Genetic Structure Among Pinyon Pine Beetle Populations (Scolytinae: *Ips confusus*)

ANTHONY I. COGNATO,[1] APRIL DAWN HARLIN,[2] AND MARC L. FISHER

Department of Entomology, Texas A&M University, College Station, TX 77843–2475

Environ. Entomol. 32(5): 1262–1270 (2003)

ABSTRACT The bark beetle, *Ips confusus* (LeConte), provides an opportunity to study the influence of past geologic events and host use on lineage diversification of a herbivorous insect. This beetle mainly feeds on two pinyon pine (*Pinus*) species, but it has been collected from other conifer species. Isolation by host may contribute to lineage diversity and population structure. Alternatively, the repeat fragmentation of pine and beetle populations during the Pleistocene may explain population structure. We performed cladistic and nested clade analyses of mitochondrial DNA cytochrome oxidase I (mtDNA COI) sequence data from 95 *I. confusus* individuals collected from two hosts, *P. monophyllae* and *P. edulis*, and an atypical host, spruce (*Picea pungens*), from 10 western United States populations. Thirty-one most-parsimonious trees of 15 COI haplotypes demonstrated little association between host species and monophyletic groups. Thus, host use does not appear to contribute to genetic structure among populations. Nested clade analysis revealed three main haplotype lineages, each associated with eastern, southwestern, and western geographic localities. Estimation of the time to the most recent common ancestor places the start of lineage divergence during the Pleistocene (≈836,000 yr ago). Past glaciation events better explain genetic population structure among *I. confusus* populations.

KEY WORDS Coleoptera, phytophagous, host use, population genetics, nested clade analysis

THE PINYON PINE BARK beetle, *Ips confusus* (LeConte), presents an opportunity to study the influence of geographic history and host use on herbivorous insect population structure. This beetle is endophytophagous of dead and dying pinyon pine trees (*Pinus edulis* Engelmann and *P. monophylla* Torrey & Fremont) throughout the southwestern United States. Host selection behavior is poorly understood, but bark beetles tend to be host specific (Byers 1995). Secondary chemistry and physical properties of resin vary greatly among pine species and conifer genera (Gijzen et al. 1993, Byers 1995). Adaptations associated with consumption of these chemicals also tend to be host specific, which limits the ability of beetles to exploit alternative hosts. *I. confusus* rarely uses other pine species within its range (Lanier 1970).

In addition, pheromone production, which is, in part, the oxidation of host secondary chemicals, of *I. confusus* constraints the opportunity to use novel conifer hosts. Upon finding a suitable host, a male bores through the outer bark and into the phloem to excavate a nuptial chamber. Pheromones are produced as the male feeds, which attracts other males and females to the host (Wood 1982). Multiple females enter a nuptial chamber and mate, and each constructs an egg gallery. The brood completes its development under the bark, and the resulting adults may reinfest the brood tree. Three to four generations of this species can occur in 1 yr (Furniss and Carolin 1992). Individuals, which have completed their development in the late summer and autumn, gregariously hibernate under the bark of standing host trees (Chansler 1964).

The intricacies of their biology may influence the intraspecific genetic structure of *I. confusus.* A recent study of mitochrondrial cytochrome oxidase I DNA (mtDNA COI) sequence differences among *Ips* populations showed genetic differentiation between *I. confusus* individuals collected from *P. monophylla* and *P. edulis* (Cognato and Sperling 2000). Host selection may influence genetic divergence, as observed with other bark beetle populations (Kelley et al. 1999, Kerdelhué et al. 2002). In addition, recent collection of pinyon pine bark beetles from an unrelated nonhost, blue spruce (*Picea pungens* Engelmann) (A.I.C., unpublished data), suggests that host infidelity may contribute to genetic variation among *I. confusus* populations. Genetic isolation and deme formation may follow given continued use of the nonhost by subsequent surviving individuals (Mopper 1996). Given

[1] Address correspondence to: a-cognato@tamu.edu
[2] Department of Wildlife & Fisheries Sciences, Texas A&M University, College Station, TX 77843–2475.

0046-225X/03/1262–1270$04.00/0

enough time, this isolation would generate population structure (Avise 2000).

Nucleotide variation within mtDNA has revealed phylogenetic patterns among post-Pleistocene populations for many insects (Hewitt 1996), including *Ips* species (Cognato et al. 1999, Stauffer et al. 1999). Repeated fragmentation of pine populations during the Pleistocene (Betancourt 1987) provides an alternative explanation for lineage diversification among *I. confusus* populations. Hence, a more thorough exploration of genetic diversity, which accounts for host and geographic variation, is necessary to address possible causes of bark beetle lineage diversification (Cognato et al. 1999, Kelley et al. 1999, Kerdelhué et al. 2002).

Inferences of population history and gene flow based on measures of genetic variation (e.g., analysis of molecular variance [Fst], Wright 1951, Excoffier et al. 1992) or phylogenetic methods (Avise 1989) can be misleading. Measures of genetic variance do not account for historical relationships among populations (Bossart and Pashley 1998). In such cases, a lack of lineage sorting caused by recent coalescence will overestimate gene flow, even if populations are no longer exchanging effective migrants (Slatkin and Maddison 1989). Conversely, some phylogenetic methods (e.g., Avise 1989) do not account for allelic variance and are a poor predictor of gene flow. Other phylogenetic methods that consider allelic variation are limited by the requirement of phylogenetic resolution (Slatkin and Maddison 1989). Nested clade analyses developed by Templeton (1995) and Templeton et al. (1987, 1992) combine genetic variance and phylogenetic reconstruction to provide statistical tests for association of haplotype variation and geography. Specifically, this approach uses a statistical parsimony network and the number of steps linking haplotypes to examine the spatial distribution of haplotypes (Templeton 1998, Posada et al. 2000). Statistically significant associations are interpreted using an inference key (Templeton 1998) based on patterns predicted by coalescence theory (Crandall and Templeton 1993, Castelloe and Templeton 1994). With this technique, one can examine both population structure and history, and employ statistical tests of different predictions related to population range expansion and fragmentation.

