GENETIC STRUCTURE OF THE PINYON PINE BEETLE, IPS CONFUSUS (LECONTE) DURING AN OUTBREAK

By

Liu Yang

A THESIS

Submitted to Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Entomology

2012

ABSTRACT

THE GENETIC STRUCTURE OF THE PINYON PINE BEETLE, IPS CONFUSUS (LECONTE) DURING AN OUTBREAK

BY

Liu Yang

Genetic structure of phylophagous insects are formed under many factors, such as co-evolutionary effect with hosts, geographic distribution, or migration which is impacted by climatic fluctuations or natural disturbances. To investigate the impact of 2003 pinyon pine beetle outbreak on its genetic structure, we sampled in total 244 individuals from 28 populations across six states in Southwest of United States in 2001 and 2003, constructed a phylogenetic tree, compared genetic diversity within each populations before and during outbreak, calculated genetic differentiation among populations, tested genetic variations on different hierarchical levels, and performed mantel tests to test isolation-by-distance. The diversity analysis and haplotype network did not demonstrate significant differences among populations before and during outbreak. Thus the outbreak had little impact on the genetic structure of *Ips confusus*. Spatial patterns of haplotype distribution, diversity trend, AMOVA and Mantel tests indicated that the genetic structure was closely associated with geography. These results suggest that multiple short-distance dispersals among proximal populations rather than dispersal among distant populations, have shaped the genetic structure of *I. confusus* despite greater potential for long distant dispersal during outbreaks.

ACKNOWLEDGEMENTS

I thank Dr. Anthony Cognato for providing the sequences of *Ips confusus*, tutoring me with phylogeny and haplotype network constructions, and helping edit this thesis; I also thank Dr. Mike Kaufman for tutoring me with PCA analysis and Dr. Kim Scribner for advices with statistical analyses. Amanda Lorenz, Nick Barc and Rachel Olson provided valuable comments and great friendship during the whole time of writing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Insect populations often experience abnormal growth including the sudden increase in the number of populations and dispersal events during outbreaks (Christiansen *et al.* 1987). As food sources are depleted emigration of individuals to new food sources is expected. This dispersal has potential genetic consequences for the metapopulation (Vandergast *et al.* 2004; Eckert *et al*. 2008). A new population founded by many individuals from a nearby outbreak would retain much of the genetic variation found in the source population (Ibrahim *et al.* 1996). However, dispersal of few individuals to distant resources, without an established population, would result in a genetic bottleneck given that additional emigration to the new population would be likely rare. Empirical data documenting the effect of outbreaks the on genetic structure of pest species are limited to a few studies (e.g. Ibrahim *et al.* 2000, 2001; Chapius *et al*. 2008, 2009; Fonseca *et al*. 2010; James *et al*. 2011; Kobayashi *et al*. 2011; Ronnas *et al.* 2011; Tao *et al.* 2012). The majority of these studies suggest that outbreaks do not result in an increase in gene flow among nearby populations (James *et al.* 2011; Kobayashi *et al*. 2011; Ronnas *et al.* 2011). This observation may depend on the species and the regional scope of the study. For example, homogenizing effect of outbreak events on genetic variation (lower population differentiation in outbreak than non-outbreak) was not observed in worldwide populations of migratory locust (Chapius *et al*. 2008) however at the regional scale, homogenizing effect was found on the population structure among local populations (Chapius *et al*. 2009). Thus, the additional study of species prone to outbreaks could potentially reveal common genetic consequences of epidemics. A recent outbreak of a bark beetle, *Ips confusus* (LeConte) presents opportunity to document this species population genetic structure before and

during an outbreak.

Ips confusus occurs in Western US including Arizona, California, Colorado, Nevada, New Mexico, Oregon, Texas, Utah, Wyoming and approximately overlaps with the distribution of pinyon pines (Wood 1982). This bark beetle, as adults and larvae, mainly feeds on two pine species, *Pinus edulis* Engelman and *Pinus monophylla* Torrey & Fremont and they are considered as host specific despite their occasional use of other conifers (Lanier 1970; Cognato *et al*., 2003). The beetle's lifecycle is dependent on the tree. Colonization of a suitable host usually starts with the male beetles excavating an entrance tunnel followed by a nuptial chamber. While the male bores in the tree, it produces pheromones that attract more conspecifics and thus initiates a mass attack (Wood *et al*., 1967). Two- five females join the male in the nuptial chamber, mate, and each construct a gallery in which they lay eggs. After the eggs hatch, the larvae feed on inner bark, until they pupate under the bark and emerge (Furniss and Carolin 1977, Wood 1982, Eager 1999, Negron and Wilson, 2003). *Ips confusus* usually has three, sometimes four generations a year. It is not uncommon that after first emergence, the parent beetles with re-infest the tree and produce similar size broods (Wood 1982).

Ips confusus is not the most aggressive bark beetle species. It rarely feed on healthy trees and usually attacks stressed or dying individuals. Nevertheless, outbreaks can occur during times of severe drought or some other natural disaster (Furniss and Carolin 1977; Negron and Wilson, 2003). Recently, a large outbreak of *I. confusus* occurred during a prolong drought (2001–2004) in Arizona, New Mexico, Nevada, Utah, and southwestern Colorado. Ecological damage and economical loss were extensive. In 2003, during peak outbreak, approximately 15–30% of pinyons were killed throughout >1.6 million hectares

(USDA–Forest Service 2004, Breshears *et al*. 2005, Williams *et al*., 2010). By 2007, beetle populations and pinyon mortality declined to endemic levels (Williams *et al*., 2010; Halsey *et al*., 2011).

Population genetics for *I. confusus* was characterized for 100 individuals from ten populations collected in 2001 at the beginning of the outbreak (Cognato *et al*. 2003, Halsey *et al*., 2011). Using partial mitochondrial COI DNA sequences, Cognato *et al*. (2003) revealed much haplotype diversity, which was partitioned into two clades corresponding to geographic regions (California and the Rocky Mountains). These clades likely developed in Pleistocene refugia. Interestingly, of the 15 observed haplotypes, 11 were unique to a population and distributed among Californian and Rocky Mountain populations. This amount of haplotypic endemism appears typical for bark beetles (Cognato *et al.* 1999, Stauffer *et al*. 1999, Cognato *et al*. 2005a,b, Menard and Cognato 2007). Gene flow among *I. confusus* populations was recurrent throughout the species history however the amount of gene flow among present day populations is not known (Cognato *et al*. 2003). Bark beetles in general are able flyers and *Ips* species have a maximum potential of 50 km/generation unaided by wind or human transport (Jactel and Gaillard 1991). Dispersal ability is influenced by environmental conditions especially wind (Gara and Vite 1962; Byer 2000). Air currents can potentially carry beetles hundreds of kilometers (de la Giroday *et al*. 2011, Safranyik *et al*. 2010) and consequently cause genetic bottlenecks as observed for the mountain pine beetle (James *et al.* 2011). Hence, during the outbreak of *I. confusus* long distance dispersal may have increased due to the greater numbers of individuals and if so, the rare haplotypes observed in Cognato *et al*. (2003) could have potentially become fixed in other locations.

This study investigates the diversity of mitochondrial cytochrome oxidase I haplotypes among *I. confusus* individuals collected before and during a region-wide outbreak. We test the hypothesis that *I. confusus* haplotypes are not fixed for populations thus indicating little barrier to gene flow. Also, we hypothesize that rare haplotypes observed at the beginning of the outbreak are rare among populations sampled during the outbreak.

Methods and Materials

Approximately 10 live *Ips confusus* adult beetles were collected from host trees (*P. edulis* and *P. monophylla*) of 28 localities across the six states in the Southwest of United States (CA, NV, UT, CO, AZ and NM) (Figure 1, Table 1a). Each beetle was removed from the tree by using a knife and forceps, and in total 267 individuals were collected and stored in 100% ethanol for later use (details in Cognato *et al*., 2003).

Total genomic DNA was extracted from beetle thoraces, with a silica-based spin column procedure, following the manufacturer's tissue protocol as described in Cognato *et al.*, 2003. Mitochondrial cytochrome oxidase I DNA (399bp) was amplified for total 267 Individuals via polymerase chain reaction (PCR) with primers C1-J-2183 and C1–N-2611 following the methods of Cognato *et al*. (2003). Each PCR reaction consisted of: 35 ul ddH2O, 5 ul 10X TaqDNA polymerase buffer (Promega, Madison, WI), 4 ul 25 mM Promega MgCl2, 1 ul 40 mM deoxynucleotide triphosphates (dNTPs), 2 ul of each 5 mM oligonucleotide primer, 0.2 ul of Promega TaqDNA polymerase, and 1 ul of DNA template. The PCR was performed on a thermal cycler (MJ Research, Cambridge, MA) under the

following conditions: one cycle for 3 min at 95℃, .75min at 45℃, 1 min at 72℃, followed by 34 cycles of .5min at 94℃, .75 min at 45℃, 1 min at 72℃, and a final elongation cycle of 5 min at 72℃.

Unincorporated dNTPs and oligonucleotides primers were removed from PCR with a Qiaquick PCR Purification Kit (Qiagen) following the manufacture's instructions and were directly sequenced on an ABI 377 automated sequencer as described in Coganto *et al.* (2003). Both sense and antisense strands were sequenced for all individuals. Consensus sequences were arranged and inspected for nucleotide ambiguities in Sequencher (Ann Arbor, MI) and resulted in 244 sequences that were used for subsequent analyses. Sequences were complied in MacClade (Maddison& Maddison 2005) and submitted to GenBank (upon publication).

We used parsimony and Bayesian analyses to investigate potential phylogenetic signal among the beetle mitochondrial haplotypes. PAUP* ver. 4.0b10 (Swofford 2003) was used to conduct the parsimony analysis with the following settings: heuristic searches with 10 replicates and tree-bisection-reconnection branch swapping. We attempted a Bayesian analysis of these data (MRBAYES3; http:// morphbank. ebc. uu. Se/ mrbayes/) using a time reversible (GTR + C + I) model; four Metropolis-Coupled Markov chain Monte Carlo searches (one cold, three heated) which were run twice simultaneously for 20 million generations, each with sampling every 100th iteration. These runs did not complete or approach to stationarity after two weeks of computation. Hence, Bayesian analysis was not pursued.

A haplotype network was created with TCS with default algorithm (Clement *et al*., 2000) because little resolution found in the parsimony tree. Ambiguous characters were

considered as missing data and a limit of nine mutational steps were considered for the 95% plausible set of alternative parsimony networks (Clement *et al*., 2000).

Molecular diversity for mtDNA sequences was analyzed by estimating haplotype diversity (H) (Nei 1981) and nucleotide diversity (π) (Nei 1987)in two groups: populations (1-10) from 2001 and populations (11-28) from 2003. Haplotype diversity (H) was calculated as $H=n/(n-1)(1-\sum p_i^2)$ where n is the number of gene copies in the population and pi is the frequency of the ith haplotype (Nei, 1987). For codominant marker, it is the same formula for calculating expected heterozygosity; nucleotide diversity (π) measures the average nucleotide differences between all pairs of DNA sequences randomly chosen from the population. It is calculated as $\pi = n/(n-1)(\Sigma x_i x_j d_{ij})$ where n is the sample size, xi and x_j are the frequencies of haplotype i and j, and d_{ij} is the fraction of the number of nucleotide differences between two haplotypes out of total nucleotide number per haplotype (Tajima, 1993; Excoffier *et al*., 2005). Besides haplotype diversity and nucleotide diversity, the number of unique haplotypes, the number of pairwise differences, and their means and confidences were calculated (Table 2). The number of polymorphic sites in each population was calculated and the loci were identified (Table 1a). We calculated the means and variances of the pre-outbreak populations and the during-outbreak populations on molecular diversity indices (Table 2) and haplotype frequencies (Table 3), and used the t-tests (on means) and F-tests (on variances) to compare the means and variances between pre and during-outbreak in order to access the effect of the outbreak (see the result). We also investigated the relationship between haplotype diversity (H) and latitudes by plotting the regression graphs (Figure 4).

