GENETIC POPULATION STRUCTURE OF *LYGUS HESPERUS* IN THE TEXAS HIGH PLAINS

Ram B. Shrestha
Megha N. Parajulee
Stanley C. Carroll
Texas A&M AgriLife Research
Lubbock, TX

Abstract

Insect pest population genetic diversity information is essential to understanding variation in morphological, behavioral or biological characteristics, and predicting various ecological processes in pest populations. *Lygus hesperus* is an emerging insect pest of cotton in Texas. Texas High Plains insect pest genetic diversity information is lacking. A molecular population genetic diversity study was designed to determine the population genetic structure of *L. hesperus* in the region. *L. hesperus* samples (n=48) were collected from four different locations, each 40 to 65 miles apart, spanning 153 miles on a north-south axis across the Texas High Plains, with a north-to-south land elevation gradient from 3,615 feet to 2,992 feet above sea level. *L. hesperus* samples were genotyped by Polymerase Chain Reaction (PCR) and Polyacrylamide Gel Electrophoresis (PAGE) using six previously developed and characterized microsatellite markers. This study demonstrated that microsatellite markers are useful molecular tools for study of neutral genetic variation of *L. hesperus* populations in the Texas High Plains. Within the 153-mile geographic sample range, *L. hesperus* was genetically differentiated into two distinct populations, constituting northern and southern population clusters. It is hypothesized that genetic differentiation results from differences in the ecological environment such as hosts, habitat, and others; possibly due to the 623 feet north-south elevation gradient. A larger-scale landscape genetics study is planned to determine the relationships between the ecological parameters and Texas High Plains *L. hesperus* population genetic diversity.

Introduction

Lygus hesperus (Hemiptera: Miridae) is an ecologically and economically important insect in the Texas High Plains. *Lygus* spp. in the U.S. are highly polyphagous insects with a documented broad host range of more than 300 plant species. They are serious pests of many fruits, vegetables, forest trees, pasture crops, ornamental crops, cereals, legumes, and oilseed crops (Kelton 1975, Snodgrass *et al.* 1984). *Lygus* has been reported in 26 different roadside weed hosts in the Texas High Plains (Parajulee *et al.* 2003). Previous 10-year average data show that the *Lygus* complex was the second most damaging pest of cotton in the United States and the number one damaging pest of cotton in 2008 (Williams 2000-2009). In addition to being classified as crop pests, *Lygus* spp. serve as predators of small insects and bollworm eggs, and have also been reported as vectors of bacterial and fungal diseases of crop plants. Their eggs and nymphs serve as prey/hosts for many predators and parasitoids. Because *Lygus* serve multiple ecological roles, *they* should be considered an important ecological player rather than only as a field crop pest.

A previous field survey revealed that *L. hesperus* is the dominant *Lygus* species, comprising >95% of the *Lygus* population complex in the Texas High Plains (Parajulee *et al.* 2003). Morphological study of *L. hesperus* samples collected from different States has shown great intra-specific morphological variation (unpublished data). It is hypothesized that there are intra-specific differences in morphology, behavior, or biology between geographically distributed or host-specific *L. hesperus* populations within the Texas High Plains region. If there are biological differences between geographically distributed populations within a region or landscape, then regional or landscape-level pest management strategies need to be evaluated to develop a population-specific management strategy. Conducting research to compare morphological, behavioral, and biological aspects of geographically distinct populations is challenging. To ease this challenging task, the overall *L. hesperus* population needs to be divided into manageable population structures or management units. This can be done via genetic diversity analysis.

Genetic diversity becomes important when determining management units or sub-populations within species which are difficult to differentiate morphologically. Population genetic diversity studies can also shed light on how a sub-population is interacting and evolving. A variety of approaches can be used to study genetic diversity, including protein markers (allozymes), various gene sequences (mitochondrial or other functional genes), neutral markers like microsatellites, Inter Simple Sequence Repeat (ISSR), random markers such as Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), or Restriction Fragment Length Polymorphism (RFLP), and others. The microsatellite markers are also known as Simple Sequence Repeats (SSR), one of the more

common molecular markers used for genetic diversity, phylogeographic, and many molecular ecology studies. Advantages of microsatellite markers include genome-wide abundance in eukaryotes, co-dominant nature, fast evolution, high polymorphism, and inheritance from both parents. Development of new microsatellite markers is a time-consuming and costly process, but once the marker is characterized, it is easy to use in fingerprinting or genotyping. Microsatellite markers have many advantages over other markers like mtDNA and allozymes.