This study examines genetic structure of *I. confusus* populations to test for associations among beetle lineages, geographic location, and hosts. We obtained a partial mtDNA COI sequence from 95 individuals collected from 10 populations throughout the southwestern United States and California within the distribution of pinyon pine hosts. Individuals were collected from both hosts and a nonhost, blue spruce (*P. pungens*). Phylogenetic analysis of these data reveals no association of hosts with clades of individual mtDNA haplotypes. Nested clade analysis, however, suggests that population genetic structure is associated with geographic isolation.

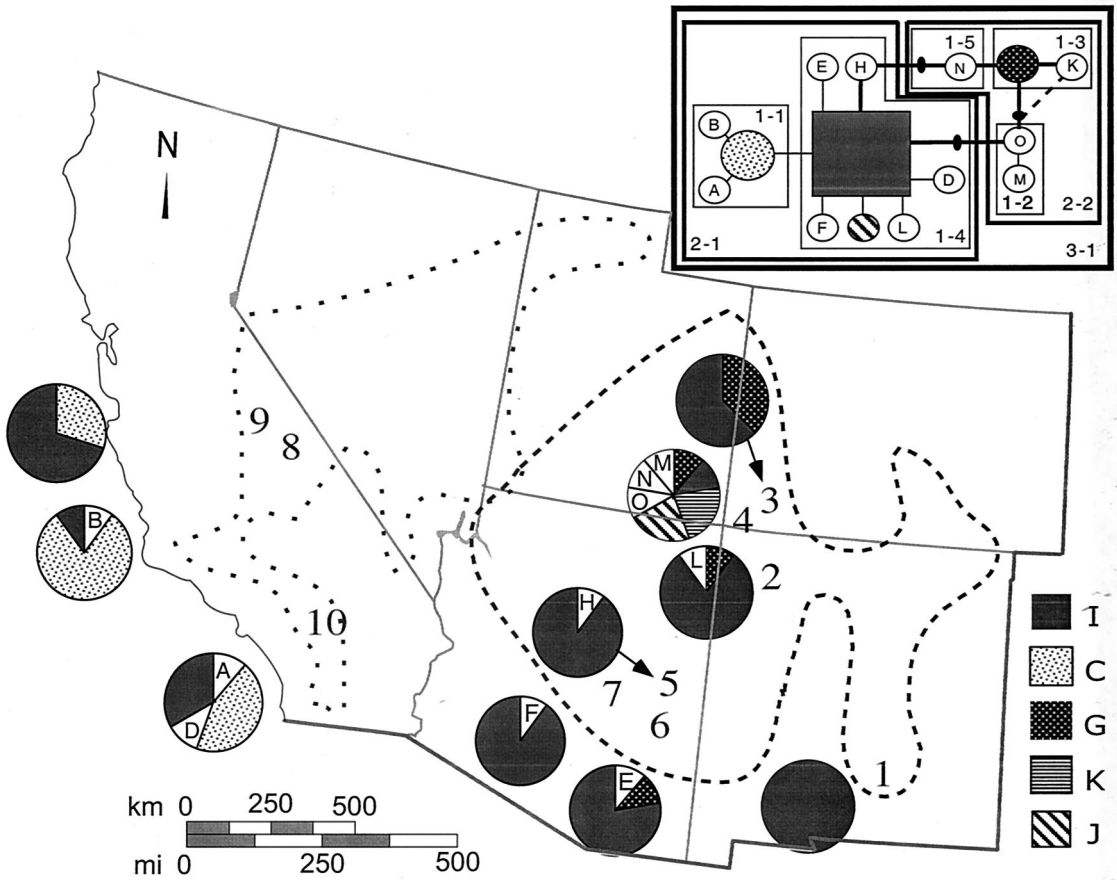
Materials and Methods

Specimens and Molecular Protocols. Eight to ten beetles were sampled from each of ten locations (Fig. 1). Live beetles were excised from infested host trees with a knife and forceps and stored in 100% ethanol. Where possible, specimens were removed from different areas of the host tree to ensure that different broods were sampled. The presence of 20 *I. confusus* individuals (1:1 sex ratio) within blue spruce was an usual discovery. These individuals were concentrated in an approximate 6 × 4-cm region of the cambium bark layer, and there was no evidence of nuptial chambers or larval galleries created by these individuals. Only 10 individuals were sampled for this study. Ten individuals collected in 1996 from Gila County, Arizona, which were curated as pinned museum specimens, were included in the study. A.I.C. identified all beetles. Total genomic DNA was extracted from beetle thoraces, which were removed from the rest of the body, with a silica-based spin column procedure (i.e., Qiamp, Qiagen, Santa Clara, CA), following the manufacturer's tissue protocol. The remaining body parts were glued to a mounting board and pinned. These specimens were vouchered in the Texas A&M University Department of Entomology insect collection.