Gene flow and genetic drift usually result in changes in species' spatial genetic structure which can be assessed by measuring the changes of haplotype frequencies before and during the outbreak. We calculated the haplotype frequencies for each haplotype in each populations, then converted the haplotype frequencies into percentages and log-ratio transformed, which were treated as variables in the Principal Component Analysis (PCA). In PCA, the variables were transformed into lower dimensional space (in our case, three dimensional, thus three Principal Components PC1, PC2 and PC3). The PC scores were produced under each category (geographic regions and time) using software JMP (a SAS product) (Table 4). PCA created axes from the variables and assigned them along the axes, so to explain the distribution of sample values (Figure 3). Then we conducted a Multivariate analysis of variance (MANOVA) of the values categorized by pre-outbreak, during-outbreak and geographic populations (Table 5).

The genetic structure among populations was analyzed by computing the hierarchal analysis of molecular variance (AMOVA) based on estimated Fsts, the exact test of population differentiation, and a Mantel test with the software, Arlequin ver. 3.0 (Weir and Cockerham 1984; Excoffier *et al*., 2005).

AMOVA test was initially performed on different levels of genetic variation and associated F-statistics for testing corresponding significance levels. The total variance (σ^2) was partitioned into covariance components(σ^2_a, σ^2_b and σ^2_c) due to differences among groups, differences among populations within group and differences within populations, respectively (Rousset 2000; Excoffier *et al.*, 2008). Since the same framework could be extended to the fixation index FST, which is identical to the F-statistics over loci (Michalakis

and Excoffier 1996), the significance of F-statistics was tested to interpret the significance of the fixation indices, by using non-parametric permutation approach with 1,000 iterations (Excoffier *et al*., 1992). In our study, the hierarchical variation analysis (AMOVA) was first conducted in one group consisting entire 28 populations 42 haplotypes, and thus σ^2_a and FST were tested by permuting haplotypes among populations; then the AMOVA test was conducted in several smaller groups with different combinations of populations, σ^2_c and within populations (FST), σ^2_b and among populations within groups (FSC), and σ^2_a and among groups (FCT) were tested respectively in each case (Table 6. a-c)

The exact test of population differentiation was conducted to test whether populations were significantly different from each other by comparing the pairwise genetic distances (Excoffier *et al*., 2005). This exact test was designed to test the null hypothesis of random distribution of *k* haplotypes among *r* populations (*k*=42, *r*=28 in our case) (Raymond and Rousset 1995), which was extrapolated from Fisher's exact test of 2x2 contingency table. The P-value was calculated by summing up the probabilities of all contingency tables that have same or smaller probabilities and with same row and column sums (Raymond and Rousset 1995).

Mantel tests were performed to test for correlation between genetic and geographic distances by evaluating the correlation coefficient (r) and the statistical significance (P-value) (Mantel 1967; Sokal & Rohlf 1995). The genetic distances between populations were estimated as Fst/(1-Fst) (Slatkin's Distance) (Slatkin 1995) under the Tamura & Nei's substitution model (Tamura & Nei 1993), at the permutations of 5000, significant level of 0.05, and Gamma value of 0. We created the geographic distance matrix by calculating the

great-circle distances among the 28 populations using the on-line geographic distance calculator (http://www.movable-type.co.uk/scripts/latlong.html). If the P-value was smaller than 0.05, then the null hypothesis of no relationship between two distance matrices was rejected. Isolation-by-distance model was tested and described by plotting pairwise Fst/(1-Fsts) against geographic distances in the eastern and western group respectively (Figure 5).

Results

Parsimony resulted in a single tree (not shown here), which was mostly unresolved except for one large clade, which corresponded 12 haplotypes that were mostly associated with CA and NV localities Haplotype network showed similar relationship among haplotypes as the parsimony tree but provided additional information on the reticulation among haplotypes (Figure 1). Eight haplotypes out of 42 were common (haps 3, 6, 7, 10, 15, 17, 24, 28) and occurred in more than one population. Haplotype 6 (Figure 1) was most common and was shared by 154 individuals throughout 26 populations. Haplotypes 7 and 10 were the second and third most frequent and shared by 25 and 13 individuals, respectively. Two haplotype networks were centered around haplotypes 7 and 10, which were clustered in two distinct groups of closely located populations, respectively (Figure 1, 2). Among the 19 haplotypes that occurred in pre-outbreak populations, only five haplotypes (haps 6, 7, 10, 15, 17) were observed in outbreak populations (Figure 1, Table 2). None of the unique haplotypes (13) observed in the pre-outbreak populations occurred in the outbreak populations.

Diversity analysis indicated that pre-outbreak populations had a higher level of

genetic variation. Mean values of the number of unique haplotypes, haplotype diversity(H), nucleotide diversity(π) and the number of pairwise differences from pre-outbreak populations(1.3, .5229, .0032 and 1.2667, respectively) were all higher compared to the outbreak populations(1.2, .4002, .0018 and .7024, respectively), but the differences were not statistically significant (Table 2). The average number of unique haplotypes per population from the two time periods was similar (1.3 and 1.2). The results suggested that this genetic diversity was endemic, and there was no obvious sign of founder from pre-outbreak populations in outbreak populations, because there was no significant genetic diversity reduction in outbreak. Interestingly, the highest diversity indices all occurred in the localities in the same region (populations 3, 10 and 14), which suggested that genetic variation was associated with geography.

The table 1b and Figure 2 showed the trend of haplotype diversity. In the central range, populations (24-28) were dominated by the more common haplotypes whereas starting from population 2 towards northwest haplotype diversity increased gradually, to population 20 where unique haplotypes comprise 50% of the haplotypes. Northeastern populations (3, 10, 13, 14) also exhibited greater haplotype diversity. There was a strong linear relationship of genetic diversity with higher latitude (Figure 4) supported the above pattern of haplotypes. What's interesting was haplotype 3 only appeared in three distant populations (1, 2 and 9) that were scattered at three different states (CA, AZ and NW, respectively), but seemingly pointed three different clusters (west, center and east). It is possible that two groups of *Ips* beetles dispersed from northwest and northeast separately, and eventually jointed in the south centerline area (where populations 2,4,5 located). .

PCA and, T-test and F-test were conducted to investigate the outbreak effect and

geographic distribution on spatial genetic structure. PCA explained the sample values distribution along two perpendicular axes: PC1 always explains most of the variance, and PC2 explains more variance and PC3 explains less variance than PC1 and PC2. Sometimes, PCA can calculated more and illustrate in 3D plot (in which new axis is perpendicular to previous) but in our case, PC1 has explained 53.1% of total variance, PC2 explained additional 22.5% and PC3 explained 5.6%. So overall, 81% of total variance was explained and thus we demonstrated the results in 2D graph (Figure 3). From the graph, we can see that the separation of the sample values was not observed along PC1, which simply means that the sample groupings were not associated with the most efficient way to address the variance among the haplotypes. However, the haplotype variance was more associated with geographic regions seen from PC2. Although it's not usual to see this, it's not surprising given the large overall variance in our dataset with high number of unique haplotypes.

T-tests and F-tests were given on the means and variances of each haplotype frequencies (haps 6, 7, 10, 15 and 17) respectively (Table 3). All five tests (on each of the five haplotypes) failed to reject the null hypotheses that there were no significant differences between the means (or variances) and compared populations. In other words, the outbreak did not effect spatial genetic structure of *I. confusus*. Similar tests were conducted in the grouping based on geographic regions on the same haplotypes (west vs center, west vs east, center vs east), and haps 6, 7 and 10, which were more common showed strong association with geographic locations (Table 3).

MANOVA on PC values indicated significant difference on geography (Table 5), therefore the haplotype patterns were different between each other at least two of the geographic regions. There was no significance effect on outbreak solely or the interaction

between outbreak and geographic location (Table 5). Populations from east and west were different based on the T-tests results on haplotype frequencies (Table 3) and subsequent tests (results not shown).

The hierarchical variation analysis, we conducted both standard global AMOVA tests and locus-to-locus AMOVA (because of the missing data), two types of AMOVA tests were not statistically significantly different thus we only present the results from the standard AMOVA tests here. The results of the one group (28 populations) analysis showed 55.68% and 44.32% of genetic variation was attributed to the variance within populations and the variance among populations (Table 6a.) and the global Fst was .4432($p < .0005$). We conducted two-group AMOVA to investigate the difference between western and eastern geographic regions (Table 6b.). The results of the two-group AMOVA showed that within population contributes about half (49.86%) of the total variation, and the remaining variance is explained by among groups (19.97%) and among populations (30.18%). All three levels were highly significant ($p < .0005$). The p-value of Fct and Va (among groups) was lower than the significance level, indicating that there were significant differences between the two groups, suggesting genetic structure among the 28 populations. A cluster of populations dominated by pure haplotype (6) was observed in the central region (Figure 2, oval area), so we separated those populations from the rest of Western group and formed a three-group AMOVA (Table 6c.). All three levels were highly significant (p-value< .0005) as we expected, which confirmed that genetic structure strongly associated with geography. A fourth AMOVA analysis was also performed to test genetic differentiation between pre-outbreak and outbreak populations (Table 6d.). Genetic variation between groups was very small (little differentiation) and not significant, which suggested genetic differentitation

among populations was not associated with outbreak effect. The genetic variation among populations and "outbreak" differentiation contributed 44.32% and 45.44% of total variance, versus the variation among populations between geographic regions was smaller but significant (30.18%, P-value <.0005),

The result of the exact test of population differentiation was shown as the matrix of pairwise Fsts (Table 7). Significant genetic differentiation Fsts were common all over the place, there were 140 out of 378 (37%) pairwise comparisons in whole 28 populations found genetically significant different (P<.05). No obvious trend of Fsts was observed between pre-outbreak and outbreak populations (Table 7a), even though the average number of significantly differentiated population in each grouping was slightly different (12.20 of pre-outbreak vs. 8.89 in outbreak), suggesting that population differentiation was not related with the outbreak. Among geographic regions however, both within Western grouping (except populations 1,6) and within central grouping showed very low genetic differentiation (Table 7b). Except populations 1,2,3,6,10,13,14 and 21, which all coincidentally located at the edge of our sampling area, the rest populations were not significantly differentiated from each other (P>.05), suggesting some gene flow. Populations 13, 16, and 27 (Fst=.8881, P-value<.05), and populations 7 and 22(Fst=-.1133, P-value>.86) were the most and the least differentiated populations respectively (Table 7), suggesting the possibility of isolation by distance, because the former two were located at the two corner of the area whereas latter two were very near to each other.

