was used to calculate each sampled individual's probability of assignment to one or more predetermined genotypic groups (K). PLINK version 1.07 (Purcell et al., 2007) was used to convert the vcf file into a format that was suitable for input into fastSTRUCTURE.

The browser-based program StructureSelector (Li and Liu, 2018) was then used to evaluate the fastSTRUCTURE outputs by initially testing K values from 1 to 30 to choose the optimal K value for our dataset. Optimal K values were determined using both a maximum marginal likelihood approach and the Puechmaille (2016) method of K selection, which accounts for unevenness in sampling distribution. For the Puechmaille (2016) method, we tested mean membership coefficient thresholds of 0.5, 0.6, 0.7, and 0.8, as recommended by the author. CLUMPAK (Kopelman et al., 2015), which is integrated into StructureSelector, was used to visualize individual assignment probabilities.

We also calculated within-population genetic diversity and the inbreeding coefficient (F_{IS}) for each collection using R/genepop. We measured within-population genetic diversity by calculating the average genotypic diversity within individuals (1-Q_{intra}) and among individuals within a collection (1-Q_{inter}) based on allele identify. Weir & Cockerham (1984) F_{IS} was measured at each locus and then averaged across individuals for each collection.

2.2.5. Mantel Test for Isolation by Distance

To test if any observed population genetic structure was consistent with the isolation by distance (IBD) model (Wright, 1943; Rousset, 1997), we used option 6 and sub-option 9 of the web implementation of Genepop version 4.2 (Raymond and Rousset, 1998; Rousset, 2008) to run the Isolde program. Isolde investigated the correlation between a semi-matrix of pairwise genetic distances and a semi-matrix of pairwise geographic distances for all pairs of collections in our dataset using a Mantel test. For the genetic

distance semi-matrix, we used the values of F_{ST} calculated by R/genepop adjusted to $F_{ST}/(1-F_{ST})$. Geographic distance was calculated as natural logarithm of the straight-line distance (in kilometers) between pairs of GPS coordinates. The adjusted values of F_{ST} and straight-line distance were extracted from the Isolde output and plotted in Microsoft Excel to calculate the slope and intercept of the linear regression and calculate the R^2 value.

2.2.6. Phylogenetic Reconstruction

We accessed the software RAxML version 8.2.10 (Stamatkis, 2014) via the CIPRES Science Gateway version 3.1 (Miller et al., 2010) to conduct the phylogenetic reconstruction. Four phylogenetic trees were generated using the GTR+γ and GTR+γ+I models of nucleotide evolution and both models were tested with and without correcting for ascertainment bias (Leaché et al., 2015). The tree with the highest maximum log likelihood score was selected as the best possible reconstruction. Since we did not sequence any appropriate outgroup, we midpoint-rooted the best tree by inferring a most recent common ancestor along the longest branch of the unrooted tree.

2.3. Results

Our dDocent run identified 170,993 homologous SNP loci, which were subsequently filtered to 7,177 loci that were of sufficiently high quality in 100% of individuals for the 292 weevils described here. R/adegenet identified 524 loci with more than two alleles, so analyses conducted using the genlight object (AMOVA, PCA) were carried out using a subset of 6,653 biallelic loci. In general, we found that there was a high degree of genetic diversity across the geographic distribution of boll weevil. Among-population

Table 2.2. Analysis of molecular variance (AMOVA) and Monte Carlo permutation test results. D. F. is the degrees of freedom. % variation is the percentage of the variation in the data explained by the corresponding hierarchical level.

	AMOVA results			Components of covariance		MC perm.
	D. F.	Sum sq.	Mean sq.	Sigma	% variation	P-value
Between sequencing year	1	48713.72	48713.7165	14.26773	0.7314032	0.52
Between collections within seq. year	12	271694.08	22641.1732	1203.85897	61.7131488	0.01
Within collections	278	203664.65	732.6066	732.60664	37.555448	0.01
Total	291	524072.44	1800.9362	1950.73334	100	

measures of genetic diversity indicated that the sampled populations were strongly divergent, yet within-population measures of genetic diversity did not indicate that the populations were genetically isolated. The hierarchical AMOVA revealed that within-population and among-population variation together explained nearly 100% of the observed variation in the dataset (Table 2.2). The year in which a particular collection was sequenced was found to be non-significant with regards any observed population genetic structure, providing evidence against a possible sequencing batch effect.

2.3.1. Population Genetic Analyses

Among-population measures of genetic diversity indicated that the sampled collections were genetically distinct and highly structured. The global F_{ST} for the entire dataset was high, as were most of the pairwise comparisons (Table 2.3). Pairs of collections with low pairwise F_{ST} values (<0.05) were found to be not significantly different by the exact conditional contingency-table test, and were considered to be indistinguishable from one another. The majority of these comparisons were among Argentine collections and among Arizona collections, but the comparison of the 2014 Texas collection and the 2014 Tamaulipas collection was also found to be non-significant. Genetic distance was

Table 2.3. Semi-matrices of pairwise comparisons of linear geographic distance (km, above the diagonal) and genetic distance (F_{ST}, below the diagonal). Superscripts on the values in the F_{ST} semi-matrix indicate the results of the exact conditional contingency table tests of population differentiation (NS indicates P>0.05 (not significant); all other pairwise comparisons are found to be statistically significant). Bold values of F_{ST} are those that are comparisons of the same geographic location, but in different years.

		Global F_{ST} 0.3858	2014					2016								2017						
			Mex				TX, USA	Arizona, USA						Mexico	TX, USA	Mexico		Argentina				
			Son.	Chi.	Dur.	Tam.	LRGV	Mt. L.	Sah.	H. 83	A. Cal.	B. (W)	B. (E)	Tam.	LRGV	Son.	Coa.	C. (GP)	C. (SP)	Salta	S. del E.	For.
2014	Mexico	Sonora	-	444.78	667.06	1196.52	1250.73	549.9	511.43	507.65	483.52	452.21	447.37	1197.53	1213.87	12.61	716.88	7984.8	8027.09	7675.75	8048.82	7928.97
		Chihua.	0.3718	-	326.23	793.8	840.3	661.94	644.38	632.56	637.83	550.67	535.61	794.68	803.34	449.76	380	7765.56	7801.03	7459.02	7845.84	7688.04
		Durango	0.3383	0.1561	-	535.54	594.05	987.19	968.1	956.63	959.54	874.82	860.07	536.62	558.1	666.55	54.58	7439.32	7474.84	7132.8	7519.92	7362.31
		Tamau.	0.2935	0.1942	0.1541	-	66.12	1423.15	1414.2	1401.22	1414.09	1321.41	1305.48	1.18	44.48	1197.1	492.82	7082.45	7110.51	6780.09	7180.21	6982.71
	TX, USA	LRGV	0.2937	0.1915	0.1528	0.0103^{NS}	-	1459.16	1451.78	1438.64	1452.96	1359.62	1343.63	64.94	37.05	1251.7	553.2	7069.33	7096.15	6767.8	7169.83	6965.81
2016	Arizona, USA	Mt. Lem.	0.2274	0.4959	0.4752	0.4439	0.4429	-	41.59	42.3	75.39	115	128.33	1423.86	1423.6	562.07	1041.38	8424.51	8460.99	8117.46	8501.69	8349.47
		Sahuarita	0.2392	0.5181	0.4984	0.4686	0.4674	0.0213^{NS}	-	13.66	34.13	93.89	109.39	1414.94	1415.94	523.51	1022.49	8402.51	8439.61	8095.21	8478.29	8329.46
		Hwy. 83	0.253	0.5293	0.509	0.4784	0.4771	0.023^{NS}	0.0003^{NS}	-	40.33	81.89	97.09	1401.95	1402.83	519.81	1010.98	8391.8	8428.75	8084.55	8467.92	8318.29
		Agua Cal.	0.2187	0.4779	0.4573	0.4286	0.4279	0.0023^{NS}	0.0253^{NS}	0.026^{NS}	-	94.56	110.56	1414.84	1416.91	495.44	1014.05	8390.23	8427.88	8082.7	8464.72	8318.96
		Bis. (West)	0.25	0.521	0.5021	0.4701	0.4678	0.0383^{NS}	0.036^{NS}	0.031^{NS}	0.041^{NS}	-	16.04	1322.15	1323.66	464.7	929.15	8310.63	8347.4	8003.46	8387.25	8236.65
		Bis. (East)	0.2651	0.5345	0.5141	0.4799	0.478	0.0456^{NS}	0.04^{NS}	0.0332^{NS}	0.0473^{NS}	0.0026^{NS}	-	1306.23	1307.68	459.91	914.37	8296.46	8333.1	7989.35	8373.4	8222.09
	Mexico	Tamau.	0.3335	0.232	0.2028	**0.1009**	0.095	0.4426	0.4676	0.4761	0.4276	0.4661	0.4752	-	43.55	1198.12	493.92	7082.12	7110.16	6779.77	7179.92	6982.31
	TX, USA	LRGV	0.423	0.399	0.3742	0.2607	**0.2557**	0.5099	0.5298	0.5372	0.4941	0.5292	0.5379	0.1964	-	1214.88	517.91	7098.18	7125.44	6796.34	7197.68	6995.97
2017	Mexico	Sonora	**0.0872**	0.4007	0.3834	0.3762	0.376	0.1423	0.1413	0.1501	0.1405	0.1567	0.1736	0.3896	0.4329	-	715.95	7977.15	8019.61	7668.05	8040.76	7921.9
		Coahuila	0.4232	0.254	0.1454	0.3073	0.3028	0.5084	0.5287	0.5357	0.4949	0.5279	0.5379	0.2697	0.3936	0.4157	-	7385.72	7421.05	7079.28	7466.79	7308.19
	Argentina	Chaco (GP)	0.4794	0.5135	0.4754	0.3312	0.3245	0.523	0.548	0.5568	0.5084	0.5308	0.5415	0.2618	0.3849	0.4403	0.4462	-	113.59	310.36	260.71	348.72
		Chaco (SP)	0.471	0.5249	0.4769	0.3234	0.3169	0.5207	0.546	0.5555	0.5047	0.531	0.5432	0.2553	0.3845	0.4362	0.4443	0.0156^{NS}	-	375.71	359.25	258.67
		Salta	0.4774	0.5362	0.4872	0.3359	0.3289	0.5218	0.5464	0.5558	0.5056	0.5314	0.5435	0.2667	0.3952	0.4358	0.4484	0.0393^{NS}	0.0289^{NS}	-	434.15	448.18
		S. del E.	0.4596	0.5152	0.4701	0.3239	0.3175	0.5083	0.5341	0.5434	0.4929	0.518	0.5305	0.2578	0.3834	0.428	0.4431	0.0138^{NS}	0.0027^{NS}	0.0174^{NS}	-	609.05
		Formosa	0.4913	0.5293	0.4896	0.3438	0.3376	0.5298	0.5551	0.5641	0.5153	0.5373	0.548	0.272	0.3972	0.4465	0.4542	0.0274^{NS}	0.0424^{NS}	0.0682	0.0419^{NS}	-

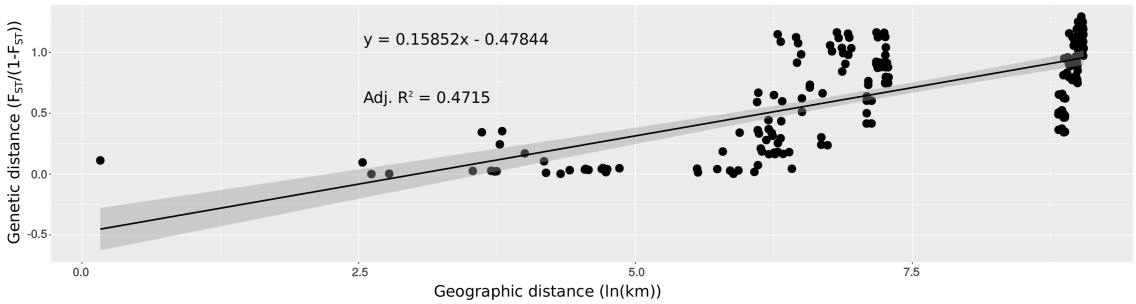

Figure 3. Genetic distance (values of F_{ST} from Table 2.3 adjusted to $F_{ST}/(1-F_{ST})$) plotted as a function of geographic distance (linear values from Table 2.3 adjusted by a natural log transformation). Inset shows the formula for the regression line and the adjusted R^2 value.

significantly correlated with geographic distance, and the adjusted R^2 value indicated that an assumption of IBD explained 47.15% of the variation in the dataset (Fig. 3). Despite the majority large pairwise values of F_{ST}, both the PCA and the population assignment probability test identified groups of collections that were more closely related to one another than they were to other collections (Figs. 4, 5). Though these tests did not agree exactly, both results were roughly consistent with what would be expected under an IBD model. Principal components 1 and 2 together explained 44.97% of the observed variation in the dataset (Fig. 4). The population assignment probability test was optimized at K=6 under a maximum likelihood model and at K=13 using the Puechmaille (2016) method (Fig. 5). The observed patterns of genotypic assignment at K=6 and K=13 are similar, except that the 2016 Tamaulipas collection was assigned to the same genotypic group as other collections when K=6, but assigned to a unique genotypic group when K=13. In both analyses, the Argentine weevil collections clustered tightly with each other and were distinct as a group from all other sampled

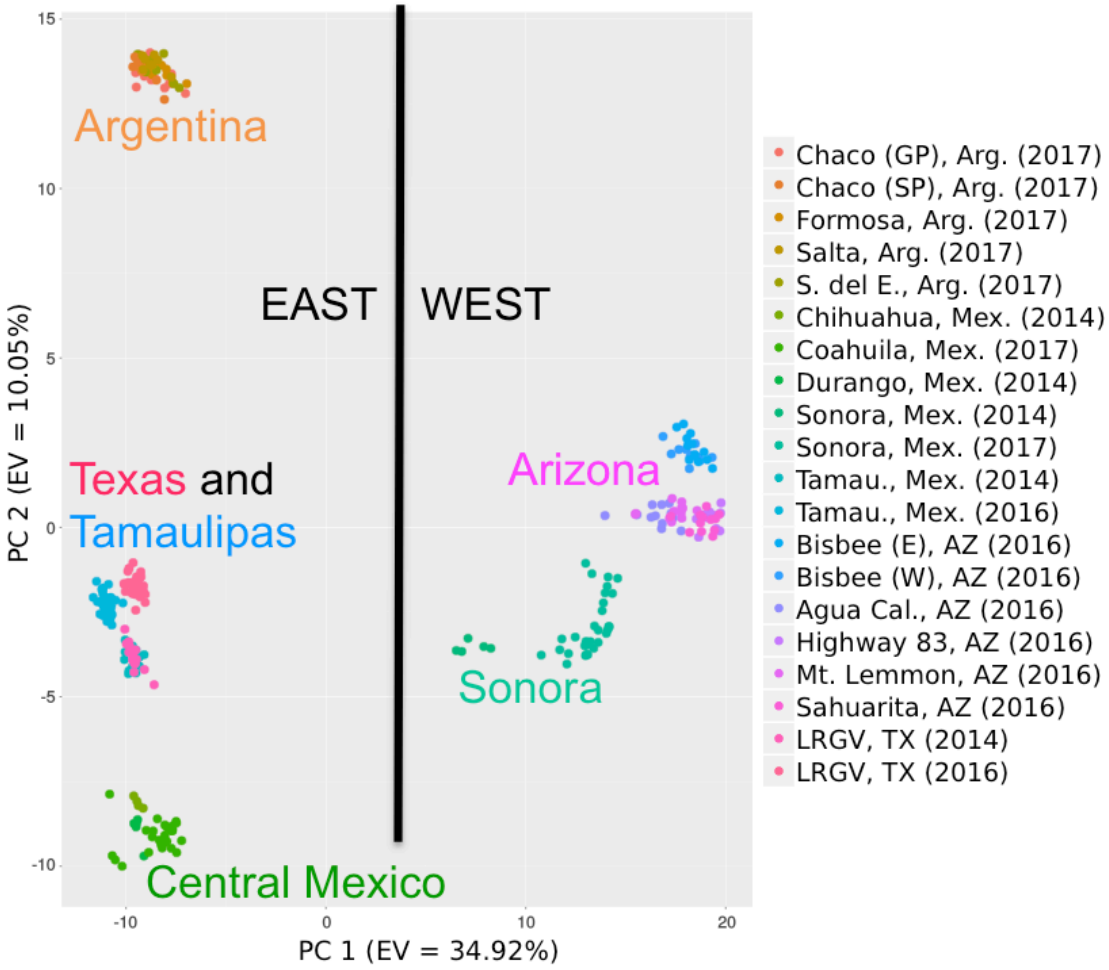

Figure 4. Principal component 2 (PC 2) plotted as a function of principal component PC 1. Eigenvalues (EV) indicate the proportion of the observed variation that is explained by the corresponding PC. PC 1 and PC 2 together explain 44.97% of the total variance. Black bar indicates separation of clusters associated with eastern and western lineages along PC 1 axis.