A region of ≈400 nucleotides of mtDNA COI, beginning at the equivalent of nucleotide 2194 of *Drosophila yakuba* mtDNA COI (Clary and Wolstenholme 1985), was amplified via the polymerase chain reaction (PCR) with primers C1-J-2183 and C1-N-2611 following the methods of Cognato and Sperling (2000). Each PCR reaction contained: 35 μl ddH_2O, 5 μl 10× *Taq*DNA polymerase buffer (Promega, Madison, WI), 4 μl 25 mM Promega $MgCl_2$, 1 μl 40 mM deoxynucleotide triphosphates (dNTPs), 2 μl of each 5 mM oligonucleotide primer, 0.2 μl of Promega *Taq*DNA polymerase, and 1 μl 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°C, 0.75 min at 45°C, 1 min at 72°C, followed by 34 cycles of 0.5 min at 94°C, 0.75 min at 45°C, 1 min at 72°C, and a final elongation cycle of 5 min at 72°C.

Unincorporated dNTPs and oligonucleotides were removed from PCR with a Qiaquick PCR Purification Kit (Qiagen) and were directly sequenced on an ABI 377 automated sequencer after a BigDye (Applied Biosystems, Foster City, CA) fluorescent chemistry reaction. Both sense and antisense strands were sequenced for all individuals. A reference sequence was submitted to GenBank (AY306137).

Sequence Analysis. Alignment of individual sequences was straightforward because of complete amino acid conservation. No nucleotide insertions or deletions were observed. Cladistic and branch support analyses of the sequence data were implemented with the program PAUP (Swofford 1998). Cladograms were generated under a parsimony optimality criterion. A heuristic search of potential cladograms was performed with 50 replicates of random stepwise ad-



Population	State:County	Latitude, Longitude	Host	Haplotype (individuals)
1. NM1.edul	New Mexico:Otero	32:53N, 105:30W	*P. edulis*	I (10)
2. NM2.edul	New Mexico:Sandoval	36:01N, 106:57W	*P. edulis*	G (1), I (8), L(1)*
3. CO1.edul	Colorado:Dolores	37:45N, 108:00W	*P. edulis*	G (3), I (5)
4. CO2.edul	Colorado:Montezuma	37:28N, 108:29W	*P. edulis*	G (1), I (1), J (2) K(2), M(1), N (1) O(1)
5. AZ1.S	Arizona:Greenlee	33:38N, 109:20W	*P. pungens*	H(1)*, I (9)
6. AZ2.edul	Arizona:Greenlee	33:10N, 109:23W	*P. edulis*	E (1)*, G (1), I (7)
7. AZ3.edul	Arizona:Gila	33:36N, 110:15W	*P. edulis*	F (1)*, I (9)
8. CA1.mono	California:Mono	38:05N, 119:10W	*P. monophylla*	B (1)*, **C** (8), I (1)
9. CA2.mono	California:Inyo	37:15N, 118:10W	*P. monophylla*	**C** (3), I (7)
10. CA3.mono	California: San Bernardino	34:18N, 116:49W	*P. monophylla*	A (1)*, **C** (4), D (1)*, I (3)

Fig. 1. Geographic location of *I. confusus* populations sampled in western United States (numbers) and corresponding haplotypes (pie graphs to the left). Ranges of the host plants *P. monophylla* (dotted line) and *P. edulis* (dashed line) are restricted between 610 and 3030 m elevation. *P. pugens* occurs at >3030 m elevation in the Southern Rocky Mountains. Asterisk indicates unique haplotypes, and bold text emphasizes the restriction of haplotype C to California. For comparison, the nested clade of haplotypes is shown above the map. Major lineages L1, L2, and L3 correspond to nested clades 1–1, 1–4, and 2–2, respectively. See Fig. 4 for details of the nested clade analysis.

Haplotype	2212	2326	2332	2335	2365	2392	2449	2470	2521	2527	2556	2566	2581	2590
I	C	A	A	C	A	T	T	T	A	A	T	T	C	G
A	*	G	*	*	*	*	*	A	*	*	*	*	*	*
B	*	G	*	T	*	*	*	*	*	*	*	*	*	*
C	*	G	*	*	*	*	*	*	*	*	*	*	*	*
D	*	*	*	*	T	*	*	*	*	*	*	*	*	*
E	*	*	*	*	*	*	*	*	*	G	*	*	*	*
F	*	*	*	*	*	C	*	*	*	*	*	*	*	*
G	*	*	G	*	*	*	C	*	G	*	*	*	*	A
H	*	*	*	*	*	*	*	*	*	*	*	*	*	A
J	*	*	*	*	*	*	*	*	*	*	*	*	A	*
K	*	*	G	*	*	*	C	*	G	*	G	*	?	?
L	*	*	*	*	*	*	*	*	*	*	*	C	*	*
M	T	*	*	*	*	*	C	*	G	*	*	*	*	*
N	*	*	G	*	*	*	C	*	*	*	*	*	*	A
O	*	*	*	*	*	*	C	*	G	*	*	*	*	*

Fig. 2. Haplotype differences at nucleotide positions. Haplotype I is the reference sequence, GenBank accession AY306137. Numbering corresponds to homologous nucleotide positions in *D. yakuba* COI (Clary and Wolstenholme 1985). Invariant nucleotide positions are not shown. An asterisk indicates the same nucleotide as the reference sequence, and "?" indicates missing data.

dition of haplotypes and branch swapping via tree bisection-reconnection. All other settings were default. Bootstrap proportions were determined with 500 replicates and default PAUP settings. Branch support (Bremer 1994) for individual nodes was assessed using constraint trees generated with TreeRot (Sorensen 1996). Nucleotide diversity (π, Nei 1987) was calculated with Matrix 2.0 (Posada 2001).