The global Fst (.4432) across entire 28 populations was significant indicating that populations were highly differentiated from each other and there was restriction of gene flow among all populations, whereas global Fsts of pre-outbreak and outbreak populations

were similar (.3480 and .5172 respectively). In the AMOVA separating pre- and during outbreak populations (Table 6d), Fct (among pre-outbreak and outbreak populations) was only .0042, showing little differentiation, which agreed the similar global Fst values that outbreak has very limited impact on populations differentiation. Also, there were no obvious geographic patterns of significant differentiation.

Mantel tests were performed to test isolation-by-distance (Table 8, Figure 5). The first Mantel test was performed on all 28 populations, and there was no significant correlation between two matrices (r=.1102, P=.1316); then we performed Mantel tests on Western, Central and Eastern group corresponding to geographic regions respectively as before, and found strong indication of isolation-by-distance in Western region(r=.5188, P=.0004) and partially in Eastern region (r=.4321, P=.0434), but not in central area (r=.2131, P=.1910) (Table 8.).

Discussion

Ips confusus is not the most aggressive bark beetle and thus its appearance is related to the presence of weakened or dying host trees. Drought produces large amount of stressed trees, which would provide perfect habitats and food sources for the beetles, in other words, create the conditions for bark beetle outbreaks. Along with the increasing drought severity index from 2000, the annual area killed by bark beetle started to increase dramatically in 2001, peaked in 2003 and dropped to an endemic level by 2007 (Williams *et al.*, 2010). During that period of time, *I. confusus* outbreak occurred in the six states (Breshear *et al*., 2005; USDA-Forest Service 2004; Williams *et al*., 2010).

Our study investigated the outbreak effect on the genetic structure of this species by

building the haplotype network, comparing the genetic diversity and genetic variation distribution between pre-outbreak and outbreak populations, and rejected the hypothesis that outbreak impacted the genetic structure of *I. confusus*. The six states we sampled covered the geographic range of pinyon pines of United States (Little 1971), and on each tree we sampled an individual from separate broods (leading by one single male beetle) so not to bias the sampling of mtDNA haplotypes .

Changes in haplotype frequency are reliable indicator of gene flow that could result in alterations of spatial genetic structure. The statistical tests in our study showed no significant differences between haplotype frequency changes before and during outbreak. In addition, most haplotypes were unique and those from pre-outbreak populations were not found in greater abundance in outbreak populations: Forty-two haplotypes were found in 244 individuals; 34 haplotypes were unique which was consistent with the haplotype diversity observed among other scolytine and some insect species (Menard and Cognato, 2007, Kobayashi *et al.*, 2011). The genetic diversity of pre-outbreak and outbreak *I. confusus* populations was similar to the diversity observed with endemic and epidemic populations of mountain pine beetle (Chapuis *et al*., 2008 and 2009). Also, *I. confusus* populations were isolated by distance. The distribution of the observed genetic diversity suggested that Pleistocene geology shaped genetic structure and short-distance dispersal accounted for beetle migration.

Although, we did not observe an association between the distribution of genetic variation and outbreak status, genetic variation was associated with geography. The three genetic clusters revealed by phylogenetic tree, AMOVA analyses (Table 3) and Mantel tests (Table 8, Figure 5) are associated with the western, southwestern, and eastern range of the

beetle as observed in Cognato *et al*. (2003). A significant association between inter-population variance in haplotype frequency and geographic distance (Isolation by distance) was observed within western region (P-value=.0004) and part of eastern region (P-value=.0434). The wide distribution of common haplotypes (i.e. haplotypes 6 and 7) and the pattern of genetic variability association with geography showed in AMOVA and Mantel tests is likely due to Pleistocene geologic events observed for other North American scolytine species (Cognato *et al*. 1999, Kelly *et al.* 1999, Cognato *et al.* 2005, James *et al.*, 2011; Tsui *et al.*, 2012). It is not well understood how the Pleistocene effected the distribution of genetic variation among *I. confusus* populations. However, beetle populations likely followed the distribution of their tree hosts to lower altitudes and latitudes in the colder climate (Cognato *et al*. 2003).

Bark beetles' attack or new colonization occurs when a single male beetle successfully bores into a host tree and produces pheromone, which attracts female and other male beetles fly and join. After the mating, younger generation finishes its growth in the inner bark, then bores out of the bark and fly to next targeting host tree. Therefore habitat connectivity helps to mediate beetle colonization (Robertson *et al*., 2009), despite short-distance dispersal, which may explain the similar frequencies of common haplotypes in proximal populations (populations 6, 7, 21, 22, Figure 2). However landscape features (e.g., mountains and treeless areas) likely impact beetle migration (Aukema *et al*., 2008; Chen and Walton 2011). For example, the Shoshone Mountains lying between populations 19 and 20 are a possible barrier to beetle movement evidenced by the different haplotypic composition in these proximal populations. Long-distance dispersal (< 50 km), as observed with other bark beetles (Chen and Walton 2011; Lowe 2009) is possible but it is likely

uncommon given that distribution of haplotypes is better explained by the influence of Pleistocene geography. Landscape features throughout time and Western North America (de la Giroday *et al*., 2011) likely influenced the dispersal of *I. confusus* by curbing long-distance dispersal, and shaped the current population genetic structure.

There is much evidence for the effect of climate change on insect populations (e.g. Carroll et al. 2004; Robinet and Roques 2010). Increased favorable environmental conditions (e.g. increase of stressed trees in our case promote an increase in the population size, which increases the likelihood of an outbreak. Our study, as others suggest that drought promotes multiple independent outbreaks among some herbivorous insects (e.g., Ronnas *et al*. 2011). However, intrinsic factors, such as physiology and behavior, mostly influence the dispersal ability of the insects and hence short distance dispersal mediates gene flow. Extrinsic stochastic factors, such as wind and humans, may become more important to long distance dispersal once outbreak populations grow to a critical size and number (e.g., Safranyik *et al*. 2010, de la Giroday *et al*. 2011). As a consequence, long-term drought or global warming will likely promote increased gene flow among *I. confusus.*

APPENDICES

Appendix A (Tables)

Table 1a. Sampling information.

Population ID, Location (county names), Latitude, Longitude, elevation, host tree, the number of individuals(N_I) collected in each population, the number of haplotypes (N_H) in each population and the number of polymorphic sites (N_P) (the number of loci that has more than one allele per locus) were showed in this table. White Pine= WP, Little Antelope summit= l.a. sum.

Pop ID	Location	Latitude	Longitude	Elevation (ft)	Host tree	N_I	N_H	N_P
Pre-outbreak populations (from 2001)								
1	San Bernardino, CA	34°18'N	116°49'W		*P. monophylla*	9	6	5
2	Greenlee, AZ	33°10'N	109°23'W		*P. edulis*	8	4	7
3	Dolores, CO	37°45'N	108°00'W		*P. edulis*	7	4	6
4	Greenlee, AZ	33°38'N	109°20'W	>3030	*P. pungens*	9	2	1
5	Gila, AZ	33°36'N	110°15'W		*P. edulis*	10	2	1
6	Mono, CA	38°05'N	119°10'W		*P. monophylla*	10	3	2
7	Inyo, CA	37°15'N	118°10'W		*P. monophylla*	9	3	3
8	Otero, NM	32°53'N	105°30'W		*P. edulis*	6	1	0
9	Sandoval, NM	36°01'N	106°57'W		*P. edulis*	10	3	5
10	Montezuma, CO	37°28'N	108°29'W		*P. edulis*	8	6	5
Outbreak populations (from 2003)								
11	Huerfano, CO	37°30'N	104°42'W	1976	*P. edulis*	10	3	3
12	Fermont, CO	38°22'N	105°41'W	1948	*P. edulis*	10	3	2
13	Rio Blanco, CO	39°41'N	108°48'W	2122	*P. edulis*	10	5	4
14	Duchesne, UT	40°08'N	110°29'W	1953	*P. edulis*	9	5	10
15	Tooele,UT	40°00'N	112°17'W	1740	*P. monophylla*	7	3	2
16	WP, nr Baker, NV	39°01'N	114°12'W	1989	*P. monophylla*	9	1	0
17	WP, nr Ely, NV	39°03'N	114°37'W	2206	*P. monophylla*	10	2	1
18	WP, nr l.a. sum.	39°24'N	115°28'W	2279	*P. monophylla*	10	3	2
19	Lander, NV	39°27'N	116°45'W	1980	*P. monophylla*	10	2	1
20	Churchill, NV	39°15'N	117°48'W	1828	*P. monophylla*	10	6	4
21	Douglas, NV	38°48'N	119°44'W	1648	*P. monophylla*	8	4	3
22	Esmeralda, NV	37°25'N	117°38'W	2030	*P. monophylla*	6	3	2
23	Clark, NV	36°16'N	115°32'W	1788	*P. monophylla*	9	3	2
24	Washington, UT	37°26'N	113°30'W	2055	*P. monophylla*	8	1	0
25	Iron, UT	37°40'N	113°00'W	1885	*P. edulis*	7	1	0
26	Washington, UT	37°17'N	113°06'W	1500	*P. monophylla*	8	2	2
27	Coconino, AZ	36°51'N	112°16'W	1879	*P. edulis*	9	1	0
28	Coconino, AZ	35°24'N	111°35'W	2106	*P. edulis*	8	3	6

Table 1b. Sampling information.

Population ID, haplotypes and the haplotype number (in the bracket) in each populations were listed in the table.

Pop ID	Haplotype (#)
1	Hap 1 (1); hap 2 (3); hap 3 (1); hap 4 (1); hap 6 (2); hap 7 (1)
2	Hap 3 (3); hap 6 (3); hap 9 (1); hap10 (1)
3	Hap 3 (1); hap 5(1); hap 6 (3); hap 10(2)
4	Hap 6 (8); hap 8 (1)
5	Hap 6 (9); hap 11(1)
6	Hap 7 (8); hap 6 (1); hap 12 (1)
7	Hap 6 (5); hap 7 (3); hap 13 (1)
8	Hap 6 (6)
9	Hap 6 (8); hap 10 (1); hap 15 (1)
10	Hap 10 (1); hap 14 (2); hap 16 (2); hap 17 (1); hap 18 (1); hap 19 (1)
11	Hap 6 (8); hap 17 (1); hap 20 (1)
12	Hap 6 (8); hap 21 (1); hap 22 (1)
13	Hap 10 (6); hap 23 (1); hap 24 (1); hap 25 (1); hap 39 (1)
14	Hap 6 (2); hap 10 (2); hap 24 (1); hap 26 (3); hap 27 (1)
15	Hap 6 (5); hap 28 (1); hap 29 (1)
16	Hap 6 (9)
17	Hap 6 (9); hap 28(1)
18	Hap 6 (8); hap 7 (1); hap 30 (1)
19	Hap 6 (6); hap 7 (4)
20	Hap 6 (5); hap 7 (1); hap 31 (1); hap 32 (1); hap 33 (1); hap 34 (1)
21	Hap 6 (1); hap 7 (5); hap 35 (1); hap 36 (1)
22	Hap 6 (3); hap 7 (2); hap 37 (1)
23	Hap 6 (7); hap 15 (1); hap 38 (1)
24	Hap 6 (8)
25	Hap 6 (8)
26	Hap 6 (7); hap 40 (1)
27	Hap 6 (9)
28	Hap 6 (6); hap 41 (1); hap 42 (1)

Table 2. Molecular diversity information.