weevil collections. Arizona weevil collections clustered together in the PCA, but the Bisbee collections clustered separately from the more western collections. Two non-overlapping clusters representing the 2014 and 2016 Sonora collections were much more similar to the Arizona weevils than they were to other collections. Individuals associated

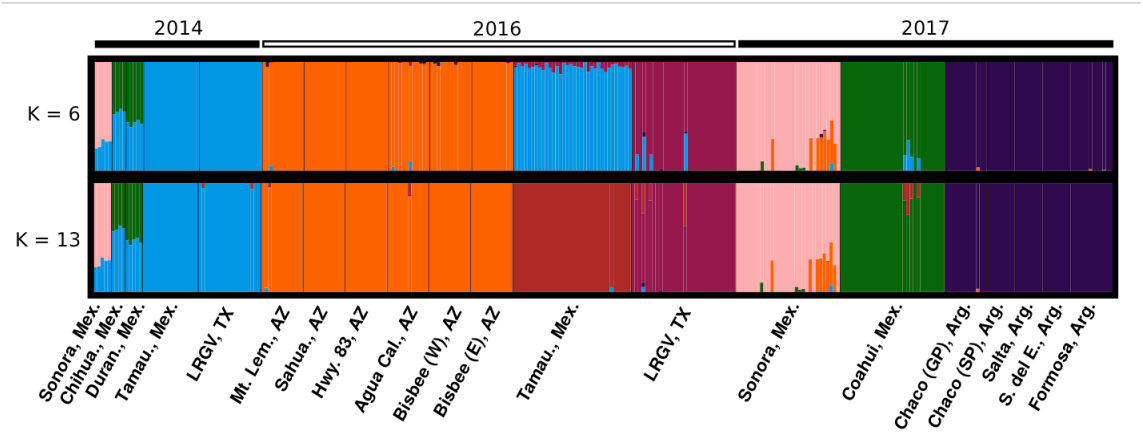

Figure 5. Results of the population assignment probability test visualized as distruct plots by CLUMPAK via StructureSelector. Bar plots are shown for two optimal values of K; using a maximum likelihood framework (6) and using the Puechmaille (2016) method (13). Each bar represents a single individual and each color represents 1 of K genotypic groups. The proportion of an individual's bar that is a certain color represents the probability that that individual belongs to that genotypic group.

with both Sonora collections were assigned to the same unique genotypic group with high probability, but 2016 samples displayed a low probability of assignment to thesame genotypic group as the Arizona weevils, whereas some 2014 samples displayed a low probability of assignment to a different genotypic group. All central Mexico collections (Chihuahua and Durango in 2014 and Coahuila in 2017) clustered together in the PCA. In the assignment test, the Coahuila collection was assigned to its own genotypic group and the Chihuahua and Durango collections had a moderate probability of assignment to that same group. However, the Chihuahua and Durango collections also had a moderate probability of assignment to another genotypic group to which the 2014 LRGV (Texas and Tamaulipas) samples were assigned. The results of the PCA and population assignment probability test were in least agreement with regards to the 2016 Texas

collection; in the assignment test, individuals were predominantly assigned to a unique genotypic group, but those same individuals clustered closely with the 2014 Texas and Tamaulipas collections and the 2016 Tamaulipas collection in the PCA.

Within-population genetic summary statistics varied greatly across space and time, but all collections exhibited a signal of outcrossing (Table 2.1). Population average values of genetic diversity ($1\text{-}Q_{intra}$ and $1\text{-}Q_{inter}$) were greatest in Sonoran collections and lowest in Argentine boll weevil collections. The inbreeding coefficient (F_{IS}) was negative for all collections, indicating heterozygote excess. F_{IS} was most negative in Argentine collections, and closest to zero in Texas and Tamaulipas collections. Geographic localities sampled in multiple years (Texas, Tamaulipas, Durango/Coahuila, and Sonora) did not necessarily maintain similar values of genetic diversity or F_{IS} across the time period between sampling dates. The Texas collections, in particular, yielded greatly different values of F_{IS} when comparing 2014 to 2016.

2.3.2. Phylogenetic Reconstruction

Of the four tested models, the tree with the highest log likelihood was constructed using the GTR+γ+I model and there was no ascertainment bias found in the data. The unrooted RAxML tree showed a topology with two major clades separated by a long branch (Fig. 6). One clade, hereafter referred to as the western lineage, consisted of the Arizona and Sonora weevils (all collections) and the other, the eastern lineage, consisted of all other collections. The tree was midpoint rooted along the long branch separating the two lineages. Both lineages were strongly supported with bootstrap support values of 100. Other major clades with strong bootstrap support were the monophyletic Arizona group,

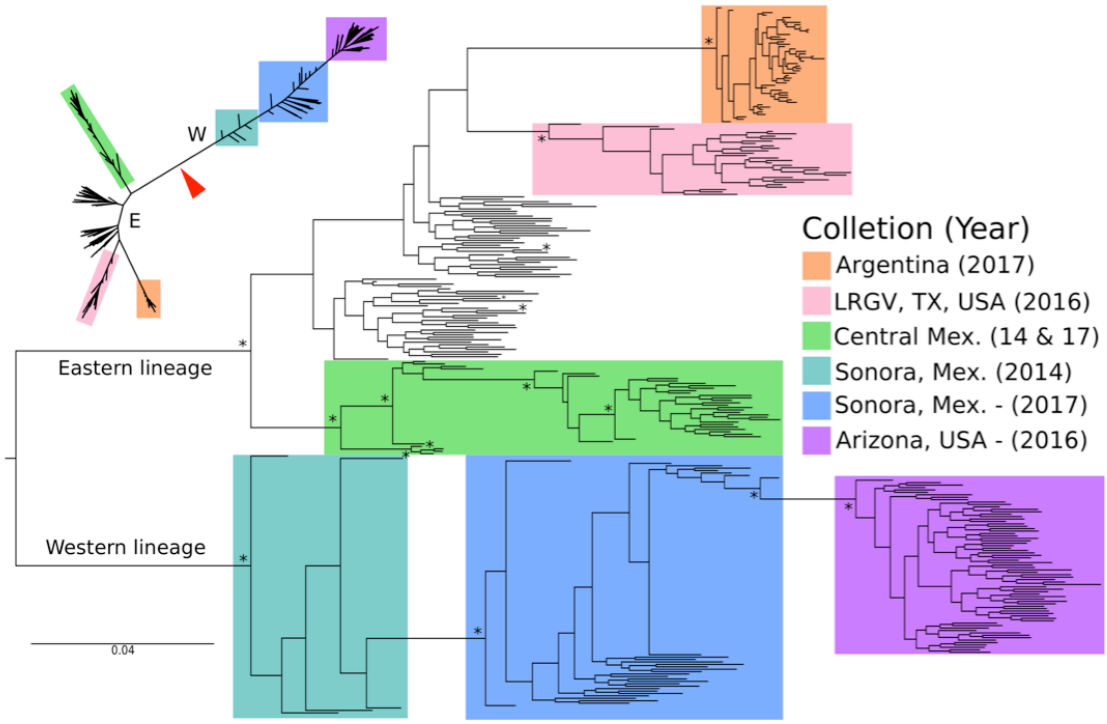

Figure 6. Midpoint-rooted RAxML phylogenetic reconstruction of all 292 sampled weevils. Asterisks indicate nodes with ≥95% bootstrap support. Highlighted groups are monophyletic groups (though some are nested within other groups) that have strong bootstrap support and that correspond with one or more collections in the study. Inset (top left) shows the unrooted tree with red arrow indicating the long branch upon which the tree was midpoint-rooted. Western (W) and eastern (E) lineages are denoted on both the unrooted and midpoint-rooted tree.

which was nested within the western lineage, and the monophyletic central Mexico group, which was found to be sister to a weakly supported group consisting of all other individuals in the eastern lineage. The Argentine weevils and the 2016 Texas weevils each formed monophyletic groups with strong support that were sister to each other and nested within the eastern lineage. Most other clades did not show strong bootstrap support, and even within the strongly supported groups, individual relationships did not show strong support.

2.4. Discussion

This study was the first investigation of boll weevil population genetic structure using a population genomic approach that sampled thousands of informative SNP markers from across the species' entire genome. It was also only the second population genetic study to include boll weevils from collections in both North and South America. Overall, our results indicated that boll weevil populations were genetically diverse and highly structured on a broad geographic scale. While the observed values of F_{ST} (Table 2.3) would be considered quite high for some systems, such numbers are consistent with previous measurements made for boll weevil populations, with values as high as 0.5 being recovered from a variety of different traditional population genetic markers (Scataglini et al., 2000; Kim and Sappington, 2006; Alvarado et al., 2017). Nonetheless, the high F_{ST} values calculated here indicating genetic divergence were seemingly contradictory to the negative values of F_{IS} (Table 2.1), which indicated outcrossing populations. Some of this discrepancy could be explained by the geography of our population sampling; though the observed genetic structuring was generally consistent with an IBD model (Fig. 3), there were large geographic gaps (Fig. 2), which may have been genetic populations that were intermediate to pairs of populations we studied but were not sampled. In addition to missing populations of weevils associated with commercial cotton, another explanation may be that the observed geographic structure exists and that our sampling scheme excluded boll weevil populations associated with wild cotton or non-cotton hosts. In addition to other species of *Gossypium*, boll weevil has been documented in association with other Malvaceae including *Hampea* spp.,

Cienfuegosia spp., and *Thespia populnea* (Cross, 1973; Cross et al., 1975; Burke et al., 1986). Since our samples were mostly obtained from areas of commercial cotton production, we may have neglected populations associated with these alternative hosts that may exchange gene flow with cotton-associated populations of boll weevil, allowing for the observed geographic structure but creating a signal of outcrossing within the sampled populations due to incomplete sampling.

2.4.1. Phylogeography and Implications for Taxonomy

Our results supported the two-form hypothesis of boll weevil variants, and we support the taxonomic status of the eastern (*A. g. grandis*) and western (*A. g. thurberiae*) forms of boll weevil in North America as distinct genetic subspecies. This conclusion is supported by the PCA and phylogenetic reconstruction. Principal component 1 (PC 1), which explained nearly 35% of the observed variation in the dataset, clearly separated individuals into eastern and western clusters (Fig. 4). Though there was some variation along the PC 1 axis between the western clusters consisting of Sonora and Arizona collections, there was a much larger gap between those clusters and the individuals associated with the eastern clusters. Nearly all of the variation between eastern clusters occurred along the PC 2 axis. The phylogenetic reconstruction provided more strong support for the existence of two distinct genetic lineages (Fig. 6). The longest branch of the unrooted tree separated the Sonora and Arizona collections from all other collections, and when midpoint-rooted along that long branch, there was strong bootstrap support for both the eastern and western lineages, which were reciprocally monophyletic.

Though *A. g. thurberiae* has historically been regarded as a host-associated variant, our study adds to the growing body of evidence that *A. g. thurberiae* is more likely to be a geographic variant. We recovered a close genetic relationship between western boll weevil populations collected from commercial cotton in Sonora and populations collected from wild cotton in Arizona. Though there was some significant genetic differentiation between these groups, the observed difference could be attributed to IBD (Fig. 3). Other recent studies have also recovered this same relationship and shown that the subspecies currently recognized as *A. g. thurberiae* can be found in association with other *Gossypium* species besides *G. hirsutum* (Kim and Sappington, 2006; Kuester et al., 2012; Alvarado et al., 2017). Additionally, weevils collected from *G. hirsutum* that have been morphologically identified as *A. g. grandis* have been genetically identified as *A. g. thurberiae*, and morphological characteristics are notoriously unreliable for variant assignment in this species (Roehrdanz, 2001; Barr et al., 2013). Our data supported the hypothesis that the two major boll weevil lineages are more likely to be the product of geographic isolation rather than the product of any host-associated differentiation. Thus, in our support for the two-form hypothesis, we stress that the two subspecies be regarded as geographic variants of boll weevil that are independent of host plant associations.

We found significant geographic substructure within the eastern boll weevil lineage (*A. g. grandis*). In particular, collections from north-central Mexico (Chihuahua, Durango, and Coahuila) that clustered together in the PCA and formed a strongly supported monophyletic group nested within the eastern lineage should be given a closer look. The Sierra Madre Occidental mountain range has been previously described as a likely

geographic barrier to gene flow that gave rise to the western and eastern boll weevil variants (Kuester et al., 2012; Alvarado et al., 2017). Here, we suggest that the Sierra Madre Oriental mountain range may also act as a geographic barrier that inhibits gene flow between boll weevil populations in central Mexico and the other eastern populations (Fig. 2A). The collections from central Mexico were found to be sister to all other collections within the eastern lineage, but the genetic distance between them is not as great as the distance between the western and eastern boll weevil, so it would be premature to suggest elevation of this group as a third subspecies. The reduced relative genetic distances suggest that the isolating effect of the Sierra Madre Oriental topography may not be as strong as it is for the Occidental range in the west, but it may also suggest that the divergence is simply more recent. Nonetheless, we support the subspecies status of the eastern boll weevil as *A. g. grandis* with the caveat that the central Mexico boll weevil may represent an incipient divergence.

Consistent with Scataglini et al. (2006), we found that Argentine boll weevils were more closely related to the eastern boll weevil lineage in North America as opposed to the western boll weevil. Low genetic diversity within and among the Argentine populations suggests a likely historical bottleneck, consistent with a single introduction, but high genetic distances and the large gap in our geographic sampling between Mexico and Argentina made it impossible to ascertain the source of the introduction. In terms of genetic distance, the Argentine populations were as distant from the eastern boll weevil as the eastern boll weevil was from the western variant in North America (Table 2.3). Since we did not have representative sampling between

Mexico and Argentina, it was impossible to ascertain whether intermediate genotypes existed between the eastern boll weevil in North America and the Argentine boll weevil. Importantly, the genetic isolation observed between the eastern variant in North America and the Argentine weevils sampled here was consistent with what is expected from IBD under the assumption that introgressing populations exist between them in Central America and northern South America. More extensive geographic sampling of boll weevil populations across the species' distribution will be critical to fully understanding the biology and genomic impact of the range expansion.

We identified temporal changes in genetic variation between populations collected from the same geographic location in multiple years. Though such an observation could be attributed to a possible sequencing batch effect, we believe that there was sufficient evidence to rule out this confounding factor. First, we specifically tested the year of sequencing as a source of genetic variation, and it was found to be not significant (Table 2.2). Second, if year effects were due to sequencing error, then we would expect to have seen an even distribution of error across the dataset; we did not. In central Mexico, where collections were made in 2014 and 2017, we saw no effect of time in the PCA, and the individuals formed a monophyletic group in the phylogenetic reconstruction. On the opposite end of the spectrum, in Sonora, which was also sampled in 2014 and 2017, we saw complete separation of PCA clusters. Genetic changes between sampling years were also observed in Texas and Tamaulipas between 2014 and 2016. These changes were particularly intriguing, because though the genetic distances between collections were relatively small, individuals were assigned with high

probability to completely different cluster in the population assignment test (Fig. 5). Similar temporal changes in genetic variation have been previously documented on a similar time scale in boll weevil populations in parts of Texas and Mexico using microsatellites (Choi et al., 2011), so changes in populations are not entirely unexpected over time. As management and eradication programs reduce populations in commercial fields, there is a possibility that the weevil populations undergo strong bottlenecks or complete local extirpation. In case of the later, individuals developing on wild or volunteer plants may eventually recolonize commercial fields thereby creating a founder effect.