Nested Clade Analysis. The frequency and spatial distribution of the mtDNA haplotypes were used to test the null hypothesis of random association of maternal lineages and geographic location. A statistical parsimony network (Templeton et al. 1992) of the 95 COI sequences was created with the program TCS (v.1.13, Clement et al. 2000). Haplotypes in the 0-step network were nested in a hierarchical series of 1-step, 2-step, etc., clades until the entire network was nested into a single clade following the rules outlined in Templeton et al. (1992). The resulting nested network was used to perform statistical tests of geographic association of nested clades with the program GeoDis (v. 2.0, Posada et al. 2000). Nested clade distance (D_c) is the average distance (km) between the center of geographic distribution of clades (or haplotypes) within a nested level and the geographic center of the distribution of all members of that clade that bear a particular haplotype. D_c was calculated for each clade (or haplotype) within the nested network. Similarly, the nested clade distance (D_n) was calculated as the geographic distance between the geographic center of distribution of the members of a clade and the next level nested clade. D_c and D_n, as well as the geographic distance between members of interior and tip clades (or haplotypes) (I-T), were randomly permuted 10,000 times to create a null distribution for comparison with observed values. Latitude and longitude coordinates were used to construct a pairwise distance matrix to the nearest kilometer between collection sites (Fig. 1). This distance matrix was used by the program GeoDis to estimate D_c, D_n, and I-T distances. Historical processes were inferred with a standardized key (Templeton 1998, http://zoology.byu.edu/crandall_lab/geodis.htm) based on the distances within clades that were significantly larger or smaller than expected at the $P \leq 0.05$ level.

Divergence Time Estimates. The method of Tang et al. (2002) was used to estimate the time since divergence between those clades in the nested network that showed significantly nonrandom geographic distribution of haplotypes. This method uses the average number of pairwise differences between sequences and an assumed rate of molecular substitution to estimate the time to the most recent common ancestor (TMRCA) based on the assumptions of infinite sites and equal probability of substitution among sites. Cladistic analysis of multiple data sets suggests that *I. hoppingi* Lanier is the most recent ancestor of *I. confusus* (Cognato and Vogler 2001). Sequences were corrected for transition/transversion bias with a two-parameter rate matrix (Kimura 1980) because most functional proteins show a bias toward third codon transitions. Time in millions of years ($\hat{T}$) was estimated from these distances with the formula $\hat{T} = \frac{D}{2\mu l}$, where μ is the number of substitutions per site, and l is the length of the gene in base pairs. D is the average pairwise distance between two clades, L and R, and is estimated by

$$D = \frac{1}{pq} \sum_{i \in L} \sum_{j \in R} d_{ij},$$

where clades L and R contain sequences $(l_1, l_2, \ldots l_p)$ and $(r_1, r_2, \ldots r_q)$ respectively, $p + q = n$, and d_{ij} is the number of nucleotide differences between sequences i and j. The rate of nucleotide substitution for this locus (μ) was assumed to be 0.023 substitutions per site per million years following previous estimates of substitution rates for this locus among insects (Brower 1994).

Results

A total of 15 mtDNA COI haplotypes (Fig. 2) was found among 95 *I. confusus* individuals, which equates to a moderate haplotype diversity ($h = 0.58$). Nucle-

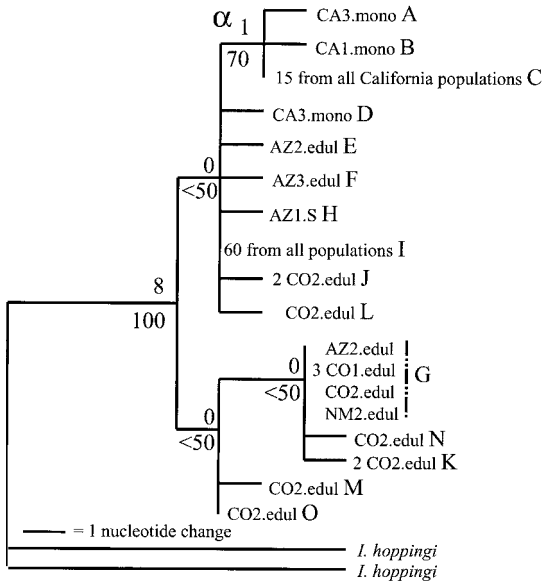


Fig. 3. One of 31 parsimonious trees for 95 *I. confusus* individuals sampled from 10 populations in western United States. *I. hoppingi* individuals were designated as outgroups. Numbers above and below lines are branch support measures and bootstrap proportions, respectively. Rescaled consistency index = 0.8. Population designations as in Fig. 1, and haplotype designations (A-O) as in Fig. 2. A numeral preceding a population designation indicates more than one individual. Clade α is discussed in text.

otide diversity (π) was low (0.011 ± 0.015), corresponding to a mean pairwise nucleotide difference among individuals of 3.1 ± 1.3. The uncorrected sequence difference among haplotypes ranged from 0.25 to 1.6%. Substitutions at third codon positions comprised all nucleotide changes. Populations averaged 2.9 haplotypes; however, population NM1 was represented by only one haplotype, while seven haplotypes were present in population CO2 (Fig. 1). Over half of the populations contain a unique haplotype (Fig. 1).