The number of haplotypes, the number of unique haplotypes, and a list of the unique haplotypes in each populations were summarized in this table, and diversity indices including gene diversity(H), nucleotide diversity(π) and the number of pairwise differences with means and 95% CI were listed in the table as well. μ represented mean value.

Pop ID	# of haps	# of unique haps	Unique haps	Haplotype diversity (H)		Nucleotide diversity(π)		# of pairwise differences	
				Mean	95% CI	mean	95% CI	mean	95% CI
Before-outbreak populations (from 2001)									
1	6	3	h1, 2, 4	0.8889	[0.7979, 0.9799]	0.0050	[0.0015, 0.0086]	2.0118	[0.7652, 3.2584]
2	4	1	h9	0.7857	[0.6730, 0.8984]	0.0059	[0.0019, 0.0010]	2.3464	[0.9189, 3.7739]
3	4	1	h5	0.8095	[0.6797, 0.9393]	0.0082	[0.0027, 0.0137]	3.2812	[1.3648, 5.1977]
4	2	1	h8	0.2222	[0.0560, 0.3884]	0.0006	[-0.0003, 0.014]	0.224	[-0.0654, 0.5126]
5	2	1	h11	0.2000	[0.0459, 0.3541]	0.0005	[-0.0003, 0.0013]	0.2011	[-0.0689, 0.4711]
6	3	1	h12	0.3778	[0.1965, 0.5591]	0.0010	[0.0000, 0.0022]	0.4023	[-0.0020, 0.8066]
7	3	1	h13	0.6389	[0.5131, 0.7647]	0.0024	[0.0004, 0.0044]	0.9548	[0.2390, 1.6706]
8	1	0	N	0.0000	0	0	0	0	0
9	3	0	N	0.3778	[0.1965, 0.5591]	0.0025	[0.0004, 0.0047]	1.0133	[0.2739, 1.7528]
10	6	4	h14, 16, 18, 19	0.9286	[0.8442, 1.0130]	0.0060	[0.0018, 0.0103]	2.2329	[0.8614, 3.6044]
μ	3.4	1.3		0.5229	[0, 1.0130]	0.0032	[0, 0.0173]	1.2667	[-0.0654, 5.1977]
During-outbreak populations (from 2003)									
11	3	1	h20	0.3778	[0.1965, 0.5591]	0.0015	[3.5E-05, 0.0030]	0.6036	[0.0823, 1.1249]
12	3	2	h21, 22	0.3778	[0.1965, 0.5591]	0.0005	[-0.0003, 0.0013]	0.2012	[-0.0688, 0.4712]
13	5	3	h23,25, 39	0.6667	[0.5034, 0.8300]	0.0024	[0.0004, 0.0044]	0.9632	[0.2496, 1.6767]
14	5	2	h26, 27	0.8611	[0.7739, 0.9483]	0.0106	[0.0040, 0.0172]	4.2437	[1.9220, 6.5654]
15	3	1	h29	0.5238	[0.3152, 0.7324]	0.0015	[-6.2E-05, 0.0030]	0.5752	[0.0522, 1.0982]
16	1	0	N	0.0000	0	0.0000	0	0	0
17	2	0	N	0.2000	[0.0459, 0.3541]	0.0005	[-0.0003, 0.0013]	0.2011	[-0.0689, 0.4711]
18	3	1	h30	0.3778	[0.1965, 0.5591]	0.0010	[-0.0004, 0.0022]	0.4027	[-0.0019, 0.8072]
19	2	0	N	0.5333	[0.4386, 0.6280]	0.0013	[-2.5E-05, 0.0027]	0.5365	[0.0531, 1.0199]

Table 2 (cont'd)

20	6	4	h31, 32, 33, 34	0.7778	[0.6404, 0.9152]	0.0029	[0.0005, 0.0053]	1.0764	[0.3046, 1.8482]
21	4	2	h35, 36	0.6429	[0.4588, 0.8270]	0.0019	[0.0001, 0.0036]	0.7541	[0.1381, 1.3701]
22	3	1	h37	0.7333	[0.5781, 0.8885]	0.0023	[0.0002, 0.0044]	0.8737	[0.1690, 1.5784]
23	3	1	h38	0.4167	[0.2260, 0.6074]	0.0012	[-0.0001, 0.0025]	0.4463	[0.0117, 0.8809]
24	1	0	N	0.0000	0	0.0000	0	0	0
25	1	0	N	0.0000	0	0.0000	0	0	0
26	2	1	h40	0.2500	[0.0698, 0.4302]	0.0013	[-8.9E-05, 0.0026]	0.5055	[0.0306, 0.9805]
27	1	0	N	0.0000	0	0.0000	0	0	0
28	3	2	h41, 42	0.4643	[0.2643, 0.6643]	0.0032	[0.0006, 0.0057]	1.2598	[0.3766, 2.1431]
μ	2.83	1.2		0.4002	[0, 0.9483]	0.0018	[-6.2E-05, 0.0172]	0.7024	[-0.0689, 2.1431]

Table 3. Haplotype frequency.

The haplotype frequencies of five haplotypes that occurred in pre- and during- outbreak.

Pop ID	h6	h7	h10	h15	h17
1	0.2222	0.1111	0	0	0
2	0.3750	0	0.1250	0	0
3	0.4286	0	0.2857	0	0
4	0.8889	0	0	0	0
5	0.9000	0	0	0	0
6	0.1000	0.8000	0	0	0
7	0.5556	0.3333	0	0	0
8	1.0000	0	0	0	0
9	0.8000	0	0.1000	0.1000	0
10	0	0	0.1250	0	0.1250
11	0.8000	0	0	0	0.1000
12	0.8000	0	0	0	0
13	0	0	0.6000	0	0
14	0.2222	0	0.2222	0	0
15	0.7143	0	0	0	0
16	1	0	0	0	0
17	0.9000	0	0	0	0
18	0.8000	0.1000	0	0	0
19	0.6000	0.4000	0	0	0
20	0.5000	0.1000	0	0	0
21	0.1250	0.6250	0	0	0
22	0.5000	0.3333	0	0	0
23	0.7778	0	0	0.1111	0
24	1	0	0	0	0
25	1	0	0	0	0
26	0.8750	0	0	0	0
27	1	0	0	0	0
28	0.7500	0	0	0	0

Pop ID		Hap 6	Hap 7	Hap 10	Hap 15	Hap 17
Pre outbreak	Average	0.5270	0.1244	0.0636	0.0100	0.0125
	Variance	0.1283	0.0676	0.0092	0.0010	0.0016
During outbreak	Average	0.6869	0.0866	0.0457	0.0062	0.0056
	Variance	0.0944	0.0322	0.0219	0.0007	0.0006
West	Average	0.5303	0.2803	0	0	0
	Variance	0.0980	0.0733	0	0	0
Center	Average	.9004	0	0	.0159	0
	Variance	.0113	0	0	.0018	0
East	Average	.5481	0	.1325	.0091	.0205

Table 4. PC scores derived from haplotype frequencies of each populations.

Population ID	Outbreak	Geo region	PC1	PC2	PC3
1	Pre	West	0.1096	0.0111	0.0718
2	Pre	East	0.0574	0.0714	0.0059
3	Pre	East	0.0514	0.0859	-0.0395
4	Pre	East	-0.083	0.0033	0.0013
5	Pre	Center	-0.0851	0.0028	0.0008
6	Pre	West	0.2187	-0.1591	-0.0131
7	Pre	West	0.0381	-0.0822	-0.0118
8	Pre	East	-0.1045	-0.0018	-0.0035
9	Pre	East	-0.0529	0.0297	-0.0044
10	Pre	East	0.185	0.1031	0.1203
11	During	East	-0.0633	0.0093	0.0144
12	During	East	-0.0645	0.0077	0.0058
13	During	East	0.2087	0.1609	-0.0844
14	During	East	0.1106	0.0973	-0.0443
15	During	East	-0.0468	0.0124	0.0118
16	During	West	-0.1045	-0.0018	-0.0035
17	During	West	-0.0855	0.003	0.0017
18	During	West	-0.0489	-0.0264	-0.0044
19	During	West	0.0349	-0.1007	-0.0182
20	During	West	0.0216	-0.0106	0.0137
21	During	West	0.1958	-0.1276	-0.0048
22	During	West	0.0349	-0.1007	-0.0182
23	During	Center	-0.0802	0.004	0.002
24	During	Center	-0.1045	-0.0018	-0.0035
25	During	Center	-0.1045	-0.0018	-0.0035
26	During	Center	-0.0802	0.004	0.002
27	During	Center	-0.1045	-0.0018	-0.0035
28	During	Center	-0.0539	0.0104	0.009

Table 5. Multivariate analysis of variance (MANOVA)

Model	Wilk's Lambda		Pillai's Trace		Hotelling-Lawley		Roy's Max Root	
	Value	Prob>F	Value	Prob>F	Value	Prob>F	Value	Prob>F
Outbreak	F test: Prob>F=.5871							
Geo region	.3602	.0016*	.7322	.0028*	1.5194	.0010*	1.3259	.0004*
Outbreak* geo region	.8602	.7885	.1428	.7761	.1591	.8014	.1333	.4419

Table 6a. Analysis of molecular variance (AMOVA) in one group.

Variance among populations (Va) and variance within populations (Vb), percentages of each variation (%) and associated F-statistics, with significance level (P-value).

Source of variance	Variance components	Percentage of variation	Fst	P-value
Among populations	.3641 Va	44.32	.4432	<.0005*
Within populations	.4574 Vb	55.68		
total	.8218			

Table 6b. Analysis of molecular Variance (AMOVA) in two groups.

Variance among groups (Va), variance among populations within groups (Vb) and variance within populations (Vc), and corresponding fixation indices Fct, Fst, Fsc, respectively, and the associated F-statistics with p-values.

Source of variance	Variance components	Percentage of variation	Fixation indices	P-value
Among groups	.1832 Va	19.97	.2000 Fct	<.0005*
Within groups among populations	.2768 Vb	30.18	.3771 Fsc	<.0005*
Within populations	.4574 Vc	49.86	.5014 Fst	<.0005*

Table 6c. Analysis of molecular variance (AMOVA) in three groups.

Variance among groups (Va), variance among populations within groups (Vb) and variance within populations (Vc), and corresponding fixation indices Fct, Fst, Fsc, respectively, and the associated F-statistics with p-values.

Source of variance	Variance components	Percentage of variation	Fixation indices	P-value
Among groups	.1156 Va	20.64	.2064 Fct	<.0005*
Within groups among populations	.0750 Vb	13.37	.1685 Fsc	<.0005*
Within populations	.3696 Vc	65.99	.3401 Fst	<.0005*

Table 6d. Analysis of molecular variance (AMOVA) in outbreak effect.
Variance among groups (Va), variance among populations within groups (Vb) and variance within populations (Vc), and corresponding fixation indices Fct, Fst, Fsc, respectively, and the associated F-statistics with p-values.

Source of variance	Variance components	Percentage of variation	Fixation indices	P-value
Among groups	0 Va	0	0 Fct	.6110
Within groups among populations	.3702 Vb	45.44	.4473 Fsc	<.0005*
Within populations	.4574 Vc	56.14	.4386 Fst	<.0005*

Table 7a. Pairwise Fsts among populations on outbreak effect.