Though *L. hesperus* is widely distributed across North America, its genetic diversity is largely unknown. One of the reasons for this information gap is a lack of molecular markers developed for this species. A previous study reported on the development of microsatellite markers or SSR markers for *L. hesperus* which proved to be useful for genetic diversity studies of this species (Shrestha *et al.* 2007). Since microsatellite markers for this species have already been developed and characterized, they can be readily and easily used in a larger-scale *L. hesperus* genetic diversity study. The goal of this study was to determine the regional level *L. hesperus* genetic diversity, however; in this study, discussion is confined to *L. hesperus* genetic diversity in the Texas High Plains region. Specifically, the objective of the study was to determine the genetic population structure of *L. hesperus* in the Texas High Plains.

Materials and Methods

Lygus sample collection and identification

L. hesperus samples were collected from various hosts (mostly from alfalfa, kochia and Russian thistle) at four Texas locations: Lamesa (Dawson County), Idalou (Lubbock County), Plainview (Hale County), and Happy (Swisher County). These Texas High Plains locations were selected along a north-south axis along Interstate Highway 27 (Fig. 1). The north-south sampling axis spanned roughly 150 miles, spanning the region, which is commonly characterized as being virtually the world's largest contiguous patch of cotton production. The elevation gradient of the sampling area ranged from 3,615 feet above sea level in the north (Happy, TX) to 2,992 feet in the south (Lamesa, TX). *Lygus* adult insect samples were collected using sweep nets and preserved by drying at room temperature until DNA extraction. Each *L. hesperus* sample was observed under a stereo microscope, and distinguished from other species for identification to species level.

DNA extraction

Among the four different DNA extraction protocols tested, the MasterPure™ Complete DNA and RNA Purification Kit method performed the best for *L. hesperus* genomic DNA extraction (Bastola *et al.* 2009); therefore this was the selected method for this study. *L. hesperus* genomic DNA from 48 randomly selected individuals from each site was extracted using a MasterPure™ Complete (MPC) DNA and RNA Purification Kit (EPICENTRE® Biotechnologies). The DNA extraction protocol provided by the kit manufacturer was followed. Whole *Lygus* insects were pulverized briefly in 300 µl MPC Tissue and Cell Lysis Solution using a bead-beater machine, after which 5 µl of Proteinase K from 25 ng/µl stock was added to each sample. Samples were then incubated at 65°C for fifteen minutes, after which 175 µl of MPC Protein Precipitation Reagent was added. Samples were then precipitated with 500 µl of 100% Isopropanol. DNA quality and quantity were assessed both by optical photospectrometry and agarose gel electrophoresis. Then, samples were diluted to make 50ng/µl and stored at -20°C until used in genotyping by Polymerase Chain Reaction (PCR).

Genotyping and visualization

Lygus samples were genotyped by PCR amplification of the targeted SSR sequence using six selected SSR probes (Table 1). Standard PCR protocol was followed with SSR probe specific temperature profile for PCR amplification. The alleles were determined by vertical polyacrylamide gel electrophoresis of PCR amplification products (1 mm thick) as described by Wang *et al.* (2003). Then the allele bands were visualized by fast silver staining procedure as described by Benbouza *et al.* (2006). The gels were scored manually as co-dominant marker.

Data analysis

The SSR data were analyzed to determine an average number of alleles per locus, the expected and observed heterozygosis of the polymorphic loci, and Shannon's Diversity Index (I) of the SSR markers. Deviations from Hardy–Weinberg Equilibrium (HWE) (Guo and Thompson 1992) were tested for each locus and population using ARLEQUIN® software (Schneider et al. 2000); with 1000 permutations. The Linkage Disequilibrium (LD) was calculated using POPGEN® software (Raymond and Rousset 1995a). Expected and observed heterozygosity for each locus and the genetic diversity indices or fixation indices (F statistics) were calculated. The degree of population differentiation was calculated at three hierarchy levels using Analysis of Molecular Variance (AMOVA).

