2006-07-12

Phylogeographic Analyses of Obligate and Facultative Cave Crayfish Species on the Cumberland Plateau of the Southern Appalachians

Jennifer Elizabeth Buhay
Brigham Young University - Provo

PHYLOGEOGRAPHIC ANALYSES OF OBLIGATE AND FACULTATIVE CAVE CRAYFISH SPECIES ON THE CUMBERLAND PLATEAU OF THE SOUTHERN APPALACHIANS

by

Jennifer E. Buhay

A dissertation submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Integrative Biology

Brigham Young University

August 2006

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Jennifer E. Buhay

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

_____	_____
Date	Keith A. Crandall, Chair
_____	_____
Date	Russell B. Rader
_____	_____
Date	Matthew F. Bekker
_____	_____
Date	Jack W. Sites
_____	_____
Date	Horton H. Hobbs III

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Jennifer E. Buhay in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

_____ _____

Date Keith A. Crandall
Chair, Graduate Committee

Accepted for the Department

_____

Larry St. Clair
Department Chair

Accepted for the College

_____

Rodney Brown
Dean, College of Biology and Agriculture

ABSTRACT

PHYLOGEOGRAPHIC ANALYSES OF OBLIGATE AND FACULTATIVE CAVE CRAYFISH SPECIES ON THE CUMBERLAND PLATEAU OF THE SOUTHERN APPALACHIANS

by

Jennifer E. Buhay

Department of Integrative Biology

Doctor of Philosophy

Cave systems and their unique biota are widely viewed as highly endangered, yet very little is known about basic life history, ecology, distributions, habitat requirements, and evolutionary relationships of subterranean species. The crux of the problem in cave studies is the assumption that traditionally defined morpho-species represent distinct evolutionary lineages. Convergence is exhibited in the morphologies of many animal groups, vertebrate and invertebrate, which leads to confusion in diagnosing species' boundaries, geographic distributions, gene flow routes, and imperilment. This dissertation research includes phylogeographic analyses of freshwater cave-dwelling crayfishes in the Southern Appalachians, a global hotspot of subterranean biodiversity. By examining population structure in light of habitat, geology, geography, and hydrology, we can better provide conservation direction for these groundwater species.

Chapter one introduces a method, Nested Clade Phylogeographic Analysis (NCPA), used to investigate hypotheses about historical and current population structures within species. Using a statistically-testable framework, NCPA can elucidate historical speciation patterns and current routes of gene flow using genetic sequence data of thoroughly-sampled species. Using diverse examples, the chapter details the methodology of building haplotype networks, performing the geographic analyses, inferring past and contemporary evolutionary patterns and processes, and delineating species' boundaries.

Chapter two examines two competing hypotheses regarding conservation status of cave-dwelling species using a wide-ranging group of obligate subterranean crayfish species on the Cumberland Plateau's western escarpment. Using a population genetic approach, cave crayfish exhibited moderate to high levels of genetic diversity and attained large population sizes over their evolutionary histories. Phylogeographic analyses revealed that this crayfish assemblage originated along the northern end of the Cumberland Plateau and in leading-edge small steps, colonized southward and accumulated diversity along the way. Current species' boundaries do not match traditional morpho-species designations and also do not match current hydrological units.

Chapter three explores phylogeography and habitat differences within the facultative cave-dwelling crayfish species *Cambarus tenebrosus*. This freshwater species is unique in that it inhabits surface and subsurface karst environments, has an unusually large distribution, and exhibits troglomorphism with reduced eyes and elongated limbs. Using sequence data from over 100 sampled localities, mostly along the Cumberland Plateau, *C. tenebrosus* appears to have inhabited surface and subsurface biomes

throughout its evolutionary history. Additionally, this species shows extremely high levels of genetic diversity and NCA revealed significant phylogeographic structure within the species, but there was no significant relationship between habitat and genetic structure.

Chapter four examines the obligate cave crayfish assemblage, genus *Cambarus*, subgenus *Aviticambarus*, which ranges across the southernmost area of the Southern Appalachians, which is known to contain the highest species diversity of obligate terrestrial animals in the United States. The *Aviticambarus* assemblage is only currently known from 58 caves in Alabama and Tennessee, and with samples from half of the known sites, this study uncovered additional lineages previously obscured by convergent morphology. These species show low levels of genetic diversity and populations that do not appear to be expanding. Species' boundaries are supported by geologic and phylogeographic information, but not current drainage basin boundaries.