Cladistic analysis of *I. confusus* and *I. hoppingi* individuals resulted in 31 most-parsimonious trees based on 14 parsimony-informative nucleotide positions (Fig. 3). Only one clade (α) was resolved in a strict consensus of these trees (Fig. 3). Although resolution among individuals is low, little homoplasy is observed (rescaled consistency index, RC = 0.8). Low branch support is most likely a consequence of few informative characters. The supported clade (α) contains individuals from California populations that are represented by three haplotypes. Relationships of the remaining individuals from all populations and host plants are unresolved. We failed to observe any association among clades, between individual mtDNA haplotypes and host species (Fig. 3).

The statistical parsimony analysis confirmed the results of the cladistic analysis, but provided more phylogenetic resolution among haplotypes (Fig. 4). The parsimony network was subsequently nested into a 3-step network defined by three major geographic lineages, L1, L2, and L3, with haplotypes distributed in California, Arizona/New Mexico, and Colorado, respectively (Figs. 3 and 4). Alternate equally parsimonious connections between L2 and L3 were observed within the 0-step network. These reticulations could not be resolved unambiguously, but multiple alternative resolutions did not affect the inferences of population history derived from nested clade analysis.

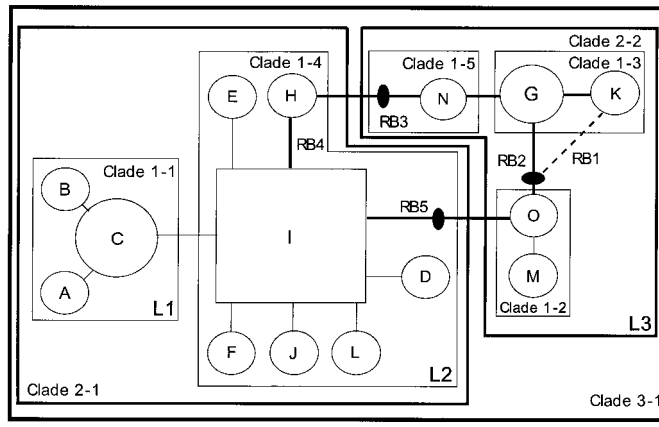


Fig. 4. Nested network of *I. confusus* haplotypes. Hollow circles and squares indicate individual haplotypes with frequencies proportional to size. A square designates the haplotype with the greatest root probability. Filled ovals represent hypothetical intermediate haplotypes inferred from statistical parsimony analysis. Lines between haplotypes and intermediates represent one nucleotide substitution. Heavy lines indicate reticulations within the network. RB1-RB5 designate alternative points within the network that could be broken to resolve network ambiguity. Dashed line indicates a network connection that was broken before nesting of haplotypes. Reticulations were resolved following the rules of Templeton et al. (1992) or Crandall and Templeton (1993). Other reticulations were left unresolved during analysis of and did not alter inferences derived from network patterns. Major lineages, L1, L2, and L3 correspond to nested clades 1–1, 1–4, and 2–2, respectively.

Table 1. Summary of inferences regarding demographic events deduced from clades with significant nested clade values

Clade	χ^2	Nested clades	D_c	D_n	Chain of inference	Demographic event
1–4	n.s. ($P = 0.08$)	J (TIP)	n.s.	n.s.	1 yes, 2 yes, 3 no, 4 no	Restricted gene flow, isolation by distance
		F (TIP)	n.s.	n.s.		
		E (TIP)	n.s.	n.s.		
		D (TIP)	n.s.	n.s.		
		L (TIP)	n.s.	n.s.		
		H (INT)	n.s.	n.s.		
		I (INT)	n.s.	n.s.		
		I-T	513.41, >, $P = 0.031$	n.s.		
2-1	$P < 0.0001$	1-1 (TIP)	267.63, <, $P < 0.001$	751.32, >, $P < 0.001$	1 yes, 2 yes, 3 yes, 5 no, 6 yes, 13 yes	Long distance colonization
		1-4 (INT)	511.66, <, $P < 0.001$	583.23, <, $P < 0.001$		
		I-T	244.03, >, $P = 0.003$	−168.10, <, $P < 0.001$		
Total cladogram	$P < 0.001$	2-1 (TIP)	622.13, >, $P = 0.039$	613.93, >, $P = 0.007$	1 yes, 2 no, 11 yes, 12 no	Contiguous range expansion
		2-2 (TIP)	157.69, <, $P < 0.001$	519.04, <, $P = 0.035$		

TIP = tip clade, INT = interior clade, n.s. = nonsignificant, I-T = interior tip. Greater-than or less-than symbols indicate a D_c or D_n value that is significantly larger or smaller than expected if haplotypes were distributed randomly. *P* values indicate probability that the D_c or D_n estimated from the data is observed by chance. Inferences were drawn from the nested clade analysis interpretation key of Templeton (1998). The steps in the chain of inference can be examined by comparison with this key.