Fixation indices, F_{IS} (heterozygosity among individuals within a population), F_{ST} (heterozygosity due to population subdivision), and F_{IT} (heterozygosity due to total population) were calculated as described by Hartl and Clark (1997). The genetic distances matrix for all pair-wise population comparisons was calculated using Nei and Li's distances. Then an Unweighted Pair Group Method with Arithmetic (UPGMA) cladogram of populations was constructed using the genetic distances (Sneath and Sokal 1973). Population genetic structure was determined by Bayesian cluster analysis using Structure® 2.3.2 software (Pritchard *et al.* 2000) using multilocus genotype data of *L. hesperus* individuals. The microsatellite allele information of 192 individuals and 6 loci were used to run 10 iterations of admixture ancestry model simulation with 10000 long burn-in periods, 10000 MCMC reps after burn-in, 1-6 K-population clusters.

Results and Discussion

Marker characters

The microsatellite markers used in this study showed high allele diversity with average Shannon diversity index (I) value of 1.04 and a 4.1 average number of alleles per locus (Table 2). The average heterozygosity of the 6 loci was 0.57 and allele frequency in all loci was significantly different from the Hardy-Weinberg's equilibrium. These six microsatellite markers revealed the intra-specific genetic variation among the *Lygus hesperus* individuals within landscape level, thus these markers can be used to study the large scale regional level phylogeographic studies in the future.

Population genetic diversity

The exact test (Raymond and Rousset 1995b) of population genetic differentiation, using *L. hesperus* haplotype information, revealed that *Lygus* samples from different locations were significantly differentiated (Table 3). The Analysis of Molecular Variance (AMOVA) showed that 88.7% of the total genetic variation contributed by among individuals within a population, and 11.3% of the total genetic variation contributed by among individual between population genetic variations of *L. hesperus*. The overall population genetic diversity (F_{ST}) value was 0.113 (P=0.000001*).* The genetic analysis indicated that *L. hesperus* from Texas High Plains regions is not a randomly mating single population, but rather genetically differentiated populations. The allele frequencies of *L. hesperus* geographical populations were significantly deviated from the HWE which indicates the *L. hesperus* population might have gone through some natural genetic selection process. The individual based genetic population structure analysis was conducted to further evaluate the genetic differentiation result.

Genetic population structure

Pair-wise Nei's genetic distances between different geographic populations of *L. hesperus* were calculated and used to develop the (UPGMA) mean cladogram (Fig. 2). The southern most populations of *L. hesperus* from Lamesa and Idalou belonged to a single clade and *Lygus* from the more northern locations (Happy and Plainview) were in two separate clades. The linear regression analysis of pair-wise genetic distance and the geographical distances between *L. hesperus* showed no significant relation between the geographical distance and the genetic differences of *L. hesperus,* but when the intercept was forced to zero, the regression analysis showed a low level of positive relationship (Fig. 3). As the geographical distance increased from 73 miles towards 153 miles, the genetic distance also increased from 0.05 to 0.12. Further analysis of the additional data (from more locations and with a higher number of microsatellite markers) is needed to develop a better conclusion on whether there is any relation between the geographical distance and the genetic differentiation of *L. hesperus.* The genetic population structure was also analyzed by Bayesian cluster analysis using Structure® 2.3.2 software. The admixture model based analysis showed that the probability of having two population clusters (K=2) given the data, the maximum log-likelihood [i.e., *Ln P (D)*], was highest (i.e., lowest negative) at K=2 (Fig. 4). Therefore, it is most likely that *L. hesperus* populations from the Texas High Plains have at least two relatively distinct genetic populations. The bar chart of Q proportion (probability of individual assigning into one or another population cluster) showed clearly there were two genetically distinct *L. hesperus* populations (Fig. 5). Further analysis of the population sub-structure of each geographic sample revealed that the more northern population (Happy, TX) was obviously different from the southern Lamesa population (Fig. 6).