ACKNOWLEDGEMENTS

John Cooper and Tom Barr - I consider you my academic grandfathers and it is an honor to crawl, swim, climb, and every now and then walk in your footsteps. I hope that I answered some of the questions you also traveled miles underground trying to figure out. This project would not have been possible without the hoards of cavers who find, map, and explore caves every weekend. Gerald Moni, Thany Mann, and Kevin Toepke stuck with me over the years, though sometimes only as far as the cave entrance. That's if we didn't get lost trying to find the cave, get the truck stuck, or be run off by landowners - there was never a dull moment. Bob, Anne, Clinton, Matthew, Birdie, Pupper, Meow, Basil, and Slurp Elmore gave me a home away from Utah - thank you for everything.

To all the Crandall Lab visitors, collaborators, and students, you are an incredible bunch of people and I hope our paths cross again. Liz Sinclair, Loubna Tazi, Jenny Pramuk, Brice Noonan, James Finlay, and Rebecca Scholl - thank you for your friendship, time, support, and advice. Thus, you made this crazy time in my life enjoyable indeed! Henry Robison provided fun getaway trips to Arkansas complete with snakes, poison ivy, and ticks along with great Southern cooking, endless laughter, and driving around aimlessly looking for burrows. The BYU faculty and staff, and especially my committee members, were always helpful and supportive - thank you for letting me find my own way. Funding for my research was generously provided for by the National

Science Foundation, Cave Research Foundation, National Speleological Society, Cave Conservancy Foundation, United States Fish and Wildlife Service, Brigham Young University Graduate School and Integrative Biology Department.

Lastly, to my advisor Keith - thank you for always being in my corner (and just next door and across the ocean), for being a safety net when I was bouncing off the walls in my usual manner, and for being an example of a great mentor and advisor. You are stuck with me for good (that's not a threat, it's a promise!).

Do not go where the path may lead.

Go instead where there is no path and leave a trail.

Ralph Waldo Emerson

TABLE OF CONTENTS

CHAPTER 1

NESTED CLADE PHYLOGEOGRAPHIC ANALYSIS FOR CONSERVATION GENETICS*

INTRODUCTION

Genetic sequence data have become widely used in evaluating the unique relationship between geography and evolutionary history for conservation of species. Traditional methods, such as bifurcating trees and Wright's F statistics, often fall short in detailing past and contemporary events and contribute little intra-specific information (Posada & Crandall 2001; Pearse & Crandall 2004). Phylogenetic techniques, when applied in lower level systematic studies, show poor resolution, often resulting in polytomies and ambiguous connections (Crandall *et al.* 1994). This is particularly the case when species have recently diverged or have complicated metapopulation structure, in which case, bifurcating trees do not have the ability to accurately depict their evolutionary history (Posada & Crandall 2001). Despite this lack of resolution, broad geographic patterns can still be elucidated for older taxa using phylogenetic approaches. The field of phylogeography began by overlaying phylogenies onto geography and making broad inferences about evolutionary histories of species and populations (Avise

---

*This paper was accepted for publication: Buhay, J. E., K. A. Crandall, and D. Posada. 2006. Nested Clade Phylogeographic Analyses for Conservation Genetics. In: *Population Genetics for Animal Conservation.* Cambridge University Press.

1989). This approach, however, does not provide the opportunity to 1) statistically test the null hypothesis of no geographic association between populations, 2) test whether samples (number of individuals and collection localities) are sufficient, or 3) infer historical and contemporary processes and patterns that dictate current genetic variation (Carbone & Kohn 2004). However, approaches such as Nested Clade Analysis (NCA: Templeton *et al.* 1995), also known as Nested Clade Phylogeographic Analysis or NCPA (Templeton 2004), provide a statistical framework in which to test hypotheses about historical events and current population structure within species.

Indeed, conservation of a species is highly dependent on understanding the processes and the patterns that gave rise to the current phylogeographic composition of each unique taxon. The NCPA approach also has important applications to species delimitation and diagnosis, as it can be used to test for exchangeability and genealogical "exclusivity" (Crandall *et al.* 2000). In this chapter, we detail the methodology of the Nested Clade Phylogeographic Analysis of haplotype trees in phylogeographic studies and its application to a wide range of issues in conservation biology. Using examples from some published NCPA studies, we will discuss the method and its applications to conservation and to the study of population history within species.