2.4.2. Considerations for Management in the US and Mexico

Our results provided strong evidence supporting the two-form hypothesis of boll weevil variants wherein there exist two geographic subspecies: *A. g. thurberiae* in the west and *A. g. grandis* in the east. It will be critical to management and eradication in the US and Mexico to consider the biology of these variants when applying control measures. A primary concern for boll weevil management west of the Sierra Madre Occidental mountain range in Mexico and the western US should be to recognize that populations of boll weevil associated with wild hosts, namely *G. thurberi*, can likely act as a source for re-infestations of commercial cotton. Management and ongoing eradication efforts involving populations of the eastern boll weevil must acknowledge the contiguousness of boll weevil populations along the US-Mexico border, and that effective management will require a coordinated international effort to successfully combat the pest. Perhaps most critical for managers in both regions is to recognize the rapid rate of evolution

observed in populations of the boll weevil. Our results along with those of Choi et al. (2011) have demonstrated that rapid turnover of local genotypes can occur within a few years. Management efforts themselves are likely contributing to bottleneck and founder effects that help explain these turnovers, but allele frequency changes due to gene flow from populations that have not yet been sampled cannot be ruled out. It is thus essential that populations associated with wild hosts, volunteer cotton, and commercial cotton are monitored routinely and genetically characterized by standardized methods so that managers can better coordinate efforts and prepare against possible re-infestation events.

3. POPULATION GENOMICS TO IDENTIFY GENETIC VARIANTS AND DETERMINE PROBABLE SOURCE POPULATIONS FOR BOLL WEEVILS *ANTHONOMUS GRANDIS* BOHEMAN (COLEOPTERA: CURCULIONIDAE) FOUND IN PREVIOUSLY ERADICATED AREAS IN THE UNITED STATES

3.1. Introduction

The boll weevil (*Anthonomus grandis* Boheman, Coleoptera: Curculionidae) is a major pest of commercially cultivated upland cotton, *Gossypium hirsutum* L. (Malvaceae) in North and South America. It is known as one of the most devastating pests in United States (US) agricultural history (Cross, 1973; Lange et al., 2009). The US Boll Weevil Eradication Program has been widely successful, eliminating the boll weevil from the near entirety of the US cotton belt, but the threat of re-infestations to areas where eradication has been completed remains (National Cotton Council, 2010). The southeastern boll weevil variant (*A. g. grandis*) is a persistent problem for growers in the Lower Rio Grande Valley (LRGV) region of Texas (TX) along the US-Mexico border where eradication efforts are ongoing (Texas Boll Weevil Eradication Foundation, 2018). In the western US, the Thurberia weevil (*A. g. thurberiae*), has historically been regarded as a host plant-associated variant that utilizes Arizona wild cotton, *Gossypium thurberi* Todaro (Malvaceae). However, mounting evidence suggests that *A. g. thurberiae* is a geographic lineage rather than a host-associated lineage, and populations of the subspecies may serve as sources for re-infestations of commercial cotton in the

southwestern US and northwestern Mexico (Kuseter et al., 2012; Alvarado et al., 2017, Raszick et al., in prep.)

Recognizing source-sink dynamics can be critical to insect management, and especially eradication efforts. If possible, identification and proactive control of source populations should help prevent the establishment of, and therefore need to control, sink populations. On the other hand, failure to recognize source-sink dynamics can result in the re-infestation of previously controlled areas that repeatedly require reactive control (Hanski and Gilpin, 1991; Harrison, 1991; Zaller et al., 2008; Sword et al., 2010; Carrière et al., 2012). The risk of re-infestation and the value of being able to identify potential source populations for ongoing boll weevil eradication efforts in the US are highlighted by two recent case studies. In 2015, boll weevils were recovered from pheromone-baited cone traps maintained by the Texas Boll Weevil Eradication Foundation in Uvalde, TX, roughly 100 km from the US-Mexico border. The so-called Winter Garden production zone had been weevil free for three years prior to 2015, but by 2016, the re-infestation had spread to nearby Batesville, TX, and a total of 15,714 weevils were captured in the Winter Garden growing area that year (Texas Boll Weevil Eradication Foundation, 2018). In 2017 and 2018 (as of completion of this study), 1,292 and 6 weevils were captured respectively (Texas Boll Weevil Eradication Foundation, 2018). Though the area appears to be trending back towards eradication, managing the outbreak was a four-year effort that cost an estimated $5 million. Considering the high risk and economic impact of boll weevil re-infestation events, it is critical to management that diagnostic assays of trapped boll weevils be able to accurately

diagnose trapped specimens as belonging to one of the two geographic *A. grandis* subspecies, *A. g. grandis* and the *A. g. thurberiae,* and ideally determine a likely source population. Current methods of boll weevil variant diagnosis include calculating the morphometric ratio of the anterior profemora, and a mitochondrial DNA haplotype assay based on the cytochrome *c* oxidase subunit I gene (COI) gene, both of which are currently utilized by the US Department of Agriculture for taxonomic determination (Warner, 1966; Burke et al., 1986; Barr et al., 2013). However, both of these methods of variant assignment have limitations. The morphometric assay is limited by the occurrence of intermediate profemora ratios that could indicate either variant. The COI haplotype assay is generally effective for assigning individuals to one of the two geographic lineages, but at least one haplotype, denoted as AN4, is recovered from both *A. g. grandis* and the *A. g. thurberiae*, making it impossible to ascertain the source lineage of individuals with that haplotype.

Raszick et al. (in prep.) recently described two major geographic lineages of boll weevil in North America using a dataset of single nucleotide polymorphisms (SNPs) generated by double digest restriction site-associated DNA sequencing (ddRADseq, Peterson et al., 2012). That study revealed that, based on 7,177 SNPs, there was a western lineage that could be considered as analogous to the *A. g. thurberiae* subspecies and an eastern lineage that could be considered as analogous to the *A. g. grandis* subspecies. Importantly, individuals assigned to the *A. g. thurberiae* subspecies were collected from both *G. hirsutum* and from *G. thurberi*, suggesting that both commercial cotton and wild host plants could support populations that contribute to re-infestations,

consistent with other recent research (Kuester et al., 2012; Alvarado et al., 2017). In this study, we applied the Raszick et al. (in prep.) methodology to 56 weevils collected from the Winter Garden re-infestation and to 24 weevils collected from cotton fields in previously eradicated Hidalgo County, New Mexico. For all 80 specimens, we generated SNP data homologous to the 7,177 loci used to delineate boll weevil populations across its range by Raszick et al. (in prep). We then compared the genotypes of the 56 Winter Garden weevils to those of the 292 weevils used in the Raszick et al. (in prep) study in order to determine the likely source population for the re-infestation, and we used the 24 Hidalgo Co., NM weevils to evaluate the performance of the SNP genotyping as a taxonomic diagnostic assay relative to the morphometric and COI assays.

3.2. Materials and Methods

3.2.1. Specimen Sampling

We sequenced a total of 56 adult boll weevil specimens from the Winter Garden, TX re-infestation and 24 from Hidalgo Co., NM (Table 3.1). All samples were collected from cotton production areas using pheromone-baited cone traps that are typical for monitoring for boll weevils in eradicated areas of the US (Cross and Hardee, 1968; Cross et al., 1969b; Tumlinson et al., 1969; Hardee et al., 1971). Live specimens were transferred from the traps into vials with 95-100% molecular grade ethanol (EtOH) and transported to the laboratory where they were stored at -80°C until they were prepared for DNA isolation. The Winter Garden, TX re-infestation was sampled in 2016 and 2017, and Hidalgo Co., NM was sampled in 2017. Since most individual cone traps contained only one or a few specimens, boll weevils from traps in close proximity were

Table 3.1. Collection information and basic genetic summary statistics. Dashes indicate missing information. GPS coordinates for Uvalde collections are midpoints calculated from individual trap GPS coordinates. N was the actual number of individuals genotyped and N_e was the effective population size as estimated by NeEstimator. ∞ symbol indicates an estimate of N_e that is not significantly different from an infinitely large population. 1-Q values are average multilocus genetic diversity calculated within individuals within a collection (intra) and among individuals within a collection (inter). F_{IS} is the inbreeding coefficient.

City/Co.	State	Latitude	Longitude	Date Collected	Year	N	N_e	$1\text{-}Q_{intra}$	$1\text{-}Q_{inter}$	F_{IS}
Batesville	TX	-	-	-	2016	30	91.2	0.1283	0.1056	-0.2147
Uvalde	TX	29.180169	-99.60665	Sep 19 - Sep 28	2016	16	11.1	0.1533	0.1382	-0.1095
Uvalde	TX	29.107374	-99.646694	Apr 4 - May 15	2017	10	∞	0.1358	0.1107	-0.2267
Hidalgo Co.	NM	-	-	Sep 22 - Oct 13	2017	24	103.9	0.1629	0.139	-0.172

pooled, and GPS coordinates reported here are the geographic midpoints as calculated from the GPS coordinates of the individual traps.

3.2.2. DNA Isolation, Library Preparation, and Sequencing for SNP Genotype Assay

Specimens from Hidalgo Co., NM had a leg removed for the Barr et al. (2013) COI haplotype assay prior to the DNA isolation for the SNP genotype assay. To isolate high molecular weight DNA for the SNP assay, all specimens were mechanically disrupted by crushing in liquid nitrogen with disposable pestles. Isolated DNA was extracted from the crushed specimens using the Qiagen Gentra Puregene Cell and Tissue Kit using a modified version of the manufacturer protocol (Appendix A). The molecular weight of isolated DNA was tested by electrophoresis on a 1.5% agarose gel. We considered electrophoresis bands of fragment length greater than 10,000 base pairs (bp) to be of sufficiently high molecular weight. Further purification, library preparation, and sequencing were carried out at Texas A&M AgriLife Genomics and Bioinformatics

Service (TxGen). The Agencourt AMPure XP purification system was used to further purify isolated DNA prior to double digest restriction site-associated DNA sequencing (ddRADseq) library preparation. Libraries were prepared using the NlaIII and HindIII restriction enzymes and selecting for fragment sizes ranging 250-500 bp, consistent with the methods of Raszick et al. (in prep). Size-selected fragments were then ligated with standard Illumina adapters, multiplexing indexes, and sequencing primers, except for a proprietary R1 primer that contained the 5' restriction site remnant. Finally, libraries were sequenced on an Illumina NovaSeq using 150x150 sequencing cycles.

3.2.3. Raw Sequence Processing and Integration with Existing SNP Dataset

TxGen provided demultiplexed raw reads and FastQC version 0.11.3 reports for each specimen to the authors (Andrews, 2010). FastQC reports were manually reviewed by the authors to ensure that sequences were of significantly high quality to move forward with further analyses. To filter out any potential bacterial contamination, we used Kraken version 1.1 to match sequences to the non-redundant bacterial database hosted by the National Center for Biotechnology Information (NCBI) and remove them (Wood and Salzberg, 2014). SNP-calling was carried out using dDocent version 2.5.5 (Puritz et al., 2014a; Puritz et al., 2014b). The dataset of 80 individuals was combined with the previously described dataset of 292 individuals from Raszick et al. (in prep.), and dDocent was run on all 372 individuals to ensure homology of the SNP loci. dDocent was run using a percent clustering similarity of 90% and other parameters as defaults, except that the pipeline was paused after initial trimming, and Trimmomatic version 0.38 was used to further trim sequences to a uniform length of 90 bp (Bolger et al., 2014).

dDocent was then restarted and allowed to finish normally. VCFtools version 0.1.15 was used to filter the dDocent output following the tutorial in the dDocent user guide except that in addition to the tutorial filters, SNP loci that were not present in 100% of individuals were removed (Danecek et al., 2011).

3.2.4. Population Genetic Analyses

All necessary file conversions for population genetics analyses were carried out using PGDSpider version 2.1.1.3 (Lischer and Excoffier, 2012). We used a variety of packages in RStudio version 1.1.456 to generate population genetic summary statistics for each collection and determine their relationships to the populations described in Raszick et al. (in prep.) and to each other (R Core Team, 2018). R/vcfR version 1.8.0 was used to read the filtered vcf file and prepare objects for use with other packages (Knaus and Grünwald, 2017). R/genepop version 1.0.5 was used to calculate genetic diversity based on allele identity and the inbreeding coefficient (F_{IS}) for each collection (Raymond and Rousset, 1995; Rousset, 2008). Genotypic diversity was measured as a multilocus average within individuals ($1\text{-}Q_{intra}$) and among individuals within each collection ($1\text{-}Q_{inter}$). We used NeEstimator version 2.1 to estimate the effective population size (N_e) for each of the collections using a lower rare allele frequency limit of 0.02 (Do et al., 2014).

To estimate gene flow between our collections and the Raszick et al. (in prep.) populations, we calculated pairwise values of Weir and Cockerham (1984) F_{ST} using R/genepop. The global F_{ST} of the data was also calculated. To test the significance of any observed genetic differentiation, we conducted pairwise exact conditional

contingency-table tests in R/genepop (dememorization = 1000, batches = 10, iterations = 500). We also carried out a principal components analysis (PCA) in R/adegenet version 2.1.1 in order to visualize clustering of our individuals with clusters of individuals from the genetic populations delineated in the Raszick et al. (2018) study (Jombart, 2008; Jombart and Ahmed, 2011). Since R/adegenet can only carry out a PCA with a biallelic dataset, 524 loci with more than two alleles were removed for this analysis.

We used the software fastSTRUCTURE version 1.0 to carry out a population assignment probability test wherein we calculated each individual's probability of assignment to one or more predetermined genotypic groups ($1 \leq K \leq 30$, Raj et al., 2014). The vcf file was formatted for fastStructure input using PLINK version 1.07 (Purcell et al., 2007). To choose the optimal value of K, we accessed the web browser-based program StructureSelector and implemented the Puechmaille (2016) method of K selection, which accounts for an uneven distribution of samples across populations (Li and Liu, 2018). As recommended by the original publication, we tested mean membership coefficient thresholds of 0.5, 0.6, 0.7, and 0.8. StructureSelector was also used to run the integrated version of CLUMPAK and generate a visual plot of individual population assignments (Kopelman et al., 2015).

3.2.5. Comparison of Assays for Variant Diagnoses using Hidalgo Co., NM Individuals

For the 24 individuals collected from Hidalgo Co., NM, we applied the COI haplotype assay described in Barr et al. (2013) in addition to the Raszick et al. (in prep.) SNP genotype assay. We also carried out morphological measurements to calculate the

morphometric ratios of the profemora. To briefly summarize the COI assay method, DNA was isolated from the aforementioned removed legs using the DNeasy Blood and Tissue Kit, and a roughly 700 bp fragment of the COI gene was amplified using polymerase chain reaction (PCR). See Barr et al. (2013) for PCR primers and thermocycling conditions. Amplification products were purified with ExoSAP-IT (Thermo Fisher, Waltham, MA, USA) before sequencing. Bi-directional sequencing of PCR products was performed using 3'BigDye-labeled dideoxynucleotide triphosphates (v 3.1 dye terminators, Applied Biosystems, Foster City, CA, USA) and run on an ABI 3730XL DNA Analyzer with the ABI Data Collection Program (v 2.0) at Functional Biosciences, Madison, WI. Individuals were each assigned to one of 31 different COI haplotypes. Measurements of profemora ratios were made following Burke (1968). If ratios ranged 3.0-3.4, the weevils were designated as *A. g. thurberiae*, and if ratios ranged 3.6-4.0, the weevils were designated as *A. g. grandis*. Ratios outside of these ranges were considered an inconclusive identification. These leg ratios have previously been found to correctly distinguish variants for 89.4% identifications (Barr et al., 2013). Putative taxonomic assignments were made based on the results of each assay and compared to assignments made using the SNP assay.