* p-value<.05.

Pop ID	Pre-outbreak							
	1	2	3	4	5	6	7	8
1	0							
2	0.1323	0						
3	0.2907	0.0118	0					
4	0.4643*	0.0899*	0.2696*	0				
5	0.5172*	0.1769*	0.3422*	0.0006	0			
6	0.4440*	0.4560*	0.4994*	0.7164*	0.7273*	0		
7	0.3589	0.1500*	0.2647	0.125	0.1375	0.3039*	0	
8	0.4607	0.1068	0.2613	-0.0511	-0.0588	0.7534*	0.0866	0
9	0.4317*	0.0771	0.1652	-0.0441	0	0.5333*	0.0588	-0.0588
10	0.5111*	0.2832*	0.0322	0.4768*	0.5111*	0.6026*	0.4217*	0.4434*
11	0.4707*	0.1215*	0.2513	-0.0053	0	0.6154*	0.0765	-0.0588
12	0.4929*	0.1620*	0.3227	-0.0033	0	0.6667*	0.1165	-0.0588
13	0.7803*	0.6770*	0.4126	0.8615*	0.8744*	0.8768*	0.8034*	0.8665*
14	0.4903	0.2926	0.0696	0.4463*	0.4842*	0.5485*	0.4153*	0.4114*
15	0.3924	0.0563	0.235	0.0157	0.0238	0.6318*	0.0894	-0.0244
16	0.5263*	0.1761*	0.3421*	0	-0.0112	0.7902*	0.15	0
17	0.5172*	0.1769*	0.3422*	0.0006	0	0.7273*	0.1375	-0.0588
18	0.4668*	0.1620*	0.3227	-0.0033	0	0.6078*	0.0294	-0.0588
19	0.3979*	0.2201*	0.3479*	0.2537	0.2667	0.3137	-0.0701	0.25
20	0.2705	0.0774	0.2366	0.0736	0.0972	0.2912*	-0.0383	0.04
21	0.3831	0.3857*	0.4315*	0.6167*	0.6322*	-0.0382	0.2255	0.6263
22	0.2606	0.0563	0.1988	0.1254	0.1712	0.3333	-0.1133	0.1333
23	0.4762*	0.1396	0.2888	-0.0385	0.0045	0.6553*	0.1071	-0.0511
24	0.5069*	0.1558*	0.3183*	-0.0141	-0.0242	0.7793*	0.1316	0
25	0.4852*	0.1331	0.2917	-0.0307	-0.0396	0.7672*	0.1108	0
26	0.4583*	0.1319	0.2821	0.0058	0.0119	0.6436*	0.098	-0.0403
27	0.5263*	0.1761*	0.3421*	0	-0.0112	0.7902*	0.15	0
28	0.4012	0.093	0.2107	0.0107	0.0207	0.5125*	0.073	-0.0403

Table 7a (cont'd)

Pop ID	Pre-outbreak		During outbreak					
	9	10	11	12	13	14	15	16
1								
2								
3								
4								
5								
6								
7								
8								
9	0							
10	0.3371*	0						
11	-0.0526	0.4166*	0					
12	0	0.4903*	0	0				
13	0.7672*	0.3397*	0.8232*	0.8558*	0			
14	0.3602*	0.0696	0.4231*	0.4716*	0.1723	0		
15	-0.012	0.4310*	-0.0011	-0.0555	0.8335*	0.4141	0	
16	-0.0112	0.5145*	-0.0112	-0.0112	0.8881*	0.4792*	0.0382	0
17	0	0.5111*	0	0	0.8744*	0.4842*	-0.0611	-0.0112
18	0	0.4858*	0	0	0.8558*	0.4701*	0.0083	-0.0112
19	0.1482	0.5003*	0.1905	0.2222	0.8481*	0.4747*	0.1966	0.3156
20	0.057	0.4072*	0.0778	0.0864	0.7843*	0.4219	0.0601	0.0966
21	0.4560*	0.5401*	0.5305*	0.5771*	0.8464*	0.4973*	0.5285*	0.6834*
22	0.0454	0.4027*	0.0988	0.1282	0.8092*	0.3747	0.0908	0.2174*
23	-0.0521	0.4565*	-0.0017	0.0006	0.8453*	0.4442*	0.0043	0
24	-0.0242	0.4935*	-0.0242	-0.0242	0.8818*	0.4591*	0.0204	0
25	-0.0396	0.4700*	-0.0396	-0.0396	0.8747*	0.4367*	0	0
26	-0.0086	0.4466*	-0.0024	0.003	0.8415*	0.4349*	0.0013	0.0156
27	-0.0112	0.5145*	-0.0112	-0.0112	0.8881*	0.4792*	0.0382	0
28	-0.0199	0.3543*	0.0097	0.0144	0.7769*	0.3812*	-0.007	0.0156

Table 7a (cont'd)

Pop ID	During outbreak							
	17	18	19	20	21	22	23	24
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17	0							
18	0	0						
19	0.2667	0.1026	0					
20	0.0972	0.0212	-0.0336	0				
21	0.6322*	0.5117*	0.2287	0.2237	0			
22	0.1712	0.0286	-0.0916	-0.0837	0.2122	0		
23	0.0045	-0.0265	0.2127	0.0553	0.5571*	0.1156	0	
24	-0.0242	-0.0242	0.2962	0.0805	0.6667*	0.1928	-0.0141	0
25	-0.0396	-0.0396	0.2746	0.062	0.6478*	0.1651	-0.0307	0
26	0.0119	0.003	0.2039	0.0446	0.5455*	0.103	0.0009	0
27	-0.0112	-0.0112	0.3156	0.0966	0.6834*	0.2174*	0	0
28	0.0207	0.0144	0.1443	0.0588	0.4286*	0.0505	-0.0643	0

Table 7a (cont'd)

Pop ID	During outbreak							
	17	18	19	20	21	22	23	24
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17	0							
18	0	0						
19	0.2667	0.1026	0					
20	0.0972	0.0212	-0.0336	0				
21	0.6322*	0.5117*	0.2287	0.2237	0			
22	0.1712	0.0286	-0.0916	-0.0837	0.2122	0		
23	0.0045	-0.0265	0.2127	0.0553	0.5571*	0.1156	0	
24	-0.0242	-0.0242	0.2962	0.0805	0.6667*	0.1928	-0.0141	0
25	-0.0396	-0.0396	0.2746	0.062	0.6478*	0.1651	-0.0307	0
26	0.0119	0.003	0.2039	0.0446	0.5455*	0.103	0.0009	0
27	-0.0112	-0.0112	0.3156	0.0966	0.6834*	0.2174*	0	0
28	0.0207	0.0144	0.1443	0.0588	0.4286*	0.0505	-0.0643	0

Table 7a (cont'd)

Pop ID	During outbreak			
	25	26	27	28
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25	0			
26	-0.0182	0		
27	0	0.0156	0	
28	-0.0182	0	0.0156	0

Table 7b. pairwise Fsts among populations on geography.

* p-value<.05

Pop ID	Western Region							
	1	6	7	16	17	18	19	20
1	0							
2	0.4440*	0						
3	0.3589	0.3039*	0					
4	0.5263*	0.7902*	0.15	0				
5	0.5172*	0.7273*	0.1375	-0.0112	0			
6	0.4668*	0.6078*	0.0294	-0.0112	0	0		
7	0.3979*	0.3137	-0.0701	0.3156	0.2667	0.1026	0	
8	0.2705	0.2912*	-0.0383	0.0966	0.0972	0.0212	-0.0336	0
9	0.3831	-0.0382	0.2255	0.6834*	0.6322*	0.5117*	0.2287	0.2237
10	0.2606	0.3333	-0.1133	0.2174*	0.1712	0.0286	-0.0916	-0.0837
11	0.5172*	0.7273*	0.1375	-0.0112	0	0	0.2667	0.0972
12	0.4762*	0.6553*	0.1071	0	0.0045	-0.0265	0.2127	0.0553
13	0.5069*	0.7793*	0.1316	0	-0.0242	-0.0242	0.2962	0.0805
14	0.4852*	0.7672*	0.1108	0	-0.0396	-0.0396	0.2746	0.062
15	0.4583*	0.6436*	0.098	0.0156	0.0119	0.003	0.2039	0.0446
16	0.5263*	0.7902*	0.15	0	-0.0112	-0.0112	0.3156	0.0966
17	0.4012	0.5125*	0.073	0.0156	0.0207	0.0144	0.1443	0.0588
18	0.1323	0.4560*	0.1500*	0.1761*	0.1769*	0.1620*	0.2201*	0.0774
19	0.2907	0.4994*	0.2647	0.3421*	0.3422*	0.3227	0.3479*	0.2366
20	0.4643*	0.7164*	0.125	0	0.0006	-0.0033	0.2537	0.0736
21	0.4607	0.7534*	0.0866	0	-0.0588	-0.0588	0.25	0.04
22	0.4317*	0.5333*	0.0588	-0.0112	0	0	0.1482	0.057
23	0.5111*	0.6026*	0.4217*	0.5145*	0.5111*	0.4858*	0.5003*	0.4072*
24	0.4707*	0.6154*	0.0765	-0.0112	0	0	0.1905	0.0778
25	0.4929*	0.6667*	0.1165	-0.0112	0	0	0.2222	0.0864
26	0.7803*	0.8768*	0.8034*	0.8881*	0.8744*	0.8558*	0.8481*	0.7843*
27	0.4903	0.5485*	0.4153*	0.4792*	0.4842*	0.4701*	0.4747*	0.4219
28	0.3924	0.6318*	0.0894	0.0382	-0.0611	0.0083	0.1966	0.0601

(Table 7b cont'd)

Pop ID	Western Region		Central Region					
	21	22	5	23	24	25	26	27
1								
2								
3								
4								
5								
6								
7								
8								
9	0							
10	0.2122	0						
11	0.6322*	0.1712	0					
12	0.5571*	0.1156	0.0045	0				
13	0.6667*	0.1928	-0.0242	-0.0141	0			
14	0.6478*	0.1651	-0.0396	-0.0307	0	0		
15	0.5455*	0.103	0.0119	0.0009	0	-0.0182	0	
16	0.6834*	0.2174*	-0.0112	0	0	0	0.0156	0
17	0.4286*	0.0505	0.0207	-0.0643	0	-0.0182	0	0.0156
18	0.3857*	0.0563	0.1769*	0.1396	0.1558*	0.1331	0.1319	0.1761*
19	0.4315*	0.1988	0.3422*	0.2888	0.3183*	0.2917	0.2821	0.3421*
20	0.6167*	0.1254	0.0006	-0.0385	-0.0141	-0.0307	0.0058	0
21	0.6263	0.1333	-0.0588	-0.0511	0	0	-0.0403	0
22	0.4560*	0.0454	0	-0.0521	-0.0242	-0.0396	-0.0086	-0.0112
23	0.5401*	0.4027*	0.5111*	0.4565*	0.4935*	0.4700*	0.4466*	0.5145*
24	0.5305*	0.0988	0	-0.0017	-0.0242	-0.0396	-0.0024	-0.0112
25	0.5771*	0.1282	0	0.0006	-0.0242	-0.0396	0.003	-0.0112
26	0.8464*	0.8092*	0.8744*	0.8453*	0.8818*	0.8747*	0.8415*	0.8881*
27	0.4973*	0.3747	0.4842*	0.4442*	0.4591*	0.4367*	0.4349*	0.4792*
28	0.5285*	0.0908	0.0238	0.0043	0.0204	0	0.0013	0.0382

Table 7b (cont'd)

Pop ID	Central	Eastern Region						
	28	2	3	4	8	9	10	11
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17	0							
18	0.093	0						
19	0.2107	0.0118	0					
20	0.0107	0.0899*	0.2696*	0				
21	-0.0403	0.1068	0.2613	-0.0511	0			
22	-0.0199	0.0771	0.1652	-0.0441	-0.0588	0		
23	0.3543*	0.2832*	0.0322	0.4768*	0.4434*	0.3371*	0	
24	0.0097	0.1215*	0.2513	-0.0053	-0.0588	-0.0526	0.4166*	0
25	0.0144	0.1620*	0.3227	-0.0033	-0.0588	0	0.4903*	0
26	0.7769*	0.6770*	0.4126	0.8615*	0.8665*	0.7672*	0.3397*	0.8232*
27	0.3812*	0.2926	0.0696	0.4463*	0.4114*	0.3602*	0.0696	0.4231*
28	-0.007	0.0563	0.235	0.0157	-0.0244	-0.012	0.4310*	-0.0011

Table 7b (cont'd)

Pop ID	Eastern Region			
	12	13	14	15
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25	0			
26	0.8558*	0		
27	0.4716*	0.1723	0	
28	-0.0555	0.8335*	0.4141	0

Table 8. Three Mantel tests.