Consequently, all reticulations remained unbroken during the nesting process with the exception of the reticulation between haplotypes K and G, which was unambiguously broken following the rules of Crandall and Templeton (1993). L1, L2, and L3 were significantly associated with geographic location (Fig. 1; Table 1). Lineage 2 contained the most common and widely distributed haplotype I, with the highest frequency in Arizona and New Mexico (Fig. 1). Haplotypes within L1, i.e., C, A, B, were found only in California, whereas haplotype G (L3), and haplotypes with one mutational step from G were recovered predominantly from populations in southwest Colorado. The nested clade analysis provided significant evidence of a westward range expansion ($P < 0.05$) and a long-distance colonization of maternal lineage L2 into California ($P = 0.001$) (Table 1). The western expansion of L2 was followed by increased long-distance genetic isolation between southwestern and western populations, resulting in the formation of a distinct maternal lineage in California (L1) (Table 1; Fig. 5). Estimates of TMRCA of L1 suggest the California population began isolation from other regions as early as 836,000 yr ago (Fig. 5). Within California, there appear to have been expansion and subsequent establishment of populations after the establishment of L1 in this region (Figs. 1 and 5). The average pairwise differences between L2 and L3 suggest the most recent isolation event between the Colorado and Arizona/New Mexico regions began 226,000 yr ago (Fig. 5). More recently, there has been, and continues to be, isolation by distance along a continuum of population distribution from Colorado (L3), through the southwest (L1), and into California (L2) (Table 1; Figs. 1 and 5).

Discussion

Coalescence theory suggests that a nested clade must be as old as or older than all other clades (or haplotypes) nested within it (Templeton 1998), and the haplotype with the greatest frequency and widest geographic distribution is the most likely root of the haplotype network (Castelloe and Templeton 1994). Haplotype frequency, nesting level, and position within the network can be used to infer the relative timing and geographic origin of demographic events that have affected the distribution of haplotypic variation. Therefore, the high frequency of haplotype I in Arizona and New Mexico suggests that the maternal lineages of *I. confusus* originated in the southwestern United States. Inferences from nested clade analysis and estimates of divergence times of L1, L2, and L3 suggest that L2, distributed predominately in Arizona and New Mexico, expanded its range west into California ≈836,000 yr ago.

The statistical parsimony procedure failed to unambiguously resolve the relationship between the Arizona/New Mexico and Colorado lineages. The presence of reticulated connections and multiple inferred intermediates between L2 and L3 suggests that Colorado and southwestern regions have shared a longer and more complex history, perhaps defined by multiple periods of genetic exchange followed by isolation, than either has had with California (Fig. 4). The contact between these two regions could have been in the form of migratory events, such as long-distance dispersal of females, or the expansion and contraction of these populations during periods change in host distribution. We are unable with the results of the nested clade and phylogenetic analyses to discrimi-

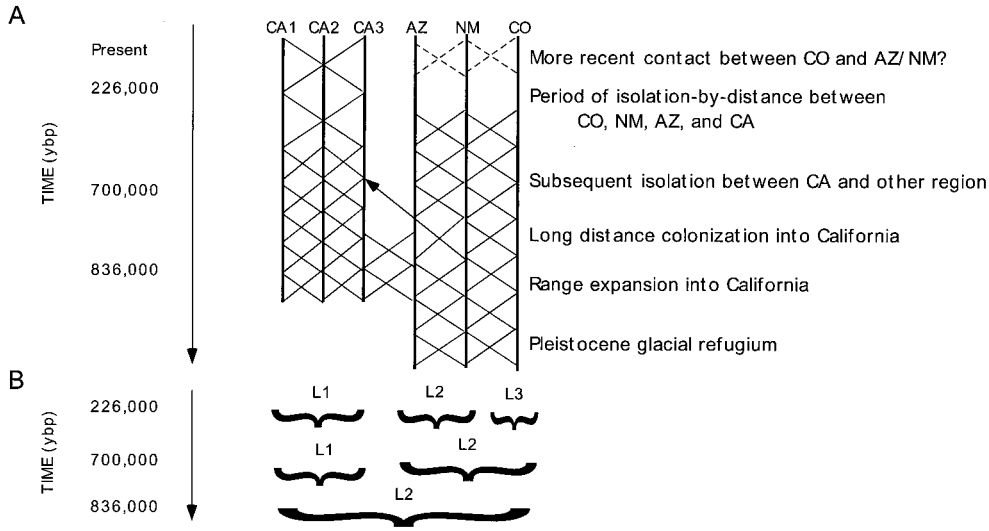


Fig. 5. Summary of *I. confusus* population history. (A) Vertical lines represent geographic regions. Crossed lines represent genetic exchange between regions. Time is presented as TMRCA in years before present (ybp). Mid-Pleistocene populations of *I. confusus* were most likely isolated to lower latitudes and altitudes of the southwestern United States and the Sonoran and Chihuahuan Deserts within a glacial refugium of *Pinus* species. The postglacial expansion of *Pinus* facilitated a coexpansion of *I. confusus*, including a westward movement of beetles into California 700,000–840,000 yr ago. Subsequent to this expansion to the west, isolation by distance resulted in genetic divergence of *I. confusus* lineages in Colorado, New Mexico, Arizona, and California. Statistical parsimony network suggests a more complicated and older shared history between Arizona/New Mexico and Colorado populations, characterized by multiple periods of isolation and contact. More recent periods of contact may be possible between these regions, but are too recent to detect with nested clade analysis. (B) Lineage divergence of *I. confusus*. Brackets define geographic distribution of lineages. Time (ybp) matches demographic events presented in (A). Westward expansion into California resulted in the divergence of L2 and the formation of a new maternal lineage L1. Multiple periods of isolation between Colorado and Arizona/New Mexico have resulted in the formation of L3; however, reestablishment of genetic exchange between these regions has complicated the evolutionary history of these two lineages.

nate between these and other potential hypotheses regarding the history of these two populations.