3.3. Results

3.3.1. Population Genetic Analyses

For all 80 specimens analyzed, we successfully recovered SNP data homologous to the 7,177 loci that were used to describe the boll weevil populations in Raszick et al. (in prep.). Multilocus genetic diversity levels within and among each of the four collections

Table 3.2. Pairwise and global F_{ST} values. Argentine, eastern, central Mexico, and Western populations are groupings of populations inferred from Raszick et al. (2018) and the re-infestation collections are those labeled with a collection year.

	Argen.	Eastern	Cen. Mex.	Western	Bat. (16)	Uva. (16)	Uva. (17)	Hid. (17)
Argentine	-							
Eastern	0.2649	-				Global F_{ST} 0.3593		
Central Mexico	0.4464	0.2165	-					
Western	0.4514	0.3787	0.4124	-				
Batesville, TX (2016)	0.425	0.1969	0.3522	0.4133	-			
Uvalde, TX (2016)	0.3622	0.135	0.3012	0.3384	0.0951	-		
Uvalde, TX (2017)	0.4237	0.1157	0.3144	0.4176	0.1307	0.0798	-	
Hidalgo Co., NM (2017)	0.5325	0.4179	0.4835	0.0283	0.4951	0.4007	0.4909	-

made for this study were similar to those obtained in the original Raszick et al. (in prep.) study, and we calculated negative F_{IS} values for all four collections, indicating heterozygote excess (Table 3.1). Estimates of effective population size ranged from 11.1 to 103.9, except for the estimate from the 2017 Uvalde, TX collection, which was not significantly different from an infinitely large population. Despite the signal of heterozygote excess, we nonetheless recovered high pairwise values of F_{ST} (Table 3.2). Additionally, the exact conditional contingency-table tests indicated that all pairwise comparisons were significantly different. The global F_{ST} for the dataset was 0.3593. The lowest pairwise values of F_{ST} were calculated when comparing the Hidalgo Co., NM collection to the western boll weevil lineage, when comparing the 2016 Batesville, TX collection to the 2016 Uvalde, TX collection, and when comparing the 2016 Uvalde, TX collection to the 2017 Uvalde, TX collection.

The PCA tightly clustered individuals from our collections with individuals described by Raszick et al. (in prep.) (Fig. 7). All individuals from the three Winter Garden collections clustered tightly together with each other and with the Raszick et al.

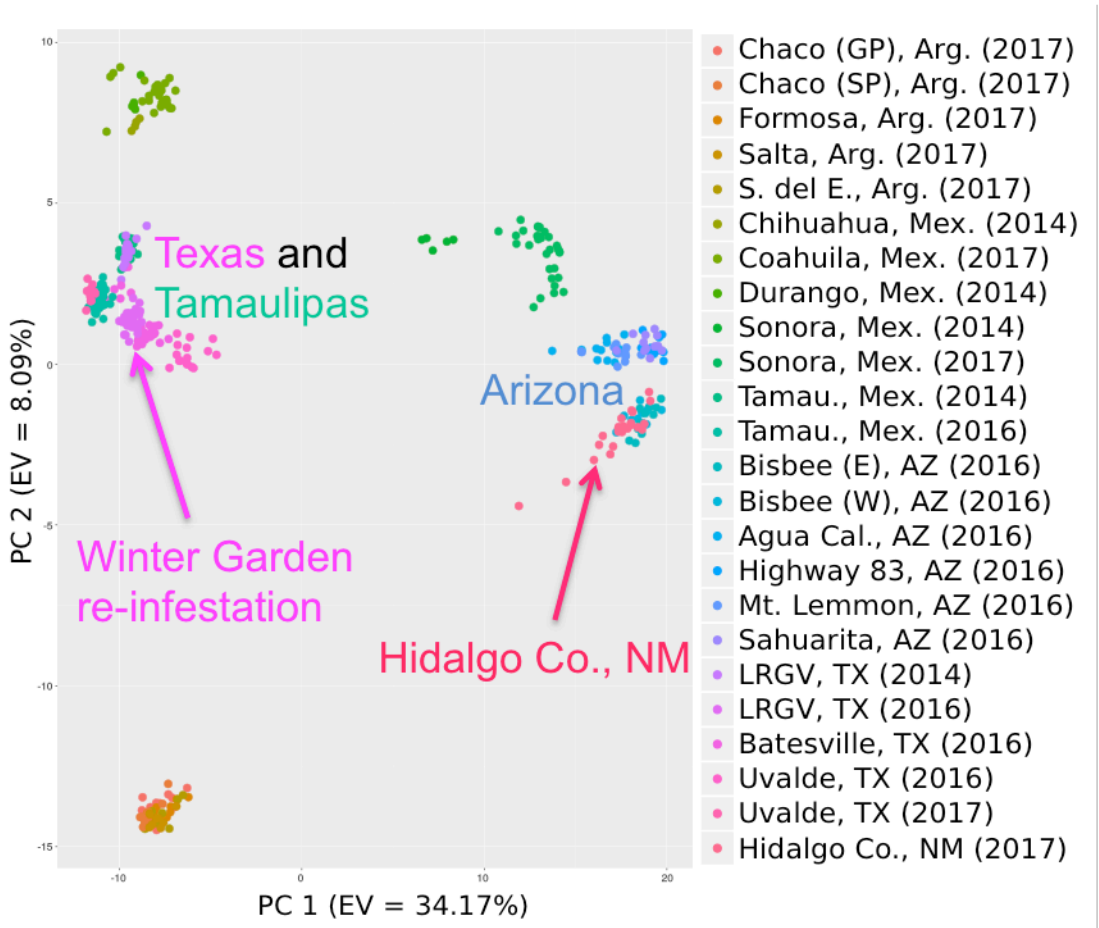

Figure 7. Results of principal components analysis presented as principal component 2 (PC 2) plotted as a function of principal component 1 (PC 1). Eigenvalues (EV) indicate the percent variation explained by the respective PCs. The Texas and Tamaulipas and Arizona groups are from Raszick et al. (in prep.)

(in prep.) individuals representing the eastern boll weevil lineage from the LGRV along the Texas-Mexico border. The Hidalgo Co., NM individuals clustered with individuals from the western boll weevil lineage, specifically those collected from Arizona wild cotton.

For the population assignment probability test (Fig. 8), the Puechmaille (2016) method of K selection determined that the optimal value was K=13 for our dataset.

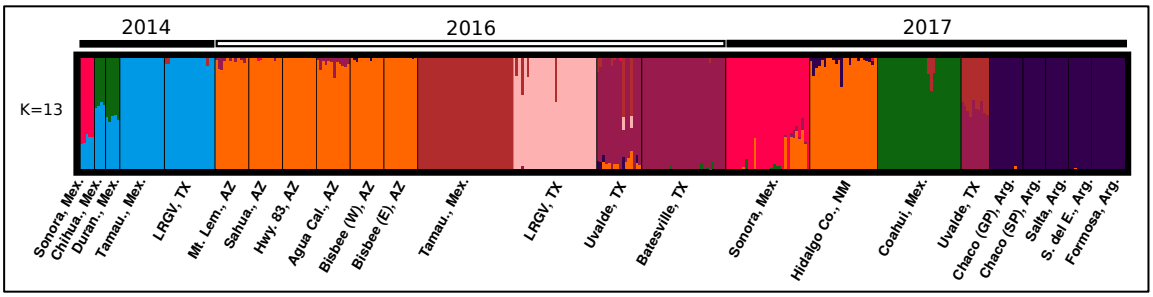

Figure 8. CLUMPAK output of population assignment probability test when K=13. Vertical bars are individuals and the different colors represent the different genotypic groups. The proportion of each individuals' bar that is a certain color indicates that individuals' probability of assignment to that genotypic group.

Individuals from Hidalgo Co., NM were assigned to the same genotypic group as the western boll weevil lineage with high probability. Individuals from the Winter Garden re-infestation were mostly assigned to their own genotypic group, though two individuals from the 2016 Uvalde, TX collection were assigned with high probability to the eastern boll weevil lineage associated with the LRGV and Tamaulipas.

3.3.2. Variant Assignments from Assays for Variant Diagnoses Using Hidalgo Co., NM Individuals

Both the morphometric and COI haplotype assays returned some inconclusive variant assignments, and the methods did not necessarily agree (Table 3.3). The morphometric assay yielded some diagnoses for both geographic variants whereas the COI assay yielded some diagnoses for a single variant. Specifically, 12 individuals yielded the ambiguous AN4 haplotype, and the remaining 12 other individuals yielded 7 different haplotypes, all associated with *A. g. thurberiae*. As described above, the SNP genotype assay unambiguously clustered all 24 individuals from Hidalgo Co., NM with the

Table 3.3. Assignment of 24 Hidalgo Co., NM weevils to the *A. g. grandis* (BW) or *A. g. thurberiae* (TW) taxonomic group based on the Barr et al. (2013) mitochondrial haplotype assay.

Collection Information		Mitochondrial Assay		Morphometric Assay	
Specimen	Collection Date	COI Haplotype	Assignment	Profemoral Ratio	Assignment
1NM-07-03	Sep 22-24, 2017	AN8	TW	3.22	TW
1NM-07-17	October 3, 2017	AN17	TW	3.28	TW
1NM-02-03	October 3, 2017	AN20	TW	3.77	BW
1NM-02-04	October 3, 2017	AN2	TW	3.79	BW
2NM-1- F1	October 6, 2017	AN3	TW	3.12	TW
2NM-1- H1	October 6, 2017	AN20	TW	3.35	TW
2NM-1- H2	October 6, 2017	AN2	TW	3.71	BW
2NM-1- A6	October 6, 2017	AN4	Inconclusive	3.54	Inconclusive
2NM-1- C6	October 6, 2017	AN4	Inconclusive	3.14	TW
2NM-1- F11	October 6, 2017	AN4	Inconclusive	3.58	Inconclusive
2NM-1- C12	October 6, 2017	AN4	Inconclusive	3.56	Inconclusive
2NM-1- E12	October 6, 2017	AN4	Inconclusive	3.32	TW
2NM-2- F3	October 6, 2017	AN17	TW	3.80	BW
NM-171113-04	October 26, 2017	AN15	TW	3.38	TW
NM-171122-04	October 13, 2017	AN24	TW	3.40	TW
NM-171122-30	October 13, 2017	AN2	TW	3.39	TW
NM-171122-19	October 13, 2017	AN4	Inconclusive	3.49	Inconclusive
NM-171122-21	October 13, 2017	AN4	Inconclusive	3.37	TW
NM-171122-34	October 13, 2017	AN4	Inconclusive	3.39	TW
NM-171122-35	October 13, 2017	AN4	Inconclusive	3.57	Inconclusive
NM-171122-36	October 13, 2017	AN2	TW	3.84	BW
NM-171122-56	October 13, 2017	AN4	Inconclusive	3.50	Inconclusive
NM-171122-65	October 13, 2017	AN4	Inconclusive	3.39	TW
NM-171122-72	October 13, 2017	AN4	Inconclusive	3.26	TW

western boll weevil lineage in the PCA, and those individuals were also assigned to the same genotypic group as the western boll weevil in the population assignment probability test.

3.4. Discussion

3.4.1. Probable Source Population for Winter Garden Re-infestation

Our results showed that the recent re-infestation of the previously eradicated Winter Garden production area of Texas could be attributed to movement of *A. g. grandis* individuals from some area along the US-Mexico border in south Texas. Low pairwise

values of F_{ST} (Table 3.2) and tight clustering in the PCA (Fig. 7) clearly indicated that the 2016 and 2017 Uvalde, TX and 2016 Batesville, TX collections were all part of the same re-infestation population and that that population was very likely to have originated from the Texas LRGV or northern Tamaulipas. Yet, contradictorily, in the population assignment probability test, all three Uvalde, TX re-infestation collections were strongly assigned to their own genotypic group, distinct from all other groups (Fig. 8). Two key exceptions here were two individuals in the 2016 Uvalde, TX collection that were assigned to the 2016 Tamaulipas population with high probability. These two individuals, combined with the small N_e for the 2016 Uvalde, TX collection, together suggest that there may have been a founder effect that could explain the observed pattern. However, if this was the case, we would expect to see a positive F_{IS} for the 2017 Uvalde, TX collection, indicating inbreeding, but that was not observed. Furthermore, individuals from the 2017 Uvalde, TX collection had a higher probability of assignment to the eastern boll weevil lineage than did individuals from the 2016 Uvalde, TX collection, except for the two previously mentioned individuals. It should be noted that our sample size in 2017 was small, which may have led to unreliable estimates of N_e, genetic diversity, and F_{IS}. Thus, it remains unclear if the Winter Garden, TX re-infestation population underwent a population bottleneck due to a founder effect. Nonetheless, we can confidently assign the source population for this re-infestation as the LRGV and Tamaulipas population, based on the relatively low observed F_{ST} values and tight clustering in the PCA. However, we should note that the details of the demographic history of boll weevil in the Winter Garden area could not be fully

elucidated using this dataset. Furthermore, our results could not rule out the possibility of multiple introductions since the 2017 population seems to be more closely related to the LRGV and Tamaulipas than the 2016 population.

3.4.2. Performance of Assays for Variant Diagnoses Using Hidalgo Co., NM Individuals

We were able to confidently assign Hidalgo Co., NM individuals to the western boll weevil variant, and the SNP genotype assay outperformed both other assays for individual variant assignment. All Hidalgo Co., NM individuals exhibited SNP genotypes that matched closely with *A. g. thuberiae* populations sampled from Arizona wild cotton and were assigned to the same genotypic group as all individuals assigned to the western boll weevil lineage by Raszick et al. (in prep). The pairwise comparison of the western boll weevil lineage to the Hidalgo Co., NM individuals also displayed the lowest F_{ST} value in the dataset. Taken all together, these results clearly indicated that the Hidalgo Co., NM individuals are of a large contiguous population of weevils in the Arizona-Sonora border area.

Of the 24 sampled individuals from the Hidalgo Co., NM re-infestation, the COI haplotype assay provided an unambiguous assignment of 12 individuals to the *A. g. thurberiae* taxonomic group. The other 50% of individuals exhibited the AN4 genotype, which is known to be found in both *A. g. grandis* and *A. g. thurberiae*. The morphometric assay performed arguably worse; though there were fewer ambiguous assignments, 5 individuals were unambiguously and incorrectly assigned to the *A. g. grandis* taxonomic group. The SNP genotype assay, on the other hand, was able to

successfully assign 100% of individuals to the *A. g. thurberiae* western boll weevil lineage as inferred by Raszick et al. (in prep) and was even able to cluster them with the Arizona populations as opposed to the southern Sonora population described in that same study (Fig. 7, Fig. 8). Thus, the SNP genotype assay performed better in that it was able to assign 100% of individuals to a subspecies whereas the COI and morphometric assays were only able to correctly assign roughly 50% of the sampled individuals. Furthermore, the SNP genotype assay was able to assign individuals to a specific geographic population within the broader taxonomic group.

4. MONTHLY AND YEARLY TURNOVER OF GENOTYPES IN LOCAL POPULATIONS OF THE COTTON FLEAHOPPER, *PSEUDATOMOSCELIS SERIATUS* (REUTER) (HEMIPTERA: MIRIDAE)

4.1. Introduction

The cotton fleahopper, *Pseudatomoscelis seriatus* (Reuter) (Hemiptera: Miridae), feeds on over 100 known host plants across its range, and can be a pest of commercially cultivated upland cotton, *Gossypium hirsutum* L. (Malvales: Malvceae) (Reinhard, 1926; Reinhard, 1929; Hixson, 1941; Ring et al., 1993; Esquivel and Esquivel, 2009). Nymph and adult stages of *P. seriatus* attack small developing cotton pre-floral buds, referred to as squares, causing their abscission and yield losses (Ewing, 1929; Powell, 1979). Though there have been numerous studies that have investigated host plant use and dispersal in this species, only three have investigated genetic variation within and among populations. Barman et al. (2012) and Antwi et al. (2015) provided limited evidence of host-associated genetic differentiation (HAD) in some populations, but a full understanding of host plant use and genetic differentiation in *P. seriatus* has been elusive. In particular, it remains unclear if a genetically differentiated host-race that is associated with cultivated cotton exists. Resolution of this uncertainty is essential for integrated pest management (IPM) because pest population dynamics, movement, and gene flow, play critical roles in the evolution of resistance to pesticides and transgenic crops (Georghiou and Taylor, 1977; Guse et al., 2002; Onstad et al., 2002).