Summary

The SSR markers previously developed were useful for this genetic diversity study because they revealed the intra-specific genetic population structure of *L. hesperus* within a relatively small geographic area (within a landscape

level) even when only six SSR markers were used. *Lygus hesperus* populations from the Texas High Plains were genetically differentiated into northern and southern populations. Further research is needed to understand the biological relevance of these genetic differentiations. The genetic difference may be associated with or implies to various important biological traits of *L. hesperus* such as biological fitness, fecundity, life table parameters, diapause physiology, photoperiodism, pesticide resistance, host selection behavior, immunity to microbes, and resistance to parasitoids. The hypothesis of association of these traits with this genetic population structure needs to be further evaluated. With the current population structure information, strong management recommendations cannot be made, but it might be wise to be careful in recommending or applying single management practices to both northern and southern *L. hesperus* population because they may respond differently to different insect management practices. We desire that further research on the phylogenetic evolutionary relationships and phylogeographic history of *Lygus* from different regions of the U.S. could be accomplished so that scientists can better understand how the genetic differentiation of *L. hesperus* might have evolved.

Acknowledgments

This project was partially funded by Cotton Incorporated Core Program, USDA CSREES RAMP, International Cotton Research Center, and Plains Cotton Growers, Inc.

References

Bastola, A., M. N. Parajulee, and R. B. Shrestha. 2009. Evaluation of different methods of nuclear DNA extraction for *Lygus hesperus* samples preserved in various storage conditions, pp 1159-1162. *In* Proc., Beltwide Cotton Conf., National Cotton Council, Memphis, TN.

Benbouza, H., J. M. Jacquemin, J. P. Baudoin, and G. Mergeai. 2006. Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. Biotech. Agron. Soc. Environ. 10:77-81.

Guo, S., and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372.

Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics. 3rd ed. Sinauer Associates, Sunderland, MA.
Kelton, L. A. 1975. The *Lygus* bugs (Genus *Lygus* Hahn) of North America (Heteroptera: Miridae). V. R. Vickery (ed.). *In* Memoirs of the Entomol. Soc. of Canada- No 95.

Parajulee, M. N., M. D. Arnold, S. C. Carroll, A. M. Cranmer, R. B. Shrestha, and P. L. Bommireddy. 2003. *Lygus* abundance on wild hosts: A survey across the Texas High Plains, pp. 970-973. *In* Proc., Beltwide Cotton. Conf., National Cotton Council, Memphis, TN.

Pritchard J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Raymond, M., and F. Rousset. 1995a. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. J. Heredity 86:248-249.

Raymond, M., and F. Rousset. 1995b. An exact test for population differentiation. Evolution 49:1280-1283.
Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: Software for population genetics data analysis. User manual ver. 2.0. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Geneva.

Shrestha, R. B., M. N. Parajulee, O. P. Perera, B. E. Scheffler, and L. D. Densmore. 2007. Characterization of microsatellite loci in the western tarnished plant bug, *Lygus hesperus* Knight (Hemiptera: Miridae). Mol. Ecol. Notes, 7:1342-1344.

Sneath, P. H. A., and R. R. Sokal. 1973. Numerical Taxonomy. Freeman and Co., San Francisco, 573 pp.
Snodgrass, G. L., W. P. Scott, and J. W. Smith. 1984. An annotated list of the host plants of *Lygus lineolaris* (Hemiptera: Miridae) in Arkansas, Louisiana, and Mississippi Delta. J. Ga. Entomol. Soc. 19:93-101.

Wang, D.., J. Shi, S. R. Carlson, P. B. Cregan, R. W. Ward, and B. W. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. Amer. 43:1828-1832.

Williams 2000-2009. Cotton insect losses-1999-2008. *In* Proc., Beltwide Cotton Conf., National Cotton Council, Memphis, TN.