Group A(all 28 populations), W(Western populations) and E(Eastern populations) with correlation coefficient(r) and p-values. Permutation=5000, * indicated the significance level (P< .05) between two matrices. E' is without pop 11, 12 which were on the edge of sampling area.

Group ID	populations	Correlation coefficient(r)	p-value
A	All 28	.1102	.1316
W	1,6,7,16-22	.5188	.0004*
C	5, 15, 23-28	.2131	.1910
E	2-4, 9-14	.1785	.1982
E'	2-4, 9-10,13-14	.4321	.0434*

Appendix B (Figures)

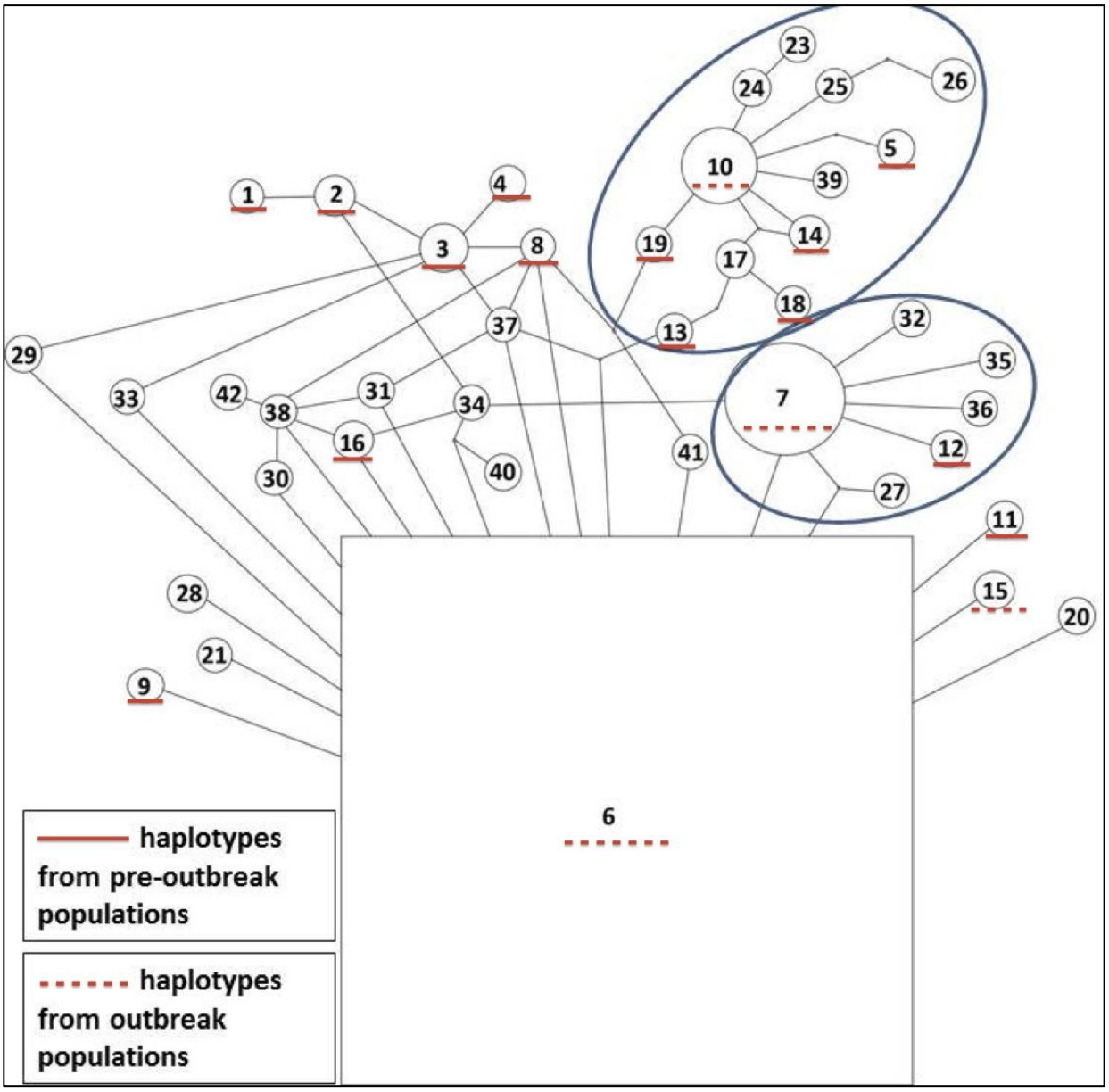

Figure 1 Haplotype network.

Each circle represents a haplotype. The square in the center represented the most common haplotype. The sizes of circles indicated the frequencies of each haplotype, the larger the more frequent. The line connecting circles and squares represent mutational steps, which is nucleotide substitutions. The nodes represented hypothetical unsampled haplotypes, either because they extinct or not sampled. The red solid bars under some sequences indicated the sequences from pre-outbreak populations (1-10) and red dash bars indicated the sequences from pre-outbreak populations and still maintained in outbreak populations. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis.

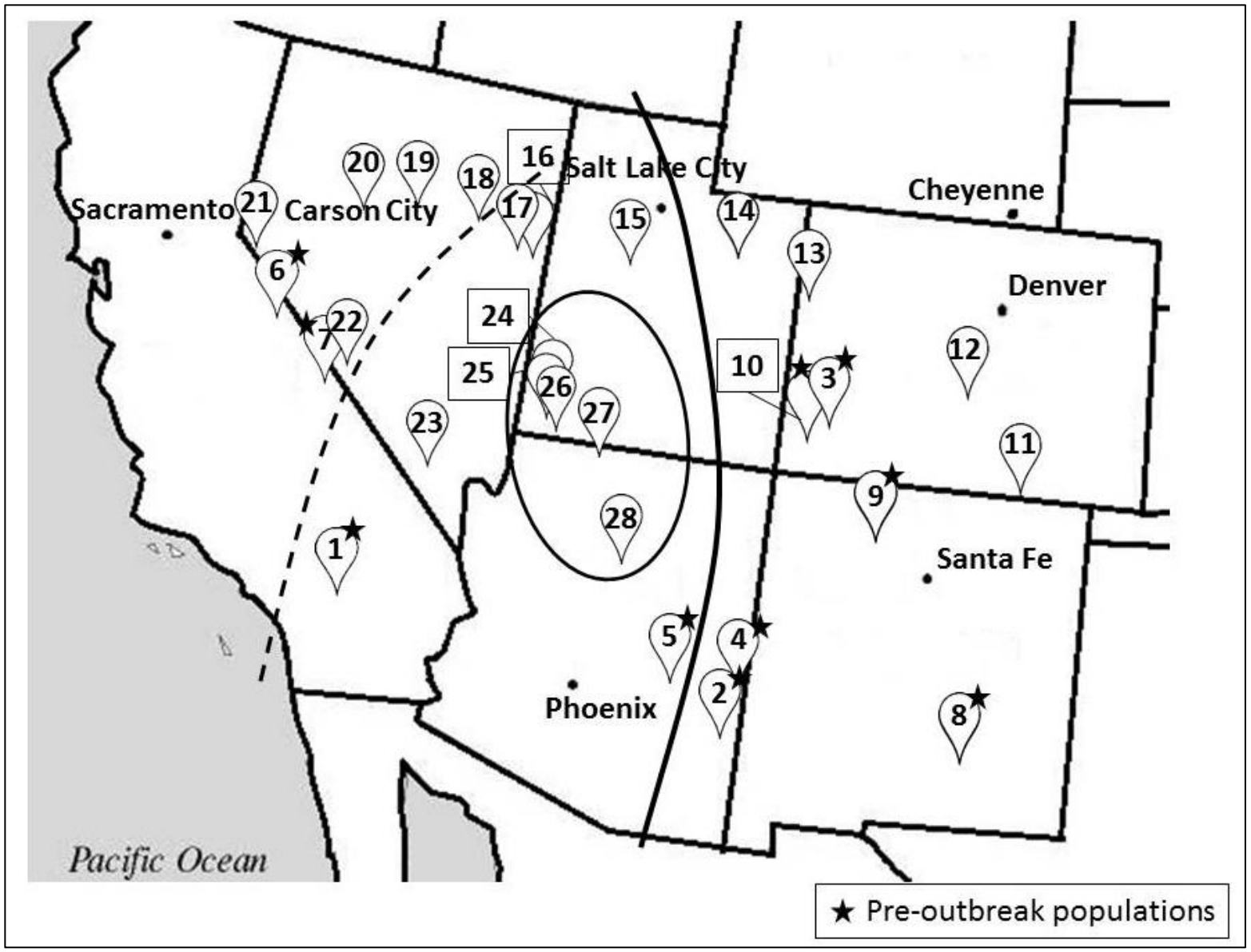

Figure 2. The map of sampling localities

Figure 2. (cont'd). Each teardrop icon located the 28 populations we sampled on the map of western and southwestern US. Population IDs were marked on top of each icons. Stars on tear-drop icon indicated pre-outbreak populations from 2001. The solid and dashed curve lines and the oval in the center divided the sampling area into three geographic regions.