Understanding gene flow between potentially divergent populations with different host plant associations is critical for IPM because new control methods, such as transgenic crop varietals, continue to be developed which rely on a refuge strategy to delay evolution of insect resistance (Gould, 1998; Tabashnik et al., 2004; Miehls, 2008; Tabashnik, 2008; Tabashnik et al., 2008). In the case of *P. seriatus*, a new transgenic cotton cultivar has been developed that is resistant to *Lygus* spp. plant bugs and may be resilient to other mirids such as *P. seriatus* (Bachman et al., 2017). If genotypes of *P. seriatus* that are associated with cotton exist and do not exchange gene flow with genotypes associated with other host plants, then the cotton-associated genotypes could potentially develop resistance to the transgenic cotton, as would be expected with any new transgenic or plant systemic control (Gould, 1998; Caprio, 2001). Thus, it is essential to understand if a natural refuge exists wherein populations exposed to the control can exchange gene flow with populations that are associated with alternative wild host plant species and are therefore not subjected to selection for resistance. Of particular interest is the role of woolly croton, *Croton* spp. (Malpighiales: Euphorbiaceae), as an overwintering host plant and possible site of admixture for populations of *P. seriatus* that are potentially associated with other host plants earlier in the season. Since commercial cotton is only available as a host plant during part of the year, populations of *P. seriatus* found on cotton must both shift from alternative host plants onto cotton at the beginning of the growing season and back to alternative hosts at the end of the growing season. It is currently unknown whether these host shifts are associated with gene flow between cotton-associated populations and populations

associated with other host plants. Since woolly croton is the primary overwintering host plant for *P. seriatus* in areas where it occurs, we hypothesized that woolly croton serves as a year-end site of admixture between populations coming from different host plants, including cultivated cotton (Hixson, 1941; Holtzer and Sterling, 1980).

We used a high-throughput sequencing (HTS) technique called double digest restriction site-associated DNA sequencing (ddRADseq) to generate a large dataset of single nucleotide polymorphism (SNP) markers in order to resolve the population genomic structure of *P. seriatus* in the Brazos Valley, Texas, a geographic location in which HAD has not been observed (Peterson et al., 2012). Previous studies have primarily utilized amplified fragment length polymorphism (AFLP) methods to examine genetic variation in populations of *P. seriatus*. Such markers, while informative, are now readily outperformed by newer methods that take advantage of HTS technology. SNP datasets derived from ddRADseq can be an order of magnitude or more larger than AFLP datasets, and are thus more sensitive to very recent or very low levels of genetic divergence. We tested whether SNP-based population genomics could detect a signal of HAD that had not been possible to detect using AFLP markers, particularly with respect to *P. seriatus* populations reproducing on cultivated cotton. We assessed the monthly and yearly turnover of *P. seriatus* genotypes collected from commercially cultivated upland cotton and from nearby wild host plants. Both Barman et al. (2012) and Antwi et al. (2015) sampled their respective study locations only once per season. This method failed to consider the impact of variation in host plant phenology, resulting in a potential temporal bias. Though both studies acknowledged that host plant phenology likely plays

a large role in determining local genotypic observations, neither was designed to parse the effects of time on genotypic distributions in *P. seriatus* populations. These studies also genotyped field-collected adults rather than juveniles, and therefore could not disentangle potentially confounding effects of adult movement from other populations and host plants versus local reproduction on specific hosts. For our study, we intentionally sampled juveniles resulting from local reproduction at the same geographic locations multiple times over the course of two years in order to investigate the genotypic turnover at those sites.

4.2. Materials and Methods

4.2.1. Sample Acquisition

We sampled a total of 561 *P. seriatus* individuals across 2015 (N=200) and 2016 (N=361). Although we collected all developmental stages, we aimed to utilize only nymphs for our study in order to ensure that individuals originated from the samples. In some cases, we were unable to obtain a suitable number of nymphs, so adults were included. Individuals associated with cultivated cotton were collected from a focal cotton field located seven miles west of College Station, TX, in 2015, and a second nearby field was added in 2016 to extend our sampling (Table 4.1). In 2015, the focal cotton field was divided into 12 sections in a grid-like manner, and on each sampling date, every section was sampled with a 38 cm diameter sweep net (20 sweeps per section; total of 240 sweeps on each sample date). Sweep net contents collected from each section were transferred into separate Ziploc bags (30 x 30 cm), and the bags were placed in a cooler equipped with ice packs. Upon return to the laboratory, bags were

Table 4.1. Locality information for all sampled populations used in this study. N is the number of individuals successfully sequenced from each locality. Senesced croton populations refer to those individuals which were reared from eggs collected from croton stems. Cultivated cotton and wild host populations were collected using sweep nets. N (Juveniles) may refer to eggs or nymphs depending on if the population is from senesced croton or from cultivated cotton or wild hosts.

Population	Latitude	Longitude	Date Collected	N (Juveniles)	N (Adults)	N (Total)
Senesced Croton 1	30° 33' 30.72" N	96° 24' 28.31" W	1-Feb-2015	33	0	33
	30° 33' 32.52" N	96° 24' 27.42" W	23-Feb-2016	27	0	27
Senesced Croton 2	30° 23' 53.69" N	96° 21' 20.40" W	1-Feb-2015	32	0	32
	30° 24' 7.64" N	96° 21' 16.21" W	23-Feb-2016	30	0	30
Senesced Croton 3	30° 33' 50.70" N	96° 30' 40.98" W	1-Feb-2015	33	0	33
	30° 33' 51.58" N	96° 30' 40.41" W	23-Feb-2016	30	0	30
Senesced Croton 5	30° 40' 46.91" N	96° 25' 35.74" W	2-Feb-2015	31	0	31
	30° 40' 48.38" N	96° 25' 35.13" W	23-Feb-2016	30	0	30
Senesced Croton 9	30° 46' 42.60" N	96° 42' 32.60" W	2-Feb-2015	30	0	30
	30° 46' 43.14" N	96° 42' 32.25" W	23-Feb-2016	31	0	31
Cultivated Cotton 1	30° 36' 38.38" N	96° 29' 42.80" W	23-Jun-2015	19	0	19
	30° 36' 37.66" N	96° 29' 43.29" W	14-Jul-2015	9	0	9
	30° 36' 38.13" N	96° 29' 43.32" W	10-Aug-2015	13	0	13
	30° 36' 38.98" N	96° 29' 39.18" W	28-Jun-2016	3	23	26
Cultivated Cotton 2	30° 36' 10.65" N	96° 30' 15.76" W	28-Jun-2016	4	22	26
Wild Host 1	30° 33' 45.87" N	96° 30' 55.82" W	28-Jun-2016	6	22	28
	30° 33' 44.58" N	96° 30' 57.09" W	9-Aug-2016	16	10	26
Wild Host 3	30° 36' 43.10" N	96° 29' 34.21" W	28-Jun-2016	5	20	25
	30° 37' 26.95" N	96° 29' 31.99" W	9-Aug-2016	21	4	25
Wild Host 4	30° 38' 26.54" N	96° 27' 24.52" W	29-Jun-2016	25	3	28
	30° 38' 21.08" N	96° 27' 20.55" W	26-Jul-2016	10	14	24
	30° 38' 25.80" N	96° 27' 21.61" W	9-Aug-2016	2	3	5

placed into a freezer in order to euthanize captured insects and stored for no more than two days. Next, *P. seriatus* individuals were separated from by-catch, and larger nymphs (>3rd instar) from each bag on each sampling date were transferred into vials containing 100% ethanol (EtOH). The vials were then stored at -80°C until DNA isolation. In 2016, instead of dividing the field into 12 sections, 20 sweep samples were each taken from

five internal (>30 m from field edge) and five external sites (<10 m from field edge) of each field for a total of 200 sweeps per field. On each sampling date, samples were collected from each side of the field (doubled up on one side) for a total of 10 sweep sites in each cotton field. Sweep net samples were handled and processed in the same manner as in 2015.

Overwintering *P. seriatus* individuals associated with senesced woolly croton were sampled from five localities within a 20-mile radius of the original focal cotton field. All five localities were sampled in both years of the study. To ensure locality fidelity for the senesced croton populations, the host plant itself was collected in February (during the overwintering period) and late-instar juvenile insects were reared from the overwintering eggs found on the plant. Croton stems from each locality were contained in separate burlap sacks and stored in a walk-in-cooler (5.5±1°C) until they were exposed to conditions to terminate diapause as described in Breene et al. (1989). Hatched nymphs from each locality were housed separately in plastic containers (~750 mL) equipped with 80 mesh organza-lined lids. Nymphs were provided green beans (*Phaseolus* spp.) replaced every other day and reared at 29±1°C with a 14:10 (L:D) h photoperiod. Additionally, shredded paper was added to each container to serve as resting sites for nymphs and to minimize movement of green beans. Cages were monitored daily for fourth and fifth instars, which were removed with an aspirator and euthanized directly in 100% EtOH. Nymph samples were then stored at -80°C until DNA isolation.

P. seriatus individuals utilizing alternative wild host plants were sampled in the second year of the study. Three localities near the cultivated cotton fields were sampled in June, July, and August 2016. Sweep net samples were collected from 5 sites within each wild host locality (10 sweeps per site; total of 50 sweeps per wild host locality). Sweep net samples were then processed in the same manner as those collected from cultivated cotton. All individuals were euthanized directly into 80-100% molecular lab grade EtOH and then stored at -80°C until preparation for DNA isolation. Individuals collected in June (representing "early summer") were included from all three sites. At wild host localities 1 and 3, individuals collected in August were used to represent "late summer" genotypes, while wild host locality 4 includes individuals from July and August due to low population density in August. In addition to insect collection, the most dominant plant species at each wild host locality were recorded (up to five species). All vegetation recordings were made during the sampling events for those localities in June, July, and August, 2016.

4.2.2. DNA Isolation, Library Preparation, and Sequencing

We used the Qiagen Gentra Puregene Cell and Tissue Kit to isolate genomic DNA from *P. seriatus* individuals with a modified protocol (Supplementary Information). DNA isolation was carried out in two batches; batch 1 contained the 200 individuals collected in 2015, and batch 2 contained the 361 individuals collected in 2016. DNA was isolated from individuals in both batches using nearly the same protocol, but 2015 samples were mechanically disrupted prior to tissue lysing using dissecting scissors whereas 2016 samples were mechanically disrupted by freezing in liquid nitrogen and crushed with

disposable pestles. Isolated DNA from all 561 individuals was verified for high molecular weight via electrophoresis on a 1.5% agarose gel. DNA sample preparation and sequencing were carried out at the Texas A&M AgriLife Genomics and Bioinformatics Service (TxGen). Isolated DNA was further purified using the Agencourt AMPure XP purification system prior to library preparation. To prepare the double digest restriction site-associated DNA sequencing (ddRADseq) libraries, purified genomic DNA was digested using the restriction enzymes PstI and MluCI and the resulting digested DNA was selected for fragment sizes ranging 250-500 base pairs (bp). Size-selected fragments were then ligated with standard Illumina adapters, multiplexing indexes, and sequencing primers, albeit with a single notable exception - the R1 reads (forward reads; those sequenced in the 5' direction) were ligated with a custom sequencing primer that contained the 5' enzyme remnant. The batch of 2015 libraries was sequenced on an Illumina HiSeq 2500 using 125x125 sequencing cycles. The batch of 2016 libraries was sequenced on an Illumina HiSeq 4000 using 150x150 sequencing cycles. Raw sequences generated by TxGen were demultiplexed prior to delivery to the authors and FastQC version 0.11.3 reports were provided by TxGen for each sample, and the authors subjectively reviewed those reports in order to ensure that the dataset was of suitable quality for use in a bioinformatics pipeline (Andrews, 2010). Raw reads were then filtered using Kraken version 1.1 to remove any potential bacterial contamination (Wood and Salzberg, 2014). Sequences with any match to the non-redundant bacterial database hosted by the National Center for Biotechnology Information (NCBI) were removed. Genomic loci containing SNPs were then identified

using the SNP-calling pipeline dDocent version 2.5.5 (Puritz et al., 2014a; Puritz et al., 2014b). dDocent was run using default parameters and procedures as described in the user manual, except as noted hereafter. dDocent was paused after trimming, and Trimmomatic version 0.38 was used to further trim sequences to a uniform length of 90 bp (Bolger et al., 2014). dDocent was then restarted and allowed to run using a percent clustering similarity of 90%. VCFtools version 0.1.15 was then used to filter the dDocent output in variant call format (vcf) in accordance with the tutorial in the dDocent user guide; however, we also filtered out loci that were not present in 100% of individuals in order to guarantee a dataset with no missing data (Danecek et al., 2011). The final filtered SNP dataset was used for all downstream analyses. Any necessary file conversions were carried out using PGDSpider version 2.1.1.3 (Lischer and Excoffier, 2012).

4.2.3. Population Genetic Analyses

The open-source software RStudio version 1.1.442 and its associated packages were used to carry out conventional population genetic analyses (R Core Team, 2018). R/vcfR version 1.7.0 was used to read the filtered vcf file and prepare objects for use with other packages (Knaus and Grünwald, 2017). R/adegenet version 2.1.1 was used to calculate the expected and observed heterozygosity (H_E and H_O) at each locus and to plot H_O as a function of H_E (Jombart, 2008; Jombart and Ahmed, 2011). R/genepop version 1.0.5 was used to calculate genetic diversity based on allele identity and the inbreeding coefficient (F_{IS}) for each population (Rousset, 2008). We calculated both the average diversity within individuals and the average diversity among individuals within a population.

R/genepop was also used to estimate gene flow by calculating pairwise F_{ST} values for all pairs of populations in the study. R/adegenet was used to carry out and plot the results of a principal components analysis (PCA) in order to visualize the spatial clustering of individual genotypes based on principal components 1 and 2. Since R/adegenet can only carry out a PCA using biallelic loci, SNP loci with more than two alleles were removed from the dataset. The software fastSTRUCTURE version 1.0 was used to calculate the probability of individual assignment to one or more predetermined genotypic groups (hereafter referred to as the "population assignment probability test") where K is the number of genotypic groups and $1 \leq K \leq 22$ (Raj et al., 2014). PLINK version 1.07 was used to convert the vcf file into a format that was suitable for input into fastSTRUCTURE (Purcell et al., 2007). The browser-based program StructureSelector was used to evaluate the fastSTRUCTURE outputs and to choose the optimal value of K for our dataset, and CLUMPAK was used to visualize individual population assignments relative to other individuals (Kopelman et al., 2015; Li and Liu, 2018). For the Puechmaille (2016) component of StructureSelector, we tested mean membership coefficient thresholds of 0.5, 0.6, 0.7, and 0.8.

4.3. Results

4.3.1. Host Plant Diversity at Wild Host Localities in 2016

The dominant plant species present at the Wild Host 1 locality were woolly croton and horsemint (*Monarda* spp.) on June 28. On that same day, Wild Host 3 was dominated by woolly croton and silverleaf nightshade (*Solanum elaeagnifolium* Cavanilles). Wild Host 4 had a mix of woolly croton, horsemint, and silverleaf nightshade on June 29 and a mix

of horsemint and woolly croton on July 26. On August 9, only woolly croton was abundant enough to be considered dominant at all three localities.