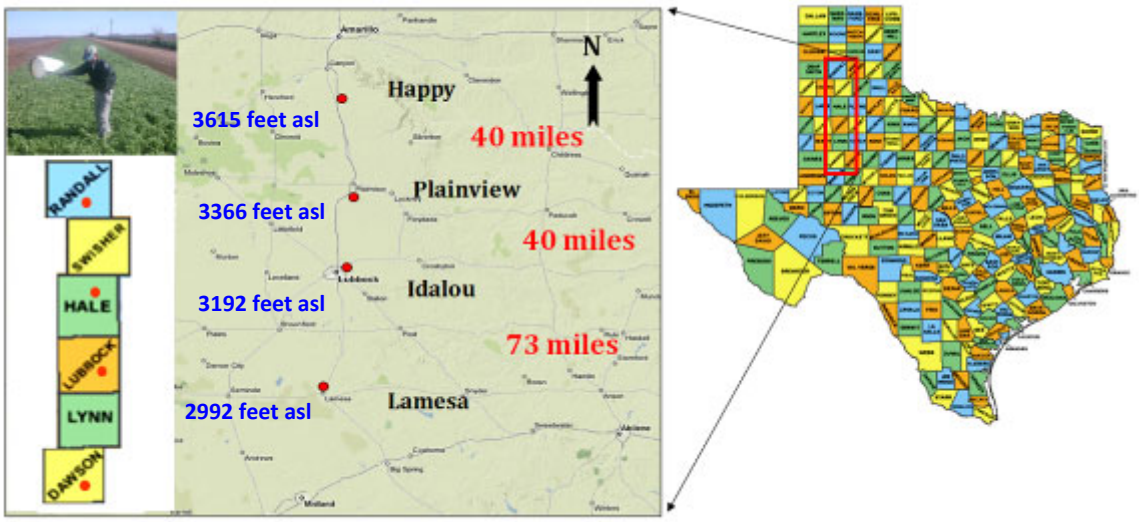

Figure 1. *Lygus hesperus* sample collection sites in the Texas High Plains. The altitude of each site is given in feet Above Sea Level (asl).

Table 1. Simple sequence repeat (SSR) primers used for genotyping *Lygus* population and their target repeat sites.

Locus	SSR Primers	Repeats	Annealing T (°C)
LhMS1	F_AAGAAACTGCTAGGAGGGGAAATG R_CCTCGTCTTCAATTCCTGTTCTTC	(AAG)4	55
LhMS2	F_AAGATTTTGATCGGACTTCGAGTG R_GTCCCTGGCAACAATGTTATCTTC	(GAA)7	55
LhMS3	F_ACTATGGGTTATGGGATGGGATCT R_AAGCAAAAGGGAGATTTCAAGAGG	(GT)6	63
LhMS4	F_ACTGTGGTTTTGTCCAGGATTGTC R_ATCAAAATCCCAAAATCGAACAGA	(ACA) 5	63
LhMS5	F_GTTTCAAAGTAAAGCGTGGACGAG R_AAATGCATGCACCTAAGTATGGGT	(CA) 16	55
LhMS6	F_GATCACTCCACTCACCTTGATCCT R_CGACGAACAGCGTTTAGCAATTAT	(AAC) 8	63

Table 2. Characteristics of the six SSR markers used in genetic diversity study of *Lygus hesperus.*

Locus	**# Alleles**	**Shannon (I)**	**H_0**	**H_e**	**HWE (*P*)**
LhMS1	4	0.49	0.27	0.27	0.2030
LhMS2	3	0.79	0.51	0.49	0.2204
LhMS3	6	1.31	0.77	0.65	0.0005
LhMS4	4	0.88	0.26	0.53	0.0000
LhMS5	4	1.25	0.61	0.69	0.0000
LhMS6	5	1.55	0.84	0.78	0.0000
Mean	4.1	1.04	0.54	0.57	-
St. Dev.	1.16	0.39	0.25	0.18	-

Table 3. Haplotype based genetic differentiation between 4 geographical populations of *Lygus hesperus.* (α=0.05).