Changes in host distribution and TMRCA of *I. confusus* lineages suggest Pleistocene distribution of pinyon pines is likely to have influenced the phylogeographic pattern of *I. confusus* maternal lineages. Estimates of TMRCA between lineages suggest diversification of *I. confusus* lineages began in the middle of the Pleistocene (Fig. 4). During this time, North America experienced cycles of intense glaciation (≈100,000 yr) with short warmer periods (≈10,000 yr) (Webb and Bartlein 1992). At glacial maximums, a cool/wet climate predominated in the southwestern United States and northern Mexico. Based on fossilized rat middens and pollen, pinyon pines were distributed at lower altitudes and latitudes and more to the east, including the present-day southern Sonoran and Chihuahuan Deserts (Thompson and Anderson 2000, Metcalfe et al. 2002) and western Nebraska (Williams et al. 2000). These data and the presence of haplotype I (Fig. 1) in all beetle populations with the highest frequency in Arizona and New Mexico suggest a past concentration of pinyon pine and beetle populations in the southestern United States and northern Mexico (Betancourt 1987). The history of L1, L2, and L3 inferred from nested clade analysis agrees with this scenario. The nested clade analysis provides evidence of a possible second refugium in California (Wells 1987, Williams et al. 2000) (Fig. 1), as well as periods of contact and isolation between Colorado and the southwest evidenced by the more complex pattern of connections between these two regions (Figs. 4 and 5).

Our data suggest host specificity has not greatly influenced lineage diversification of *I. confusus* (Fig. 3). Strong selection pressure imposed by alternative hosts would promote an exclusive association of monophyletic groups of individuals with each plant species regardless of geographic locality. Indeed, genotypic, morphological, and behavioral differences have been observed among herbivorous insect populations that feed on different hosts (e.g., Feder et al. 1988, Roininen et al. 1993, Funk 1998, Scheffer and Wiegmann 2000). However, many of these examples are of herbivores that use living hosts for several generations. Use of an individual plant will most likely have a greater effect on herbivore evolution because successive insect generations are continually exposed to a plant's particular selection pressures (Mopper 1996). Perhaps herbivores that use dead or dying hosts, for only one or few generations, are less likely to become genetically isolated because any novel selection pressure imparted by a nonhost would be diminished through the use of other hosts in subsequent generations. Thus, minimal population structure is caused by host selection.

I. confusus individuals may occasionally use a nonhost for a hibernation roost, as observed for the AZ1.S population (Chansler 1964). In such cases, a nonhost, like spruce, would impose minimal selection on the beetles, assuming that they subsequently chose a pinyon host for their brood. However, the dense congregation of 20 *I. confusus* males and females found among broods of *I. hunteri* suggests that production of aggregation pheromone occurred in spruce. Other *Ips* species can produce pheromones in nonhosts (Elkinton et al. 1980), so adults may potentially attract mates while feeding in spruce, with brood production after hibernation. Although reduced, brood production in nonhosts has been observed for bark beetle species under laboratory conditions (e.g., Langor et al. 1990). A population of *I. confusus* may persist in spruce, but genetic divergence was not detected with the mtDNA sequence we examined. More variable genetic markers, such as microsatellites and AFLPs, or fitness-specific gene loci are needed to further test this hypothesis (e.g., McMichael and Pashley-Prowell 1999, Ting et al. 2000).

Acknowledgments

We thank Laura Merrill, Scott Kelley, and Robert Trujillo for samples; Indira Kuriachan for laboratory assistance; and Janice Redio and Jessica Usener for editorial review. Reviews by Thomas Sappington and two anonymous reviewers improved this study. This study was funded by a Texas A&M University, Office of the Vice President of Research, Faculty Mini Grant.

References Cited

Avise, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution 43: 1192–1208.

Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA.

Betancourt, J. L. 1987. Paleoecology of pinyon-juniper woodlands: summary, pp. 129 139. *In* R. L Everett (ed.), Proceedings, Pinyon-Juniper Conference, United States Department of Agriculture Forest Service General Technical Report INT-215, 13–16 January 1986, Reno, NV. Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Bossart, J. L., and D. P. Pashley. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. TREE 13: 202–206.

Bremer, K. 1994. Branch support and tree stability. Cladistics 10: 295–304.

Brower, A.V.Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA 91: 6491–6495.

Byers, J. A. 1995. Host-tree chemistry affecting colonization in bark beetles, pp. 154–213. *In* R. T. Carde and W. J. Bell (eds.), Chemical ecology of insects 2. Chapmann & Hall, New York.

Castelloe, J., and A. R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. 3: 102–113.

Chansler, J. F. 1964. Overwintering habits of *Ips lecontei* Sw. and *Ips confusus* (Lec.) in Arizona and New Mexico: United States Department of Agriculture Forest Service Research Note RM-27. Rocky Mountian Forest and Range Experiment Station, Fort Collins, CO.

Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organization and genetic code. J. Mol. Evol. 22: 252–271.

Clement, M., D. Posada, and K. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9: 1657–1660.

Cognato, A. I., and F.A.H. Sperling. 2000. Phylogeny of *Ips* DeGeer species (Coleoptera: Scolytidae) inferred from mitochondrial cytochrome oxidase I DNA sequence. Mol. Phylogenet. Evol. 14: 445–460.