4.3.2. Genetic Variation Within and Among Populations

1.611 TB of raw sequence data was generated by ddRADseq. FastQC reports were consistent with what would be expected from a normal Illumina run, and thus considered to be of sufficiently high quality for further downstream analysis. dDocent identified 694,072 SNP loci, most of which were subsequently filtered, resulting in a final SNP dataset consisting of 12,653 loci with 0% missing data. This dataset was used in all of the analyses except in cases where non-binary loci were removed to create a subset for

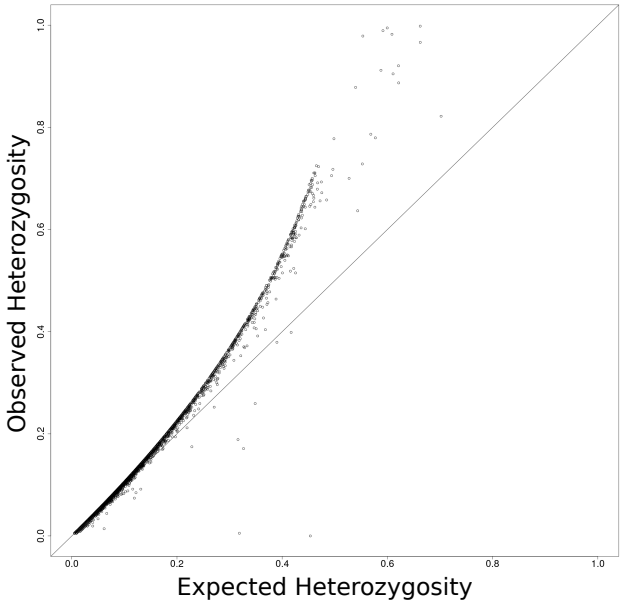

Figure 9. Observed heterozygosity plotted as a function of expected heterozygosity. Both axes range from 0 to 1, the theoretical minimum and maximum for heterozygosity. Drawn line indicates a theoretical 1-to-1 relationship wherein observed heterozygosity matches expected heterozygosity. Points above and to the left of the line indicate loci that have an observed heterozygosity that exceeds the expected.

Table 4.2. Population genetic summary statistics for every sampled locality. $1\text{-}Q_{intra}$ and $1\text{-}Q_{inter}$ are measures based on heterozygosity. $1\text{-}Q_{intra}$ is the average genetic diversity within individuals and $1\text{-}Q_{inter}$ is the average genetic diversity among individuals within a population. F_{IS} is the inbreeding coefficient. Populations labeled as "Croton" refer to the individuals collected as eggs from senesced croton.

			$1\text{-}Q_{intra}$	$1\text{-}Q_{inter}$	F_{IS}
2015	February	Croton 1	0.0576	0.0526	-0.0942
		Croton 2	0.0588	0.0531	-0.1071
		Croton 3	0.0637	0.0582	-0.094
		Croton 5	0.0652	0.06	-0.0856
		Croton 9	0.0773	0.067	-0.155
	Jn	Cotton 1	0.0893	0.0758	-0.1777
	Jl	Cotton 1	0.0846	0.0752	-0.124
	Au	Cotton 1	0.0607	0.0547	-0.1099
2016	February	Croton 1	0.0601	0.0557	-0.0798
		Croton 2	0.0621	0.0576	-0.0771
		Croton 3	0.0592	0.0553	-0.0711
		Croton 5	0.061	0.0566	-0.0792
		Croton 9	0.0606	0.0564	-0.0748
	Early Summer	Cotton 1	0.0717	0.0629	-0.1413
		Cotton 2	0.0725	0.0659	-0.1413
		Wild Host 1	0.0703	0.0638	-0.1017
		Wild Host 3	0.0699	0.0629	-0.1118
		Wild Host 4	0.0702	0.0614	-0.143
	Late Summer	Wild Host 1	0.0677	0.0618	-0.0958
		Wild Host 3	0.0715	0.0643	-0.1112
		Wild Host 4	0.069	0.0609	-0.1316

some analyses. For the majority of loci, H_O exceeded H_E (Fig. 9). The inbreeding coefficient (F_{IS}) was negative for all sampled populations (Table 4.2).

The global F_{ST} for our dataset was 0.0131. Pairwise values of F_{ST} ranged from -0.0008 to 0.0446 (Table 4.3). Higher pairwise values were generally observed when comparing 2015 populations to 2016 populations, and the highest values of F_{ST} (>0.03) were observed when comparing the focal cotton field June 2015 population to any other

Table 4.3. Pairwise F_{ST} values for all pairs of populations sampled. Populations labeled as "croton" are those sampled from senesced croton during the overwintering period.

2015	February	Croton 1	-																				
		Croton 2	0.0063	-																			
		Croton 3	0.0037	0.0099	-																		
		Croton 5	0.0099	0.0139	0.0021	-																	
		Croton 9	0.0268	0.0299	0.024	0.0045	-																
	Jun	Cotton 1	0.0342	0.0376	0.0393	0.043	0.0446	-															
	Jul	Cotton 1	0.027	0.0303	0.0116	0.0233	0.0248	0.0366	-														
	Aug	Cotton 1	0.0044	0.0176	0.0053	0.0133	0.0152	0.0429	0.0249	-													
2016	February	Croton 1	0.0095	0.0134	0.0089	0.0108	0.0265	0.0445	0.0241	0.0146	-												
		Croton 2	0.0091	0.0134	0.0084	0.0104	0.0254	0.0413	0.0229	0.0145	0.0009	-											
		Croton 3	0.0082	0.0127	0.0082	0.0101	0.026	0.0443	0.0246	0.0134	0.0026	0.001	-										
		Croton 5	0.0089	0.0133	0.0085	0.0104	0.0262	0.043	0.0233	0.0133	0.0013	0.0004	0.0002	-									
		Croton 9	0.0103	0.0144	0.0102	0.0116	0.0273	0.0446	0.0248	0.0152	0.0049	0.0063	0.0032	0.0045	-								
	Early Summer	Cotton 1	0.017	0.0196	0.014	0.0147	0.0256	0.0403	0.0224	0.0183	0.0103	0.0122	0.0124	0.0107	0.013	-								
		Cotton 2	0.0122	0.0154	0.0103	0.0113	0.0231	0.0369	0.017	0.0139	0.0054	0.004	0.0038	0.004	0.0083	0.0059	-							
		Wild Host 1	0.0105	0.0139	0.0086	0.0044	0.0031	0.0358	0.017	0.0124	0.0044	0.0057	0.0043	0.0044	0.0031	0.0054	0.003	-						
		Wild Host 3	0.0109	0.0146	0.0085	0.0042	0.0026	0.0374	0.0173	0.013	0.0046	0.0059	0.0048	0.0042	0.0026	0.0048	0.0029	-0.0002	-					
		Wild Host 4	0.0184	0.0212	0.0156	0.0123	0.0151	0.0423	0.0245	0.0207	0.0128	0.0138	0.0144	0.0123	0.0151	0.0013	0.0098	0.0085	0.0078	-				
	Late Summer	Wild Host 1	0.0103	0.0137	0.0082	0.0043	0.0012	0.0386	0.018	0.0127	0.0039	0.0057	0.0044	0.0043	0.0012	0.0071	0.004	-0.0005	-0.0008	0.0088	-			
		Wild Host 3	0.0114	0.0144	0.009	0.0052	0.0035	0.0367	0.0165	0.0134	0.0053	0.0065	0.0061	0.0052	0.0035	0.002	0.0037	0.0016	0	0.0039	0.0008	-		
		Wild Host 4	0.0148	0.0181	0.0127	0.0088	0.0114	0.0409	0.0229	0.0168	0.0095	0.011	0.0102	0.0088	0.0114	0.0003	0.0074	0.0063	0.0049	0.0008	0.0068	0.0018	-	
			Croton 1	Croton 2	Croton 3	Croton 5	Croton 9	Cotton 1	Cotton 1	Cotton 1	Croton 1	Croton 2	Croton 3	Croton 5	Croton 9	Cotton 1	Cotton 2	WH 1	WH 3	WH 4	WH 1	WH 3	WH 4
					February			June	July	August			February				Early Summer				Late Summer		
			2015								2016												

Global F_{ST} = 0.0131

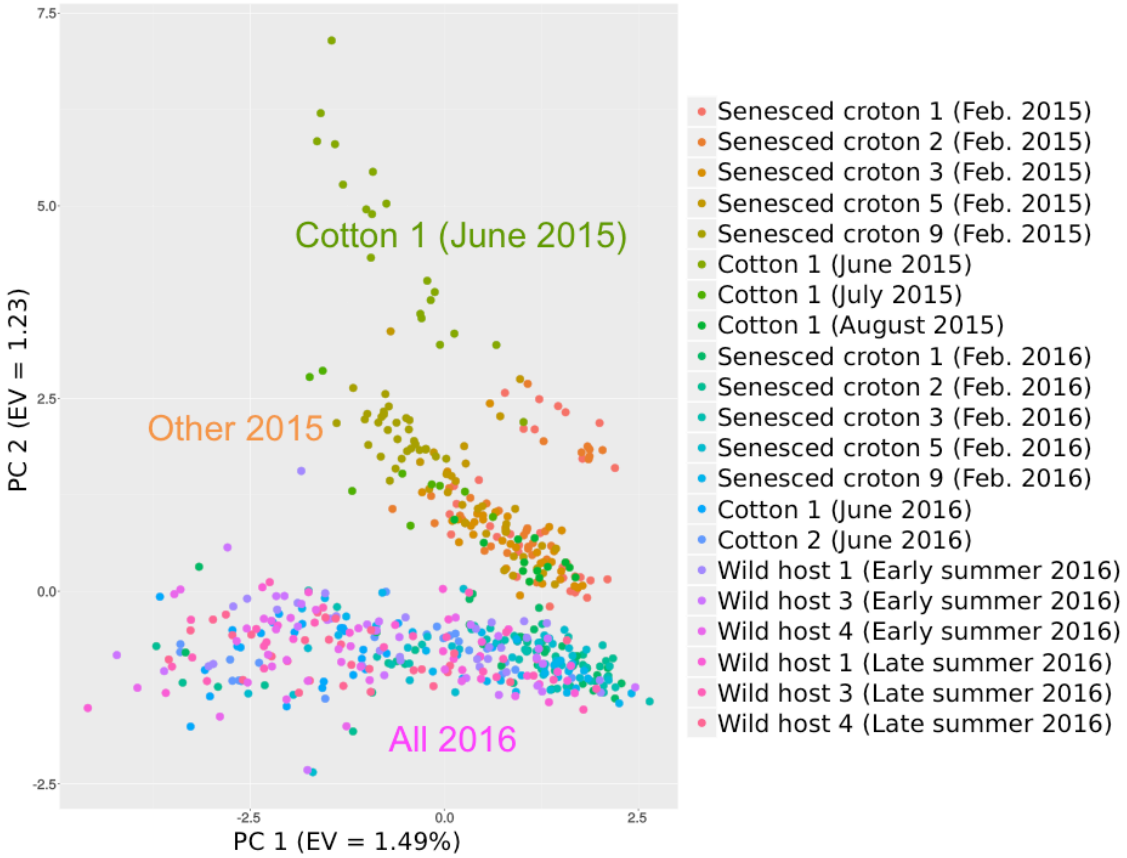

Figure 10. Results of principal components analysis presented as principal component 2 (PC 2) plotted as a function of principal component 1 (PC 1). Eigenvalues (EV) indicate percent variation explained by each respective PC.

population. For our principal components analysis, the first 14 principal components were determined to have the greatest eigenvalues before reaching an asymptote, so the PCA was conducted using only those first 14 PCs. A coordinate plot of principal components 1 and 2 revealed the presence of three distinct clusters of populations, yet some individuals in different clusters are spatially closer to one another than either is to other members of their clusters (Fig. 10). In general, all 2016 samples clustered together

and all 2015 samples clustered together except for the June 2015 focal cotton field population, which formed the third cluster.

The software fastSTRUCTURE was used to calculate the probability of individual assignment to one or more genotypic groups, where K is the number of groups. StructureSelector results indicated that the most likely actual value of K is 1 or 3. The marginal likelihood of the model is maximized at K=1, while K=3 is predicted using the Puechmaille (2016) method. The result of K=3 was consistent across all four tested mean membership coefficient thresholds. The CLUMPAK plot of population assignment probabilities at K=3 revealed the dominance of two probable genotypic groups (Fig. 11). One such group is prominent in the June 2015 focal cotton field population and occurs more frequently in 2015 whereas the second genotypic group is more prevalent in 2016. Both genotypes; however, are represented in both years of the study.

4.4. Discussion

Our results, overall, show that there is generally high gene flow between populations of *P. seriatus* infesting cultivated cotton and those associated with other host plants in the Brazos Valley of TX. Pairwise and global measures of F_{ST} were all below 0.05 (Table 4.3), which by Wright's (1951) original interpretation of the statistic indicates little to no differentiation. This conclusion is further supported by the observed negative values of within-population F_{IS} (Table 4.2) that are indicative of heterozygote excess, a pattern that is expected in outcrossing populations with high gene flow (Tapio et al., 2003). This outcome is not entirely surprising because our populations were sampled from a

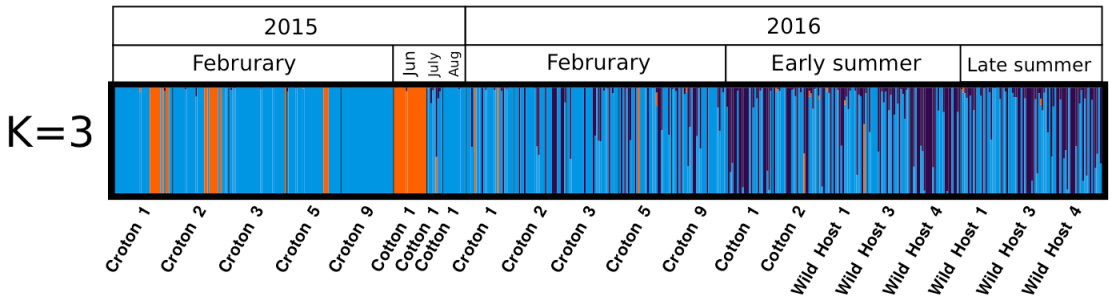

Figure 11. CLUMPAK plot of fastStructure results when the given number of genotypic groups is 3. Each color represents 1 of those three groups. Each vertical bar represents a single individual from the designated population, and the proportion of the bar that is one of the three colors represents the percent probability (from 0 to 1) of that individual's assignment to that genotypic group.

relatively small geographic area. Further, this finding is consistent with the conclusion drawn by Barman et al. (2012) that there is no signal of HAD present in these populations.

SNP datasets are very sensitive by virtue of the number of markers, enabling identification of some very slight population-level genetic differentiation. We identified three distinct genotypic groups that are differentiable despite the high gene flow detected between those groups. Previous studies that have examined population genetic structure in *P. seriatus* have largely focused on broad-scale genotypic distributions or genetic host plant associations, but our data suggests that there may be an important temporal component that contributes to the population genetic structure (Barman et al., 2012; Barman et al., 2013; Antwi et al., 2015). Our study is unique in that we sampled the same geographic locations multiple times over the course of two years, enabling detection of changes in local genotypes over time. We observed population sub-structure

that seemed to be more strongly influenced by time (year and monthly) than by any association with a particular host plant. Specifically, we observed that individuals collected in 2015 were genetically distinct from those collected in 2016. Furthermore, the individuals collected from commercially cultivated cotton at one sampling time in June of 2015 were genetically distinct from all other sampled populations. These three differentiated groups are clearly visible in our PCA plot (Fig. 10). Though there were a few non-conforming individuals, these groups were strongly supported by the non-overlapping 95% confidence ellipses associated with their respective point clusters. The population assignment probability test (Fig. 11) provides evidence for the existence of the same three genotypic groups, but arguably enables better visualization of how the groups are related. The most obvious differences were across years between 2015 and 2016. Although within each year, there is an assigned genotypic group that is present in both years, there is also one assigned genotypic group that is only associated with 2015. This unique genotypic group is composed of the individuals from commercial cotton in June of 2015, consistent with the PCA plot showing similar differentiation (Fig. 10). Interestingly, there are a few individuals sampled from senesced woolly croton populations (at least one individual from all populations except Croton 9) that were also assigned to this genotypic group. This helps to explain the overlap of the "non-conforming" individuals in the PCA and also provides further evidence that HAD is not occurring in the Brazos Valley because there are other individuals from woolly croton that were assigned to other genotypic groups. The other genotypic group to which 2015 samples were commonly assigned is represented in all croton populations and is the

dominant genotypic group in July and August 2015. This shows that there can be genotypic turnover even within a single growing season wherein the *P. seriatus* individuals present in a cotton field in the late summer may not necessarily be related to those present earlier in the season. Additionally, in the July 2016 data we first see an individual with greater than 50% percent probability of being assigned to the third genotypic group. This individual signals the beginning of the observed yearly turnover wherein a third genotypic group becomes present in all sampled host plant species by the end of the year. In 2016, the monthly turnover in cotton described from 2015 could not be observed because we were only able to obtain adequate sample size from both cotton fields at a single sampling time in June. However, our 2016 data allowed us to observe a pattern consistent with the movement of genotypic groups associated with cotton in July and August 2015 onto the overwintering host, woolly croton, detected in the February 2016 samples, and then back onto cotton in the early summer of 2016. Finally, our sampling in the Wild Host localities also allowed us to observe the persistence of *P. seriatus* genotypes on live woolly croton throughout the year and the movement of other genotypes back into live woolly croton after other host plants had begun to senesce or become otherwise less attractive.