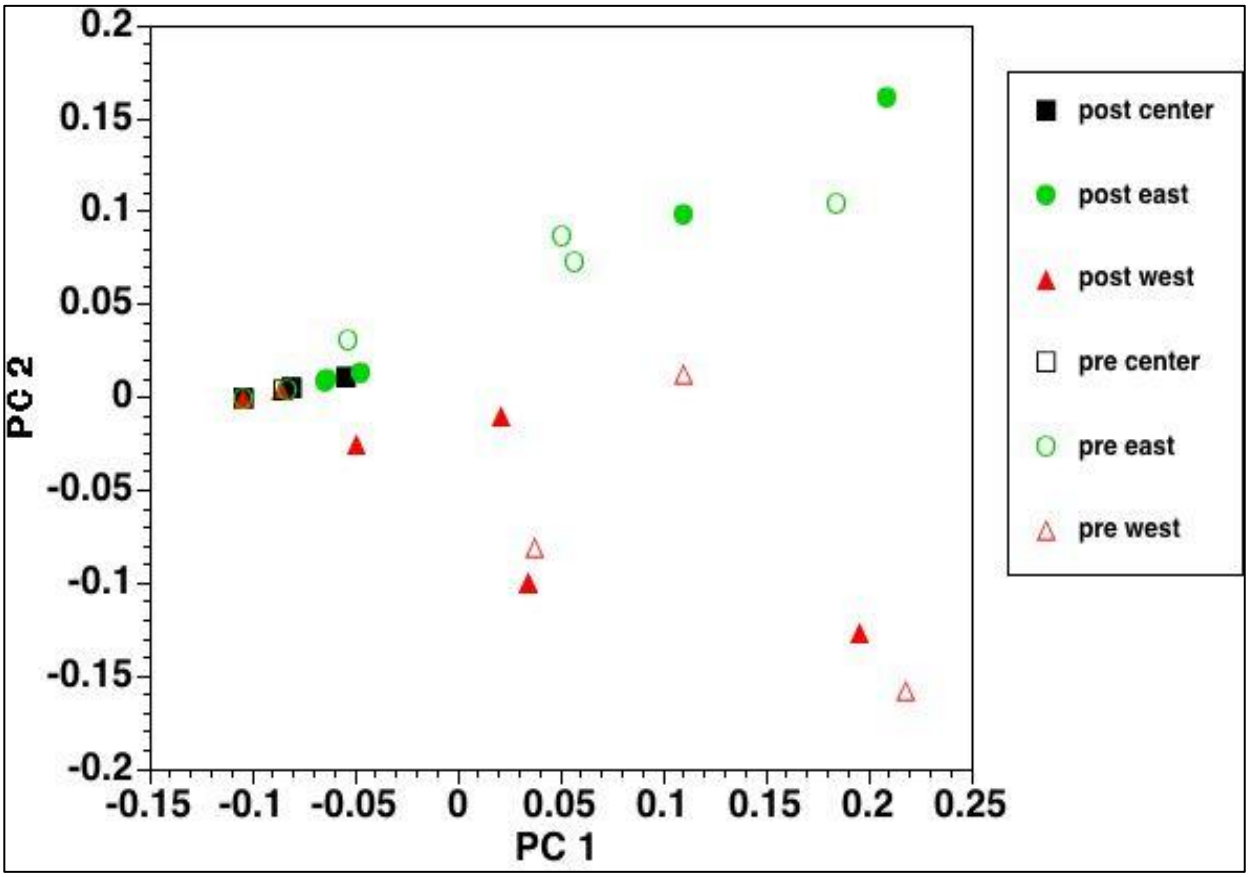

Figure 3 Principal component analysis.

Black, green and red represent three different geographic regions respectively, and solid shapes and hollow shapes differentiate during and pre-outbreak populations respectively.

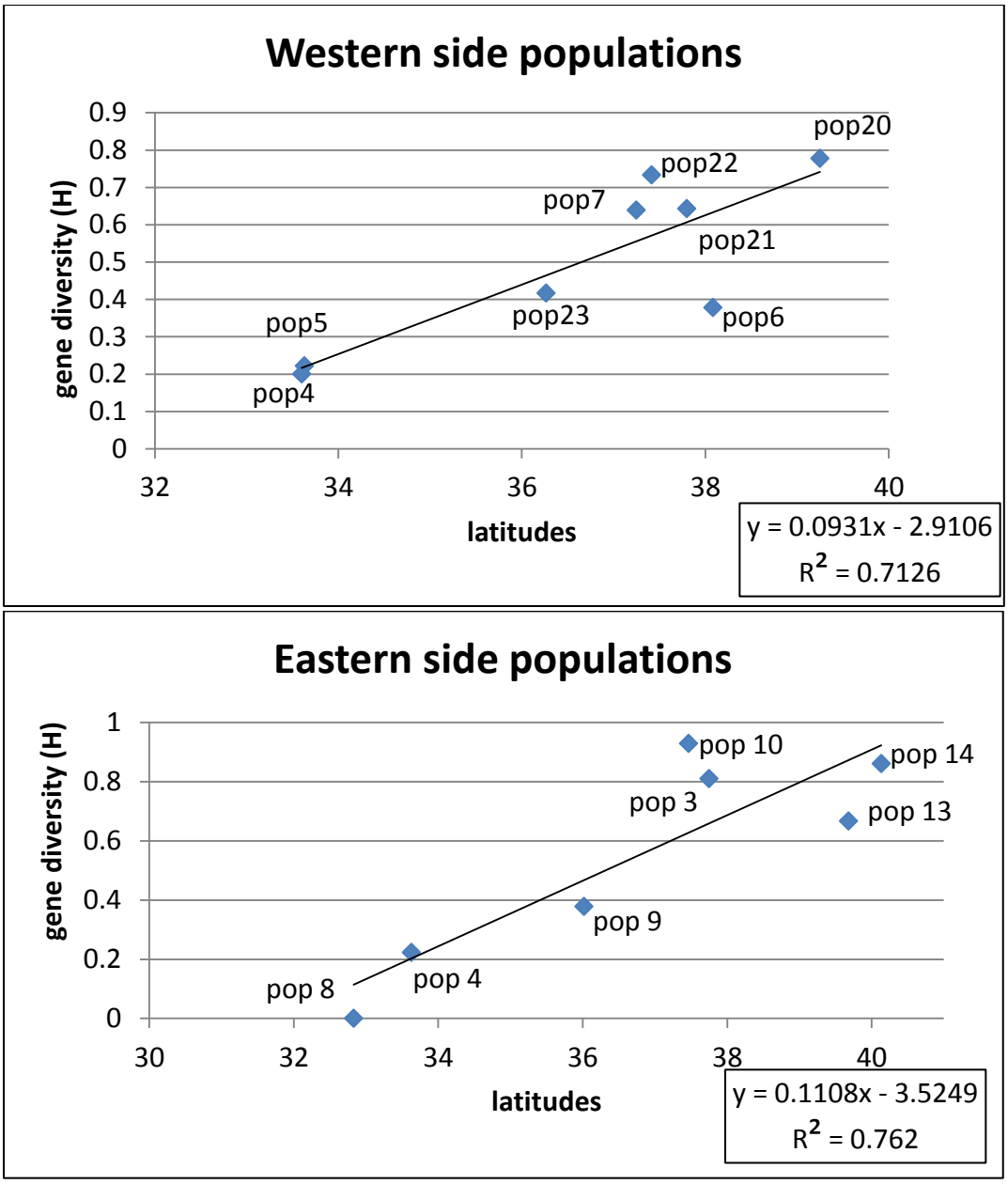

Figure 4 Regression plot of haplotype diversity (H) against latitudes.
Western group included populations 4-7 and 20-23 and the Eastern group included populations 3,4,8,9,10,13,14.

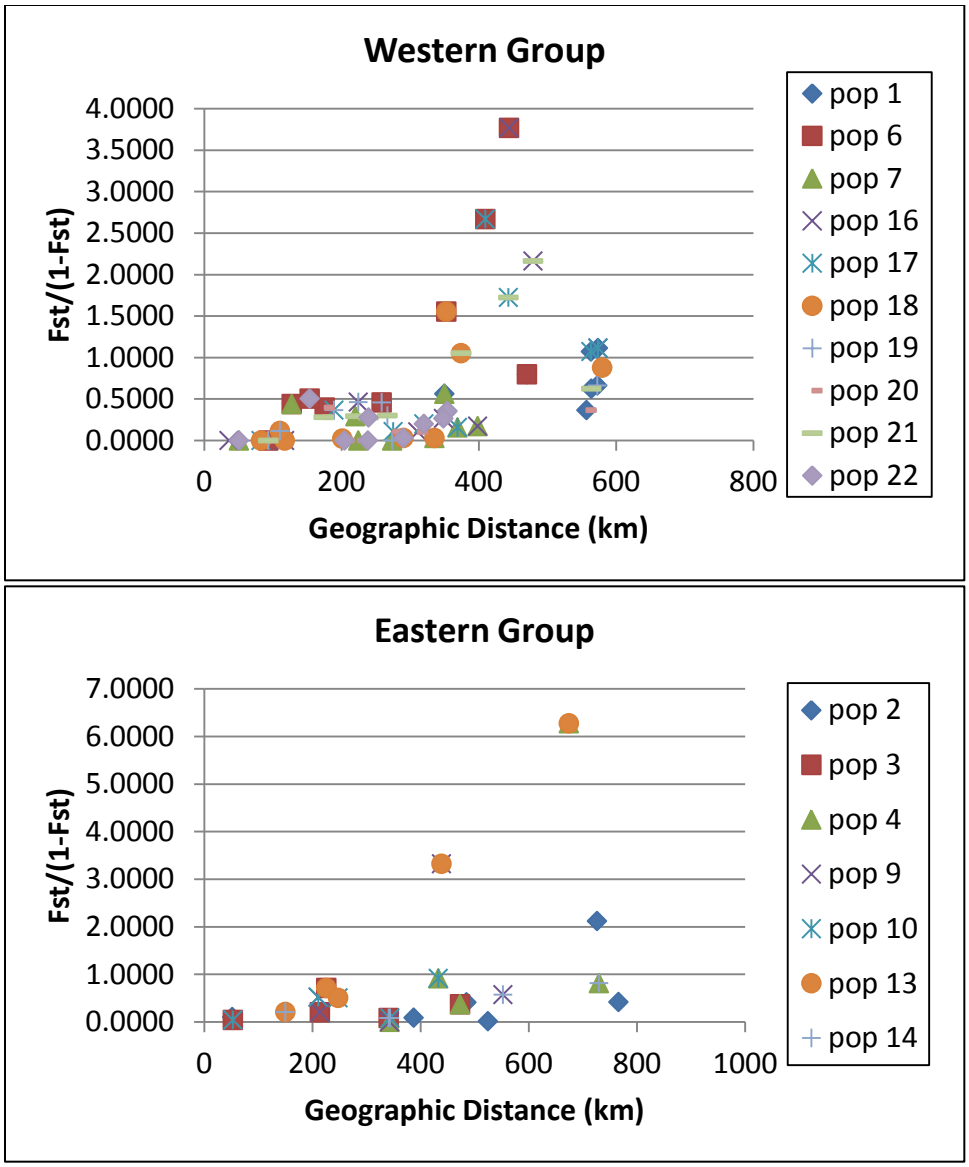

Figure 5 Mantel tests.

The relationship between geographic and genetic distances was compared for individuals in West and East geographic regions. P-values indicate a significant linear relationship between the two matrices.

REFERENCES

REFERENCES

Aukema, B. H., Carroll, A. L., Zheng, Y., Zhu, J., Raffa, K. F., Dan Moore, R., Stahl, K. and Taylor, S. W. (2008) Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate. *Ecography* 31: 348–358.

Breshears, D. D., N. S. Cobb, P. M. Price, C. D. Allen, R. G. Balice, W. H. Romme, J. H. Kastens, M. L. Floyd, J. Belnap, J. J. Anderson, O. B. Myers, and C. W. Meyer (2005) Regional vegetation die-off in response to global change-type-drought. *PNAS 102*: 15144-15148.