	Lamesa	**Idalou**	**Plainview**	**Happy**
Lamesa		0.12227	0.03412*	0.00003**
Idalou	0.12227		0.6198	0.00224**
Plainview	0.03412*	0.6198		0.004**
Happy	0.00003**	0.00224**	0.004**	

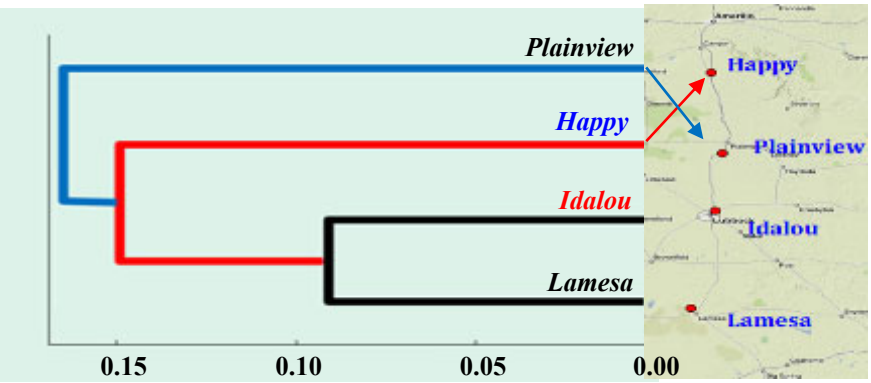

Figure 2. UPGMA Cladogram of different geographic population of *Lygus hesperus* from Texas High Plains based on Nei's genetic distances.

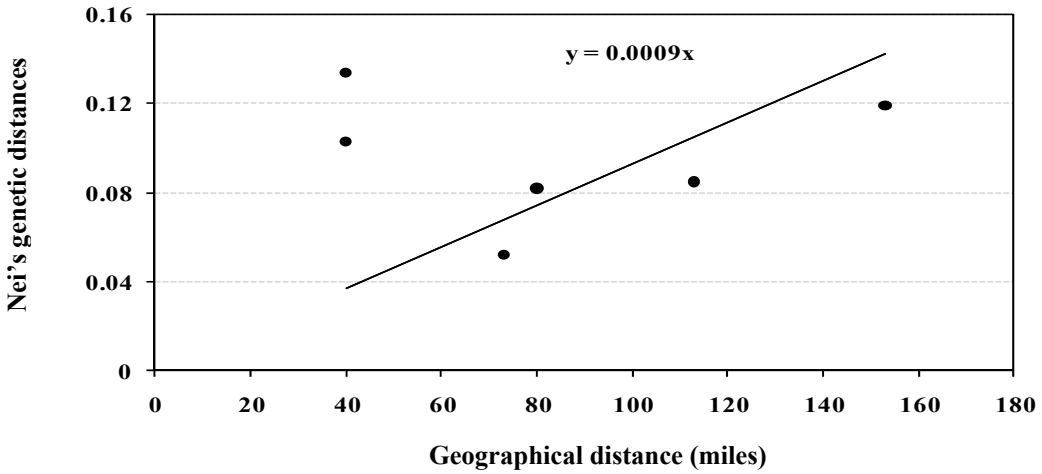

Figure 3. Relationships between the Nei's genetic distances and geographical distances of the *Lygus hesperus* samples.

Figure 4. Average log-likelihood value, *Ln P (D),* of various K-mean clustering analysis using Structure® 2.3.2 software.