Taken all together, our dataset indicates one instance of monthly turnover of local genotypes in cotton and one instance of yearly turnover. Though neither of these transitions is absolute (all three genotypic groups are represented in some manner in both years of the study, and one group is well-represented in both years), our results suggest that the time at which insect sampling occurs can have a profound effect on the detection

of genetic population structuring in populations of *P. seriatus*. However, the genotypic population structure is not maintained over time. Variation in overwintering emergence and host plant phenology seem likely to contribute to the temporal dynamics of population genetic structure observed here. Different genotypic groups may represent different cohorts that vary in emergence times from overwintering due to local environmental variation, resulting in those individuals colonizing cotton and other hosts at different times. Host plant phenology has been linked to variation in CFH colonization of plants from year to year, and is likely to play a major role in determining which genotypic groups are observed at a particular location at a particular time (Almand et al., 1976). Plant phenology may also explain the slight differentiation observed, similar to what has been documented in *Rhagoletis* flies, wherein differences in plant phenology have driven host-associated pre-mating isolation (Feder et al., 1988; Filchak et al., 2000). Alternatively, new genotypic groups may arrive in a certain location as result of a migration event. *P. seriatus* has been documented as having a dispersal range of at least 20 miles, and individuals may be able to travel even farther by taking advantage of prevailing wind currents (Gaines and Ewing, 1938). So, new genotypes may appear in local populations as a result of such movement. In either case, the effects of differentiation due to differences in host plant phenology and the introduction of new genotypes due to migration seem to be strongly counteracted by the homogenizing effects of gene flow.

Despite the observed genotypic turnover within and among years, our dataset nevertheless provides support for woolly croton as a year-end site of population

admixture, at least in areas where it occurs. Though there are differences in which genotypic groups are present from 2015 to 2016, in both years, any genotypic group associated with cotton is also found in association with woolly croton and vice versa. Furthermore, gene flow is considered to be very high between all sampled populations (Table 4.3), so we conclude that there was no apparent differentiation between cotton and woolly croton populations of *P. seriatus* in either year of the study. On a broader geographic scale, we cannot extrapolate to what occurs in areas where woolly croton does not occur. Barman et al. (2012) and Antwi et al. (2015) have demonstrated that there is a geographic pattern of HAD in *P. seriatus* wherein HAD is observed in areas where woolly croton does not occur, but is not observed in areas where it is present. Our study also supports the lack of HAD when woolly croton is present, but further investigations into patterns of genetic differentiation are still required in areas where woolly croton is absent.

5. CONCLUSIONS

The population genomics approach utilized in this dissertation was successfully applied at broad and fine geographic scales and to two very different cotton pest species. On a broad scale, we used SNP genotyping to resolve the phylogeography of the boll weevil across two continents and assign probable source populations of boll weevils occurring in previously eradicated areas of the US. On a small scale, we used the same methodology to resolve the local population genomic structure of the cotton fleahopper within a roughly 60 km area. Together, these applications demonstrated the flexibility of our approach to answer evolutionary questions and address pest management issues for a variety of systems. We also showed the power of SNP markers to both determine taxonomic identity and assign individuals to a likely population of origin even when other markers may be unable to fully do so.

5.1. Applications to Boll Weevil Management and Eradication

Our boll weevil phylogeography research has major implications for boll weevil management in North America. We revealed that populations occurring along the US-Mexico border are likely contiguous across international boundaries, particularly near the Gulf Coast. This necessitates international cooperation in order to fully address infestations occurring in that area. We also provided evidence for the two-form hypothesis of boll weevil variation over the three-form hypothesis such that there are two geographic subspecies of boll weevil, *A. g. grandis* in the east and *A. g. thurberiae* in the west. However, our research suggests that the western subspecies may not be a

truly host-associated variant as has been described in prior literature. This is an important consideration for managers in the west because they will now need to consider populations occurring on wild host plants as potential sources of infestations of commercial cotton.

In Argentina, we found limited evidence of a recent population bottleneck or founder effect, consistent with what would be expected after a recent range expansion. However, we also found that populations across the country are exchanging gene flow. This could be due to natural movement of weevils, which would imply that successful management would need to address the cotton growing regions of the country as a whole. Alternatively, weevils can be moved around by human activity. So, if human activity is contributing to gene flow between weevil populations in this country, that will need to be addressed in order to maintain successful local management.

Human movement of weevils also remains a problem for the US Eradication Program. Re-infestations of previously eradicated areas are hypothesized to have occurred due to human activity. Our source determination can help target problematic areas and prevent further re-infestations from those areas. The primary limitation to our SNP genotype assay, as implemented here, is that it takes significantly longer to complete (1-4 months) than the currently utilized diagnostic assays (less than 1 week). Thus, future work leading to the development of a rapid diagnostic SNP assay based on the research presented here will help capitalize on the power of the SNP markers without sacrificing the speed of diagnosis. To fully realize the power of a SNP approach to

phylogeography or diagnostics, future research should also focus on further collection of weevils from other geographic areas and host plants.

5.2. Applications to Insect Resistance Management for Cotton Fleahopper

With respect to developing an IRM strategy for the new transgenic control of *P. seriatus*, our research provides support for the natural refuge status of woolly croton as an overwintering host that promotes gene flow between *P. seriatus* genotypes associated with cotton and those associated with other host plants. Our results showed that while there were multiple genotypes recovered from cotton, all of those genotypes were also recovered from woolly croton populations. Additionally, we also detected a temporal turnover of local genotypes such that a given genotypic group of *P. seriatus* associated with cotton would not be expected to be exposed to control measures in cotton on a chronic basis. A critical caveat to consider is variation in the availability of woolly croton across the broader range of *P. seriatus* when making management decisions involving natural refuges. In areas where woolly croton does not occur, further work should be conducted to identify the alternative overwintering host plants that may act similarly to woolly croton as a site of year-end admixture. Beyond that, for future studies investigating population genomic structure in species with complex host associations, our results highlight the need to implement a temporal sampling scheme that accounts for potential turnover of local genotypes on yearly or even a monthly basis to avoid potentially confounding effects of spurious genetic structuring that is not maintained over time.

REFERENCES

Almand, L. K., Sterling, W. L., and Green, C. L. (1976). Seasonal abundance and dispersal of the cotton fleahopper as related to host plant phenology. *Texas FARMER Collection*.

Alvarado, A., Jones, R. W., Pedraza-Lara, C., Villanueva, O. A., and Pfeiler E. (2017). Reassessment of the phylogeography and intraspecific relationships of western and eastern populations of the boll weevil, *Anthonomus grandis* Boheman (Coleoptera: Curculionidae), in North America. *Biological Journal of the Linnean Society* 122(1): 29-45.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Antwi, J. B., Sword, G. A., and Medina, R. F. (2015). Host-associated differentiation in a highly polyphagous, sexually reproducing insect herbivore. *Ecology and Evolution* 5(13): 2533-2543.

Bachman, P. M., Ahmad, A., Ahrens, J. E., Akbar, W., Baum, J. A., Brown, S., Clark, T. L., Fridley, J. M., Gowda, A., Greenplate, J. T., Jensen, P. D., Mueller, G. M., Odegaard, M. L., Tan J., Uffman, J. P., and Levine, S. L. (2017). Characterization of the activity spectrum of MON 88702 and the plant-incorporated protectant Cry51Aa2.834_16. *PloS ONE* 12(1): e0169409.

Barman, A. K., Parajulee, M. N., Sansone, C. G., Suh, C. P.-C., and Medina, R. F. (2012). Geographic pattern of host‐associated differentiation in the cotton fleahopper, *Pseudatomoscelis seriatus*. *Entomologia Experimentalis et Applicata* 143(1): 31-41.

Barman, A. K., Sansone, C. G., Parajulee, M. N., and Medina, R. F. (2013). Population genetic structure of *Pseudatomoscelis seriatus* (Hemiptera: Miridae) in the cotton-growing regions of the United States. *Journal of Economic Entomology* 106(5): 2225-2233.

Barr, N., Ruiz-Arce, R., Obregón, O., de Leon, R., Foster, N., Reuter, C., Boratynski, T., and Vacek, D. (2013). Molecular diagnosis of populational variants of *Anthonomus grandis* (Coleoptera: Curculionidae) in North America. *Journal of Economic Entomology* 106(1): 437-449.

Bartlett, A. C. (1981). Isozyme polymorphisms in boll weevils and thurberia weevils from Arizona. *Annals of the Entomological Society of America* 74: 359-362.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-2120.

Bottrell, D. G. and Almand L. K. (1968). Evaluation of the 1967 reproductive-diapause boll weevil control program of the Texas High Plains. *Texas Agricultural Experiment Station MP* 904. 4 pp.

Bougeard, S. and Dray, S. (2018). Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software* 86(1): 1-17.

Breene, R. G., Martin Jr., W. R., and Sterling, W. L. (1989). Rearing methods for the cotton fleahopper. *Southwestern Entomologist* 14: 249-253.

Burke, H. R. (1968). Geographic variation and taxonomy of *Anthonomus grandis* Boheman. *Texas Agricultural Experiment Station, Department of Entomology Technical Report*. Mimeographed. 152 pp.

Burke, H. R., Clark, W. E., Cate, J. R., and Fryxell, P. A. (1986). Origin and dispersal of the boll weevil. *Bulletin of the Entomological Society of America* 32: 228-238.

Burt, E. C., Lloyd, E. P., and Smith, D. B. (1968). A flail machine for destroying fallen cotton squares infested with boll weevils. *Journal of Economic Entomology* 61: 1-3.

Burt, E. C., Lloyd, E. P., and Smith, D. B. (1969). Control of the boll weevil by mechanically destroying fallen infested cotton squares. *Journal of Economic Entomology* 62: 862-65.

Caprio, M. A. (2001). Source-sink dynamics between transgenic and non-transgenic habitats and their role in the evolution of resistance. *Journal of Economic Entomology* 94(3): 698-705.

Caprio, M. A., Faver, M. K., and Hankins, G. (2004). Evaluating the impacts of refuge width on source-sink dynamics between transgenic and non-transgenic cotton. *Journal of Insect Science* 4(3).

Carrière, Y., Dutilleul, P., Ellers-Kirk, C., Pedersen, B., Haller, S., Antilla, L., Dennehy, T. J., and Tabashnik, B. E. (2004). Sources, sinks, and the zone of influence of refuges for managing insect resistance to Bt crops. *Ecological Applications* 14(6), 1615-1623.

Carrière, Y., Goodell, P. B., Ellers-Kirk, C., Larocque, G., Dutilleul, P., Naranjo, S. E., and Ellsworth, P. C. (2012). Effects of local and landscape factors on population dynamics of a cotton pest. *PLoS ONE* 7(6): e39862.

Choi, S. K., Kim, K. S., Lee, H., Adamczyk, J. J., Greenberg, S. M., Westbrook, J. K., and Sappington, T. W. (2011). Temporal changes in genetic variation of boll weevil (Coleoptera: Curculionidae) populations, and implications for population assignment in eradication zones. *Annals of the Entomological Society of America* 104(4): 816-825.

Coppedge, J. R., Lindquist, D. A., Ridgway, R. L., Cowan, C. B., and Bariola, L. A. (1969). Sidedress applications of Union Carbide UC-21149 for control of overwintered boll weevils. *Journal of Economic Entomology* 62: 558-565.

Cross, W. H. and Hardee, D. D. (1968). Traps for survey of overwintered boll weevil populations. *US Department of Agriculture Cooperative Economic Insect Report* 18: 430.

Cross, W. H. and McGovern, W. L. (1969). New parasites, *Zatropis perdubius* and *Megaselia aletiae*, of the boll weevil, *Anthonomus grandis*. *Annals of the Entomological Society of America* 62: 674.

Cross, W. H., McGovern, W. L., and Mitchell, H. C. (1969a). Biology of *Bracon kirkpatricki* and field releases of the parasite for control of the boll weevil. *Journal of Economic Entomology* 62: 448-454.

Cross, W. H., Hardee, D. D., Nichols, F., Mitchell, H. C., Mitchell, E. B., Huddleston, P. M., and Tumlinson, J. H. (1969b). Attraction of female boll weevils to traps baited with males or extracts of males. *Journal of Economic Entomology* 62(1): 154-61.

Cross, W. H. and Chesnut, T. L. (1971). Arthropod parasites of the boll weevil, *Anthonomus grandis*: an annotated list. *Annals of the Entomological Society of America* 64: 516-527.

Cross, W. H. (1973). Biology, control, and eradication of the boll weevil. *Annual Review of Entomology* 18: 17-46.

Cross, W. H., Lukefahr, M. J., Fryxell, P. A., and Burke, H. R. (1975). Host plants of the boll weevil. *Environmental Entomology* 4: 19-26.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., and 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* 27(15): 2156-2158.

Dray, S. and Dufour, A. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22(4): 1-20.

Eden, W. G. (1968). First conference on test methods for resistance in insects of agricultural importance: method for the boll weevil and tentative method for spider mites. *Bulletin of the Entomological Society of America* 14: 31-37.

Esquivel, J. F. and Esquivel, S. V. (2009). Identification of cotton fleahopper (Hemiptera: Miridae) host plants in central Texas and compendium of reported hosts in the United States. *Environmental Entomology* 38(3): 766-780.

Ewing, K. P. (1929). Effects on the cotton plant of the feeding of certain Hemiptera of the family Miridae, *Journal of Economic Entomology* 22(5): 761-765.

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Feder, J. L., Chilcote, C. A., and Bush, G. L. (1988). Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* 336: 61-64.

Filchak, K. E., Roethele, J. B., and Feder, J. L. (2000). Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* 407: 739-742.

Fye, R. E. (1968a). The thurberia weevil in Arizona. *Journal of Economic Entomology* 61: 1264-1268.

Fye, R. E. (1968b). Spread of the boll weevil by drainage water and air currents. *Journal of Economic Entomology* 61: 1418-1424.

Fye, R. E., Cole, C. L., Tingle, F. C., Stoner, A., Martin, D. F., and Curl, L. F. (1968). A reproductive-diapause control program for the boll weevil in the Presidio, Texas-Ojinaga, Chihuahua area, 1965-67. *Journal of Economic Entomology* 61(6): 1660-1666.

Gaines, J. C. and Ewing, K. P. (1938). The relation of wind currents, as indicated by balloon drifts, to cotton fleahopper dispersal. *Journal of Economic Entomology* 31: 674-677.

Georghiou, G. P. and Taylor, C. E. (1977). Genetic and biological influences in the evolution of insecticide resistance. *Journal of Economic Entomology* 70(3): 319-323.