Cognato, A. I., and A. P. Vogler. 2001. Exploring data interaction and nucleotide alignment in a multiple gene analysis of *Ips* (Coleoptera: Scolytinae). Syst. Biol. 50: 758–780.

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

Crandall, K. A., and A. R. Templeton. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134: 959–969.

Elkinton, J. S., D. L. Wood, and L. B. Hendry. 1980. Pheromone production by the bark beetle *Ips paraconfusus,* in the nonhost, white fir. J. Chem. Ecol. 6: 979–987.

Excoffier, L., P. Smouse, and J. Quattro. 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., C. A. Chilcote, and G. L. Bush. 1988. Genetic differentiation between sympatric host races of the apple maggot fly, *Rhagoletis pomonella.* Nature 336: 61–64.

Funk, D. J. 1998. Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. Evolution 52: 1744–1759.

Furniss, M. L., and V. M. Carolin. 1992. Western forest insects: United States Department of Agriculture Forest Service Miscellaneous Publication No. 1339, Washington, DC.

Gijzen, M., E. Lewinsohn, T. J. Savage, and R. B. Croteau. 1993. Conifer monoterpenes: biochemistry and bark beetle chemical ecology, pp. 8–22. *In* R. Teranishi, R. G. Buttery, and H. Sugisawa (eds.), Bioactive volatile compounds from plants: ACS Symposium Series, 525. American Chemical Society, Washington, DC.

Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58: 247–276.

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

Kerdelhué, C., G. Roux-Morabito, J. Forichon, J. M. Chambon, A. Robert, and F. Lieutier. 2002. Population genetic structure of *Tomicus piniperda* L. (Curculionidae: Scolytinae) on different pine species and validation of *T. destruens* (Woll.). Mol. Ecol. 11: 483–494.

Kimura, M. 1980. A simple model for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.

Langor, D. W., J. R. Spence, and G. R. Pohl. 1990. Host effects on fertility and reproductive success of *Dendroc-*

tonus ponderosae Hopkins (Coleoptera: Scolytidae). Evolution 44: 609–618.

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

McMichael, M., and D. Pashley-Prowell. 1999. Differences in amplified fragment length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strains Ann. Entomol. Soc. Am. 92: 175–181.

Metcalfe, S., A. Say, S. Black, R. McCulloch, and S. O'Hara. 2002. Wet conditions during the last glaciation in the Chihuahuan Desert, Alta Babicora Basin, Mexico. Quaternary Res. 57: 91–101.

Mopper, S. 1996. Adaptive genetic structure in phytophagous insect populations. TREE 11: 235–238.

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

Posada, D., K. A. Crandall, and A. R. Templeton. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Mol. Ecol. 9: 487–488.

Posada, D. 2001. MATRIX 2.0: Department of Zoology, Brigham Young University. http://zoology.byu.edu/crandall_lab/programs.htm/programs.htm.

Roininen, H., J. Vuorinen, J. Tahvanainen, and R. Julkunen-Tiitto. 1993. Host preference and allozyme differentiation in shoot galling sawfly, *Euura atra.* Evolution 47: 300–307.

Scheffer, S., and B. M. Wiegmann. 2000. Molecular phylogenetics of the holly leafminers (Diptera: Agromyzidae: Phytomyza): species limits, speciation, and dietary specialization. Mol. Phylogenet. Evol. 17: 244–255.

Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from phylogenies of alleles. Genetics 123: 603–613.

Sorensen, M. D. 1996. TreeRot. University of Michigan, Ann Arbor, MI.

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

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods), 4.0b4a. Smithsonian Institute, Washington, DC.

Tang, H., D. O. Siegmund, P. Shen, P. J. Oefner, and M. W. Feldman. 2002. Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. Genetics 161: 447–459.

Templeton, A. R. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum.* Genetics 140: 767–782.

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol. Ecol. 7: 381–397.

Templeton, A., E. Boerwinkle, and C. Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila.* Genetics 117: 343–351.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III: cladogram estimation. Genetics 132: 619–633.

Thompson, R. S., and K. H. Anderson. 2000. Biomes of western North America at 18,000, 6000 and 0 ^{14}C yr BP reconstructed from pollen and packrat midden data. J. Biogeography 27: 555–584.

Ting, C. T., S. C. Tsaur, and C. I. Wu. 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus.* Proc. Natl. Acad. Sci. USA 97: 5313–5316.

Webb III, T., and P. J. Bartlein. 1992. Global changes during the last 3 million years: climate controls and biotic responses. Annu. Rev. Ecol. Syst. 23: 141–173.

Wells, P. V. 1987. Systematics and distribution of pinyons in the late quaternary, pp. 104 108. *In* R. L. Everett (ed.), Proceedings, Pinyon-Juniper Conference, United States Department of Agriculture Forest Service General Technical Report INT-215, 13–16 January 1986, Reno, NV. Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Williams, J. W., T. Webb III, P. H. Richard, and P. Newby. 2000. Late quaternary biomes of Canada and the eastern United States. J. Biogeography 27: 585–607.

Wood, D. L. 1982. The role of pheromones, kairomones, and allomones in the host selection and colonization behavior of bark beetles. Annu. Rev. Entomol. 27: 411–446.

Wright, S. 1951. The genetical structure of populations. Ann. Eugenics 15: 323–354.

Received for publication 2 April 2003; accepted 23 June 2003.