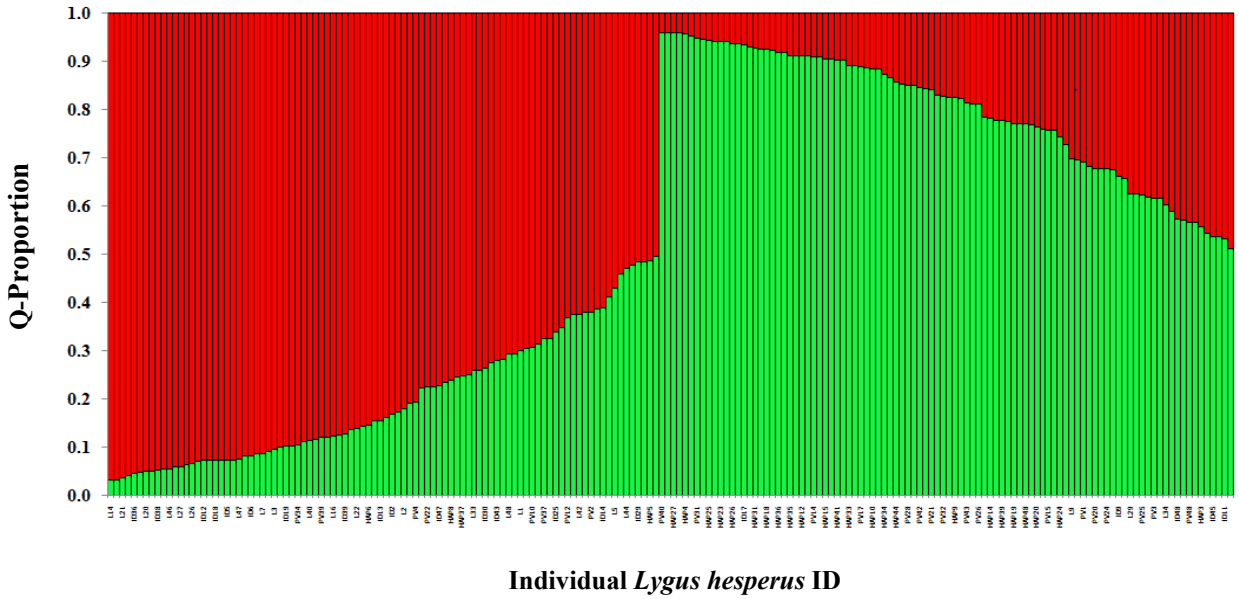

Figure 5. Genetic population structure of *Lygus hesperus* from the Texas High Plains determined by Bayesian maximum likelihood analysis of microsatellite data using Structure® 2.3.2 software.

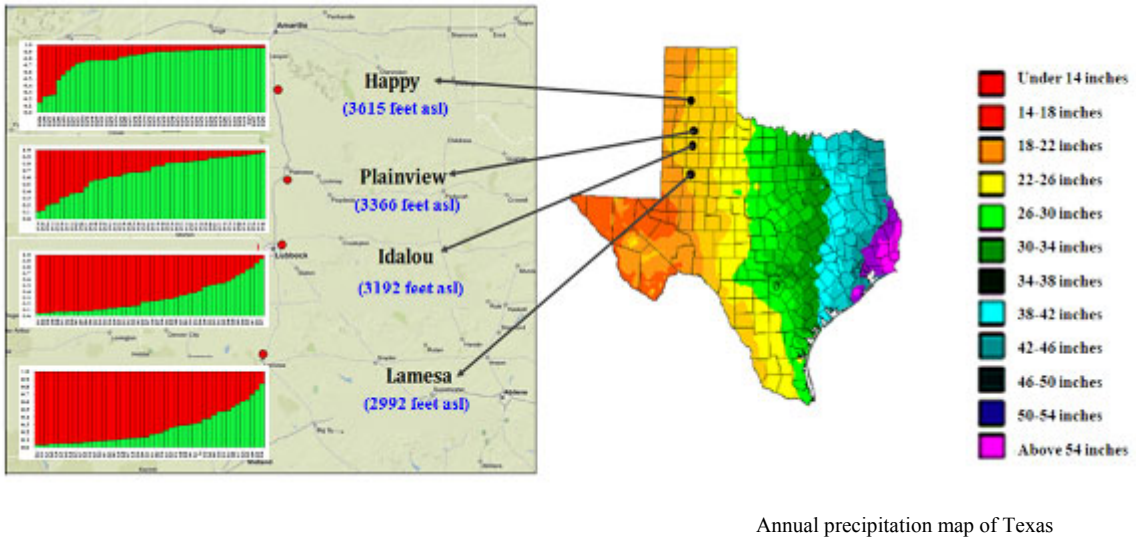

Figure 6. Genetic population sub-structure of each geographical population of *L. hesperus* from the Texas High Plains as determined by Bayesian maximum likelihood analysis of microsatellite data using Structure® 2.3.2 software. The altitude of each site is given in feet Above Sea Level (asl).