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

Guse, C. A., Onstad, D. W., Buschman, L. L., Porter, P., Higgins, R. A., Sloderbeck, P. E., Cronholm, G. B., and Peairs, F. B. (2002). Modeling the development of resistance by stalk-boring Lepidoptera (Crambidae) in areas with irrigated transgenic corn. *Environmental Entomology* 31(4): 676-685.

Hanski, I. and Gilpin, M. (1991). Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society* 42: 3-16.

Hardee, D. D., Wilson, N. M., Mitchell, E. B., and Huddleston, P. M. (1971). Factors affecting activity of grandlure, the pheromone of the boll weevil in laboratory bioassays. *Journal of Economic Entomology* 64: 1454-1456.

Harrison, S. (1991). Local extinction in a metapopulation context: an empirical evaluation. *Biological Journal of the Linnean Society* 42: 73-88.

Hixson, E. (1941). The host relation of the cotton flea hopper. *Iowa State College Journal of Science* 16: 66-68.

Holtzer, T. O. and Sterling, W. L. (1980). Ovipositional preference of the cotton fleahopper, *Pseudatomoscelis seriatus*, and distribution of eggs among host plant species. *Environmental Entomology* 9: 236-240.

Howard, L. O. (1894). A new cotton insect in Texas. *Insect Life* 7: 273.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11): 1403-1405.

Jombart, T. and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27(21): 3070-3071.

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.

Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics* 6: 208.

Kim, K. S. and Sappington, T. W. (2004). Genetic structuring of boll weevil populations in the US based on RAPD markers. *Insect Molecular Biology* 13(3): 293-303.

Kim, K. S. and Sappington, T. W. (2006). Molecular genetic variation of boll weevil populations in North America estimated with microsatellites: implications for patterns of dispersal. *Genetica* 127: 143-161.

Kim, K. S., Cano-Ríos, P., and Sappington, T. W. (2006). Using genetic markers and population assignment techniques to infer origin of boll weevils (Coleoptera: Curculionidae) unexpectedly captured near an eradication zone in Mexico. *Environmental Entomology* 35(4): 813-826.

Kim, K. S., Sappington, T. W., and Allen, C. T. (2008). Genetic profiling to determine potential origins of boll weevils (Coleoptera: Curculionidae) captured in a texas eradication zone: endemicity, immigration, or sabotage? *Journal of Economic Entomology* 101(6): 1729-1736.

Knaus, B. J. and Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* 17(1): 44-53.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15(5): 1179-1191.

Kuester, A. P., Jones, R. W., Sappington, T. W., Kim, K. S., Barr, N. B., Roehrdanz, R. L., Senechal, P., and Nason, J. D. (2012). Population structure and genetic diversity of the boll weevil (Coleoptera: Curculionidae) on *Gossypium* in North America. *Annals of the Entomological Society of America* 105(6): 902-916.

Lange, F., Olmstead, A. L., and P. W. Rhode. (2009). The impact of the boll weevil, 1892-1932. *The Journal of Economic History* 69(3): 685-718.

Leaché, A. D., Banbury, B. L., Felsenstein, J., de Oca, A. N., and Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology* 64(6): 1032-47.

Li, Y. L. and Liu, J. X. (2018). StructureSelector: a web based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources* 18: 176-177.

Lincoln, C. (1969). The effect of agricultural practices on insect habitats in a typical Delta community. *Proceedings of the Tall Timbers Conference on Ecological Animal Control by Habitat Management* 1: 13-18.

Lischer, H. E. L. and Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28(2): 298-299.

Lloyd, E. P., Tingle, F. C., McCoy, J. R., and Davich, T. B. (1966). The reproduction-diapause approach to population control of the boll weevil. *Journal of Economic Entomology* 59: 813-816.

Lloyd, E. P., Tingle, F. C., Merkl, M. E., Burt, E. C., Smith, D. B., and Davich, T. B. (1967). Comparison of three rates of application of ultra-low-volume azinphosmethyl in a reproduction-diapause control program against the boll weevil. *Journal of Economic Entomology* 60(6); 1696-1699.

Lloyd, E. P., Merkl, M. E., Tingle, F. C., Scott, W. P., Hardee, D. D., and Davich, T. B. (1972). Evaluation of male-baited traps for control of boll weevils following a reproduction-diapause program in Monroe County, Mississippi. *Journal of Economic Entomology* 65(2); 552-555.

Marin, H. C. (1981). El picudo del algodonero – treinta años de existencia en Colombia. *ICABT* 81.

Martins, W. F. S., Ayres C. F. J., and Lucena, W. A. (2007). Genetic diversity of Brazilian natural populations of *Anthonomus grandis* Boheman (Coleoptera: Curculionidae), the major cotton pest in the New World. *Genetics and Molecular Research* 6(1): 23-32.

McGarr, R. L. and Chapman, A. J. (1966). Initial field tests with methyl parathion and EPN in Mexico against the boll weevil. *Journal of Economic Entomology* 61: 1107-1108.

McGarr, R. L. and Wolfenbarger, D. A. (1969). Methyl parathion, toxaphene, and DDT used alone in combination for control of several cotton insects. *Journal of Economic Entomology* 62: 1249-50.

McGarr, R. L. and Wolfenbarger, D. A. (1970). Insecticides for control of four cotton insects in 1968. *Journal of Economic Entomology* 63: 1324-1325.

Meihls, L. N., Higdon, M. L., Siegfried, B. D., Miller, N. J., Sappington, T. W., Ellersieck, M. R., Spencer, T. A., and Hibbard, B. E. (2008). Increased survival of western corn rootworm on transgenic corn within three generations of on-plant greenhouse selection. *Proceedings of the National Academy of Sciences* 105(49): 19177-19182.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA*: 1-8.

Mistric Jr., W. J., and Covington, B. M. (1968). Effects of square removal on cotton production with reference to boll weevil damage. *Journal of Economic Entomology* 61: 1060-1067.

National Cotton Council. (2010). History of the Eradication Program. *National Cotton Council of America*.

Onstad, D. W., Guse, C. A., Porter, P., Buschman, L. L., Higgins, R. A., Sloderbeck, P. E,. Peairs, F. B., and Cronholm, G. B. (2002). Modeling the development of resistance by stalk-boring lepidopteran insects (Crambidae) in areas with transgenic corn and frequent insecticide use. *Journal of Economic Entomology* 95(5): 1033-1043.

Parencia, C. R. (1968). Control of cotton insects with an insect-collecting machine. *Journal of Economic Entomology* 61: 274-279.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7(5): e37135.

Powell R. D., (1979). Fleahopper-induced infection causes cotton square drop. *Texas Agricultural Progress* 25(2): 17.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* 16(3): 608-627.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 81(3): 559-575.

Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. (2014a). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* 2:e431

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., and Bird, C. E. (2014b). Comment: demystifying the RAD fad. *Molecular Ecology* 23: 5937-5942.

R Core Team. (2018). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197(2): 573-589.

Ramalho, F. S. and Wanderley, P. A. (1996). Ecology and management of the boll weevil in South American cotton. *American Entomologist* 42(1): 41-47.

Raszick, T. J., Suh, C. P.-C., Ruiz-Arce, R., Dickens, C. M., Boratynski, T. N., Falco, M., Song, H., Kramer, Z. A., and Sword, G. A. (in preparation) Population genomic diversity of the boll weevil, *Anthonomus grandis* Boheman (Coleoptera: Curculionidae), and its variants in the United States, northern Mexico, and Argentina.

Raymond, M. and Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248-249.

Reinhard, H. J. (1926). Control of the cotton fleahopper. *Texas Agricultural Experiment Station Bulletin* 339: 7-34.

Reinhard, H. J. (1929). The value of spring emergence records on the cotton flea hopper, *Psallus seriatus*. *Journal of Economic Entomology* 22(5): 765-768.

Ring, D. R., Benedict, J. H., Walmsley, M. L., and Treacy, M. F. (1993). Cotton yield response to cotton fleahopper (Hemiptera: Miridae) infestations on the Lower Gulf Coast of Texas. *Journal of Economic Entomology* 86: 1811-1819.

Roehrdanz, R. L. (2001). Genetic differentiation of southeastern boll weevil and thurberia weevil populations of *Anthonomus grandis* (Coleoptera: Curculionidae) using mitochondrial DNA. *Annals of the Entomological Society of America* 94(6): 928-935.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145(4): 1219-1228.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8(1): 103-106.

Rummel, D. R. and Adkisson, P. L. (1971). A two-phased control program designed for maximum suppression of the boll weevil in the High and Rolling Plains of Texas. *Journal of Economic Entomology* 62: 1122-1123.

Scataglini, M. A., Confalonieri, V. A., and Lanteri, A. A. (2000). Dispersal of the cotton boll weevil (Coleoptera: Curculionidae) in South America: evidence of RAPD analysis. *Genetica* 108(2): 127-136.

Scataglini, M. A., Lanteri, A. A., and Confalonieri, V. A. (2006). Diversity of boll weevil populations in South America: a phylogeographic approach. *Genetica* 126(3): 353-368.

Sobrinho, R. B. and Lukefahr, M. J. (1983). Bicudo (*Anthonomus grandis* Boheman nova ameaca a cotonicultura Basileira, biologia e controle. *EMBRAPA*.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312-1313.

Sterling, W. L. (1978). Fortuitous biological suppression of the boll weevil by the red imported fire ant. *Environmental Entomology* 7(4): 564-568.

Sword, G. A., Lecoq, M., and S. J. Simpson. (2010). Phase polyphensim and preventative locust management. *Journal of Insect Physiology* 56: 949-957.

Tabashnik, B. E., (2008). Delaying insect resistance to transgenic crops. *Proceedings of the National Academy of Sciences* 105(49): 19029-19030.

Tabashnik, B. E., Gassmann, A. J., Crowder, D. W., and Carriére, Y. (2008). Insect resistance to Bt crops: evidence versus theory. *Nature Biotechnology* 26: 199-202.

Tabashnik, B. E., Gould, F., and Carrière, Y. (2004). Delaying evolution of insect resistance to transgenic crops by decreasing dominance and heritability. *Journal Evolution Biology* 17: 904-912.

Tapio, M., Miceikiené, I., Vilkki, J., and Kantanen, J. (2003). Comparison of microsatellite and blood protein diversity in sheep: inconsistencies in fragmented breeds. *Molecular Ecology* 12: 2045-2056.

Texas Boll Weevil Eradication Foundation (2018). 2017 program year end summary. Texas Boll Weevil Eradication Foundation, Inc. Available online at https://www.txbollweevil.org/program.html.

Tumlinson, J. H., Hardee, D. D., Gueldner, R. C., Hedin, P. A., and Minyard, J. P. (1969). Sex pheromones produced by male boll weevil: isolation, identification, and synthesis. *Science* 166: 1010-1012.

Warner, R. E. (1966). Taxonomy of the subspecies of *Anthonomus grandis*. *Annals of the Entomological Society of America* 59: 1073-1088.

Weir, B. S. and Cockerham, C. C. (1984). Estimating F‐statistics for the analysis of population structure. *evolution* 38(6): 1358-70.

Whitcomb, W. and Britton, L. A. (1953). The control of the cotton boll weevil *Anthonomus grandis* Boh. In Venezuela. *Empire Cotton Growing Review* 30: 177-181.

Wickham, H. (2016). Ggplot2: elegant graphics for data analysis. Springer-Verlag, New York.

Wood, D. E. and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15: R46.

Wright, S., (1943). Isolation by distance. *Genetics* 28(2): 114.

Wright, S., (1951). The genetical structure of populations. *Annals of Eugenics* 15: 323-354.

Zaller, J. G., Moser, D., Drapela, T., Schmöger, C., and Frank, T. (2008). Insect pests in winter oilseed rape affected by field and landscape characteristics. *Basic and Applied Ecology* 9: 682-690.

APPENDIX A

PROTOCOL FOR ISOLATION OF HIGH MOLECULAR WEIGHT DNA FROM BOLL WEEVIL AND COTTON FLEAHOPPER

For some numbers in the following protocol, there is a larger number followed by a smaller number in parentheses. The larger number is for the isolation of DNA from boll weevil, and the smaller number was used for the cotton fleahopper version of the protocol.

Equipment and Reagents not provided with Puregene Kit

1. 1.5 ml microcentrifuge tubes
2. 3 200 mL beakers with 2% bleach, ethanol, and pure water
3. Sterilized plastic pestles
4. Incubation block
5. Foam cooler
6. Liquid nitrogen
7. Pipets and pipet tips
8. Cold 100% isopropanol
9. Cold 70% ethanol
10. Microcentrifuge
11. Crushed ice
12. Glycogen Solution

Sample Preparation – Day 1

1. Preheat an incubation block to 55°C.
2. Sterilize plastic pestles and soft forceps using 1 minute soaks in 2% bleach, ethanol, and pure water.
3. Briefly blot excess ethanol from specimens on clean absorbent paper.
4. Add 1 whole specimen each to a labeled 1.5 ml microcentrifuge tube.
5. Freeze the micro centrifuge tubes in liquid nitrogen until bubbling subsides.
6. Homogenize each tissue using a sterilized plastic pestle.
    a. Add 600(100) ul of Cell Lysis Solution to each tube and remove the pestle, taking care to leave all tissue in the tube.
7. Add 1.5 ul of Puregene Proteinase K to each tube.
8. Mix by inverting 25 times.
9. Centrifuge for 5 seconds at 16,000 x g.
10. Incubate at 55°C for overnight or until tissue has completely lysed.
11. Invert tube periodically during incubation.

a. If tissue is not completely lysed after overnight digestion, add 1.5 ul of Proteinase K and incubate for another 3 hours.

DNA Extraction – Day 2

1. Preheat incubation block to 37°C. It is better to use a second heating block if possible, as a cold block will heat to 37°C much more quickly than a block at 55°C will take to cool down.
2. Add 3.0(1.0) ul of RNase A solution to the sample.
3. Mix by inverting 25 times.
4. Incubate for 30 minutes at 37°C.
    a. During this incubation, grind a full tray of ice and put it into the foam cooler for upcoming sample incubation.
5. Incubate on ice for 1 minute to quickly cool the sample.
6. Add 200(33) ul of Protein Precipitation Solution
7. Mix thoroughly by shaking vigorously for 20 seconds.
8. Centrifuge for 3 minutes at 16,000 x g.
9. Incubate for 5 minutes on ice.
    a. During this incubation, pipet 100 ul of isopropanol into a clean 1.5 ml microcentrifuge tube.
10. Centrifuge for 3 minutes at 16,000 x g.
11. You will see a protein pellet that has formed at the bottom of the sample. Without disturbing this pellet, carefully pour the supernatant from the sample into the fresh tube with the 600 ul of isopropanol.
12. Add 1.0(0.5) ul of Glycogen Solution to the sample.
13. Mix by inverting gently 50 times.
14. Centrifuge for 5 minutes at 16,000 x g.
15. The DNA may now be visible as a small white or nearly clear pellet at the bottom of the tube. Carefully pour off the supernatant, and drain the tube completely by inverting on a clean piece of absorbent paper, taking care that the pellet remains in the tube.
16. Let tube air dry for 5 minutes.
17. Add 600(100) ul of cold 70% ethanol to the sample.
18. Invert several times gently to wash the DNA pellet.
19. Centrifuge for 3 minutes at 16,000 x g.
20. Again, carefully pour off the supernatant, and drain the tube completely by inverting on a clean piece of absorbent paper, taking care that the pellet remains in the tube.
21. Allow tube to air dry for 20 minutes or dry using a vacu-fuge (10 minutes at 30°C).
    a. During this step, preheat an incubation block to 65°C.
22. Add 50(27) ul of DNA Hydration Solution.
23. Shake the sample vigorously for 5 seconds to mix.
24. Centrifuge for 15 seconds at 16,000 x g.

25. Incubate at 65°C for 1 hour.
26. Place on orbital shaker and shake gently overnight at room temperature. Ensure cap is tightly closed to avoid leakage.

Quality Control – Day 3

1. Use a spectrophotometer to obtain concentration and purity estimates.
2. Run sample on an electrophoresis gel to determine relative integrity of DNA.