Byers, J. A. (2000) wind-aided dispersal of simulated bark beetles flying through forests. *Ecological modeling* 125: 231-243.

Carroll, A.L.; Taylor, S.W.; Régnière, J.; Safranyik, L. (2004) Effects of climate change on range expansion by the mountain pine beetle in British Columbia. *The Bark Beetles, Fuels, and Fire Bibliography.* Page 195.

Chapuis, M.P., Lecoq, M., Michalakis, Y., Loiseau, A., Sword, G. A., Piry, S. and Estoup, A. (2008) Do outbreaks affect genetic population structure? A worldwide survey in Locusta migratoria, a pest plagued by microsatellite null alleles. *Mol. Ecol.* 17: 3640–3653.

Chapuis, M.P., Loiseau, A., Michalakis, Y., Lecoq, M., Franc, A. and Estoup, A. (2009) Outbreaks, gene flow and effective population size in the migratory locust, *Locusta migratoria*: a regional-scale comparative survey. *Mol. Ecol.* 18: 792–800.

Chen, H., and Walton, A. (2011) Mountain pine beetle dispersal: spatiotemporal patterns and role in the spread and expansion of the present outbreak. *Ecosphere* 2:art66

Christiansen, E., Waring, R.H., Berryman, A.A. (1987) Resistance of conifers to bark beetle attack: searching for general relationships. *For. Ecol. Manage.* 22: 89-106.

Clement, M., Posada, D. and Crandall, K.A.(2000) TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657-1659.

Cognato, A.I., Seybold, S.J. and Sperling, F.A.H. (1999) In complete barriers to mitochondrial gene flow between pheromone races of the North American pine engraver, *ips pini* (Say). *Proc. R. Sco. Lond*. B 266: 1843-1850.

Cognato, A.I. and Harlin, A.D. and Fisher, M.L. (2003) Genetic structure among pinyon pine beetle populations (Scolytinae: *Ips confusus*). *Environmental Entomology* 32: 1262-1270.

Cognato, A. I., Gillette, N. E., Bolan˜os, R. C. and Sperling, F.A.H. (2005a) Mitochondrial phylogeny of pine cone beetles (Scolytinae, *Conophthorus)* and their affiliation with geographic area and host. *Mol. Phylogenet. Evol*. 36: 494-508.

Cognato, A. I., Sun, J.H., Anducho, M. and Owen, D. (2005b) Genetic variation and origin of

red turpentine beetles (*Dendroctonus valens* LeConte) introduced to the People's Republic of China. *Agric. For. Entomol.* 7: 87-94.

de la Giroday, H.M.C, Caroll, A.L., Lindgren, B.S. and Aukema, B.H. (2011) Incoming! Association of landscape features with dispersing mountain pine beetle populations during a range expansion event in western Canada. *Landscape Ecology* 26: 1097-1110.

Eager, T.J. (1999) Factors affecting the health of pinyon pine trees (*Pinus edulis*) in the pinyon-juniper woodlands of western Colorado. Pages 397-399 In: Monsen, S.B., Richards, S., Tausch, R.J., Miller, R. F., Goodrich, C., editors. Proc. Ecology and Management of Pinyon-Juniper Communities within the Interior West. *USDA Forest Service*, RMRS-P-9.

Eckert, C.G., Samis, K.E. and Lougheed, S.C. (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Mol. Ecol.* 17: 1170-1188.

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.

Excoffier, L., Laval, G., Schneider, S. (2005) Arlequin (Version 3.0): an integrated software package for population genetics data analysis. Ecol Bioinf 1: 47-50.

Excoffier, L. (2008) Analysis of Population Subdivision, in Handbook of Statistical Genetics. 3rd Edition. John Wiley & Sons, Ltd, Chichester, UK.

Fonseca, D.M., Widdel, A.K., Hutchinson, M., Spichiger, S.E. and Kramer, L.D. (2010) Fine-scale spatial and temporal population genetics of *Aedes japonicas*, a new US mosquito, reveal multiple introductions. *Mol. Ecol.* 19: 1559-1572.

Furniss, R.L., and Carolin, B.M. (1977) Western forest insects. *U.S. Dep. Agric. For. Serv. Misc. Publ*. 1339.

Gara, R.I. and Vite, J.P. (1962) Studies on the flight patterns of bark beetles (Coleoptera: Scolytidae) in second growth ponderosa pine forests. Contributions of Boyce Thompson Institute 21: 275-290.

Halsey, D., Guyon, J., Knight, J. Wang, S. (2011) Nevada Aerial Detection Survey Damage Areas. *USDA*. http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5358303.pdf

Holder, M., and P. O. Lewis. (2003) Phylogeny estimation: Traditional and Bayesian approaches. *Nat. Rev. Genet.* 4: 275-284.

Hudson, R. R., Slatkin, M., and Maddison, W.P., (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583-589.

Ibrahim, K.M. (2001) Plague dynamics and population genetics of the desert locust: can

turnover during recession maintain population genetic structure? *Mol. Ecol.* 10: 581-591.

Ibrahim, K.M., Nichols, R.A. and Hewitt, G.M. (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77: 282–291

Ibrahim, K.M., Sourrouille, P. and Hewitt, G.M. (2000) Are recession populations of the desert locust (*Schistocerca gregaria*) remnants of past swarms? *Mol. Ecol.* 9: 783-791.

Jactel, H., and Gaillard, J. (1991) A preliminary study of the dispersal potential of Ips sexdentatus (Boern) (Col, Scolytidae) with an automatically recording flight mill. *Journal of Applied Entomology* 112:138-145.

James, P.M.A., Coltman, D.W., Murray, B.W., Hamelin, R.C., Sperling, F.A.H. (2011) Spatial Genetic Structure of a Symbiotic Beetle-Fungal System: Toward Multi-Taxa Integrated Landscape Genetics. *PLoS ONE* 6(10): e25359.

Kelley, S.T., Mitton, J.B. and Paine, T.D. (1999) Strong differentiation in mitochondrial DNA of *Dendroctonus brevicomis* (Coleoptera: Scolytidae) on different subspecies of ponderosa pine. *Ann. Entomol. Soc. Am*. 92: 193- 197.

Kobayashi, T., Sakurai, T., Sakakibara, M. and Watanabe, T. (2011) Multiple origins of outbreak populations of a native insect pest in an agro-ecosystem. *Bulletin of Entomological Research* 101: 313-324.

Lanier, G. N. (1970) Biosystematics of North American *Ips* (Coleoptera: Scolytidae): Hopping's group IX. *Can. Entomol*. 102: 1139-1163.

Little, E.L., Jr. (1971) Atlas of United States trees: volume 1: conifers and important hardwoods. *U.S. Department of Agriculture Miscellaneous Publication* 1146, 9 p., 200 maps.

Lowe, W.H. (2009) What drives long-distance dispersal? A test of theoretical predictions. *Ecology* 90: 1456–1462.

Maddison, D.R. and Maddison, W.P. (2005) MacClade 4: Analysis of phylogeny and character evolution. Version 4.08a. http://macclade.org.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer res* 27: 209-220.

Menard, K.L., Cognato, A.I. (2007) Mitochondrial Haplotypic Diversity of Pine Cone Beetles (Scolytinae: Conophthorus) Collected on Food Sources. *Environmental Entomology* 36(4): 962-966.

Michalakis, Y. and Excoffier, L. (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*. 142: 1061-1064.

Negron, J.F., Wilson, J.L. (2003) Attributes associated with probability of infestation by the pinion ips, *Ips confusus* (Coleoptera: Scolytidae), in pinion pine, *Pinus edulis*. *West. N. Am. Nat.* 63: 440-451.

Nei, M. (1987) Molecular evolutionary genetics. Columbia University Press, New York.

Raymond, M. and Rousset, F. (1995) An exact test for population differentiation. *Evolution* 49: 1280-1283.

Robertson, C., Nelson, T.A., Jelinski, D.E., Wulder M.A. and Boots, B. (2009) Spatial–temporal analysis of species range expansion: the case of the mountain pine beetle, *Dendroctonus ponderosae*. *Journal of Biogeography* 36(8): 1446–1458.

Robinet, C. and Roques, A. (2010) Direct impacts of recent climate warming on insect populations. *Integrative Zoology* 5: 132–142.

Ronnås, C., Cassel-lundhagen, A., Battisti, A., Wallén, J. and Larsson, S. (2011) Limited emigration from an outbreak of a forest pest insect. *Mol. Ecol.* 20: 4606–4617.

Rousset (2000) Genetic differentiation between individuals. *Journal of Evolutionary. Biology* 13: 58-62.

Safranyik, L., Carroll, A. L., Regniere, J., Langor, D.W., Riel, W.G., Shore, T.L. et al. (2010) Potential for range expansion of mountain pine beetle into the boreal forest of North America. *Can. Entomol*. 142: 415-4442.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

Sokal, R.R. and Rohlf, F.J. (1995) Biometry: The principles and practice of statistics in biological research. 3rd edition. W.H. freman, New York.

Swofford, D.L. (2003) PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Stauffer, C., Lakatos, E. and Hewitt, G.M. (1999) Phylogeography and postglacial colonization routes of *Ips typographus* L. (Coleoptera, Scolytidae). *Mol. Ecol.* 8: 763-773.

Tajima F (1993) Simple methods for testing molecular clock hypothesis. *Genetics* 135:599-607.

Tamura, K. and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Ecol.* 10: 512-526.

Tao, J., Chen, M., Zong, S.X. and Luo, Y.Q. (2012) Genetic structure in the Seabuckthorn

Carpenter Moth (*Holcocerus hippophaecolus*) in China: The role of outbreak events, geographical and host factors. *PLoS ONE* 7: 1-7.

Tsui, C.K.M., Roe, A.D., El-Kassaby, Y.A., Rice, A.V., Alamouti, S.M., Sperling, F.A.H., Cooke, J.E.K., Bohlmann, J. and Hamelin, R.C. (2012) Population structure and migration pattern of a conifer pathogen, *Grosmannia clavigera*, as influenced by its symbiont, the mountain pine beetle. *Mol. Ecol.* 21: 71–86.

USDA-Forest Service (2004) Forest insect and disease conditions in the southwestern region, 2004. http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5238440.pdf

Vandergast, A.G., Gillespie, R.G and Roderick, G.K. (2004) Influence of volcanic activity on the population genetic structure of Hawaiian Tetragnatha spiders: fragmentation, rapid population growth and the potential for accelerated evolution. *Mol. Ecol*. 13: 1729-1743.

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

Williams, A. P., Allen, C. D., Millar, C. I., Swetnam, T. W., Michaelsen, J., Stilla, C. J. and Leavitt, S. W. (2010) Forest responses to increasing aridity and warmth in the southwestern United States. *Proceedings of the National Academy of Sciences of the United States of America* 107: 21289–21294.

Wood, D.L., Stark, R.W., Silverstein, R.M. and Rodin, J. O. (1967) Unique synergistic effects produced by the principal sex attractant compounds of *Ips confusus* (Leconte) (Coleoptera: Scolytidae). *Nature* 215: 206.

Wood, S.L. (1982) the bark and ambrosia beetles of North and Central America (Coleoptera: Scolytidae), a taxonomic monograph. *Great Basin Naturalist Memoirs* 6: 1359.