NETWORK APPROACHES REQUIRE THOROUGH SAMPLING

There are particular cases where species are severely endangered and there are not enough populations or individuals to sample for an in-depth phylogeographic analysis. It is these species that are most in need of protection, yet it is very difficult to gather enough samples to detail biogeographical patterns and metapopulation dynamics for management

purposes. One such example is the U.S. federally threatened freshwater bivalve (*Potamilus inflatus*), which once ranged across the entire southeastern United States but is now limited to a few rivers, including the Black Warrior and Amite Rivers (Roe & Lydeard 1998). Due to the conservation status and rarity of these freshwater clams, thorough sampling (both numbers of individuals and sampling localities within the distribution) seems impossible, and therefore, direct comparison of sequence data coupled with a multi-species phylogeny was used to assess geographic variation. Twelve nucleotide sites of a 600 base pair portion of cytochrome oxidase I showed variation between the two rivers in sample of eight individuals. A phylogenetic tree revealed distinct differences between the two rivers as well as between other *Potamilus* species. Based on these results, the authors recommended that *P. inflatus* be recognized as two separate species rather than as two disjunct populations based on the presence of genetically-diagnosable characters and a 2% sequence divergence between the Amite form and the Black Warrior form of *P. inflatus*. Although the phylogeny and unique nucleotide differences were sufficient for species' diagnosis, there is still no information about the evolutionary history (i.e., dynamics within species) of the imperiled clams. Furthermore, Crandall *et al.* (2000) have argued that mere genetic distinctiveness at neutral genetic loci is not necessarily the sole criterion for diagnosing species or evolutionarily significant units (ESUs) for conservation. Additional information on the ecological exchangeability would be desirable to further substantiate the diagnosis of distinct species.

The highly endangered Tasmanian freshwater crayfish species *Astacopsis gouldi* also presented geographic sampling difficulties due to its endangered status. This species

was historically found throughout all drainages in northern Tasmania, but overhunting has lead to local extirpations and an imperiled conservation assessment. Sinclair *et al.* (2005) sampled several drainage basins across its range, including only a few individuals per site (less than ten) as permitted by authorities. Despite the restricted sampling, a haplotype network was still able to be constructed to help determine genetic structure across the rivers that currently harbor isolated populations of the species. A phylogenetic tree of the haplotypes was uninformative for evolutionary processes within the species because there were unresolved polytomies, but the network suggested extensive gene flow and migration of the crustacean species across many drainages. Despite the inability to conduct statistical tests for significant associations between sampled sites and genetic variation (as would be provided by NCPA), conservation management plans could effectively use the haplotype network information for reintroduction and augmentation efforts across various watersheds.

For cases where the species of focus is widespread and common, NCPA can be used to understand contemporary and historical evolutionary processes and patterns. It is critical that populations across the entire distribution are sampled. Phylogeographic approaches, particularly NCPA, are dependent on both geographic sampling and the numbers of individuals at each site. Thorough sampling allows researchers to detect historical events, such as range expansions and fragmentation, as well as contemporary processes, such as ongoing gene flow and isolation. "Thorough" sampling is becoming a contentious issue for metapopulation studies, particularly how to "best" sample a taxon for a phylogeographic study or for cases where species' taxonomic status is in question (i.e. species' complexes or hybrid zones). Indeed, the sampling effort should be as

homogeneous as possible, so for example, the errors in the allele frequencies estimates are similar across the sampled range.

Issues of genetic sampling include the choice of gene (how much variation within and between species), the numbers of genes sequenced (total number of base pairs), and mitochondrial versus nuclear gene regions. The gene of choice differs between taxonomic groups, but should be variable enough to detect differences at the population level for your study species. For example, the mitochondrial 16S gene is appropriate for studies of freshwater crayfish phylogeography (Buhay & Crandall 2005; Finlay *et al*. 2006) while the CO1 gene has been used for spiders (Paquin & Hedin 2004), and ND1 was used for toads (Masta *et al.* 2003). Typically, most studies use one gene region for NCPA and we recommend that other gene regions and analytical methods be used in support of the phylogeographic inferences provided by the NCPA method (see Carstens *et al.* 2004 for comparisons of analytical methods). Importantly, using several gene regions should greatly enhance the NCPA and provides cross-validation of the resulting inferences (Templeton 2002, 2004).

Issues of geographic sampling include the numbers of individuals sampled per locality and the numbers of sampled localities across the distribution of the species. Geographic sampling seems to be the most common question about the NCPA approach. Geographic sampling was recently addressed by Morando *et al.* (2003) in a phylogeographic study of a South American lizard species complex (*Liolaemus elongatus-kriegi*). They found that inadequate geographic sampling resulted in false patterns of regional genealogical exclusivity, and therefore recommended a sample size of five to ten individuals per site for as many sites as possible. Sampling density (number

of localities across the distributional range of the focal group) should be determined based on biologically-realistic dispersal ability of the species.

A recent example highlighting important issues with geographic and genetic sampling involved a meadow jumping mouse species *Zapus hudsonius* and the U.S. federally threatened subspecies *Z. h. preblei* contained within the taxon. King *et al.* (2006) sampled large numbers of individuals at few localities (348 individuals from 14 sites) and analyzed many loci (21 nuclear microsatellites, the mtDNA control region, and the mtDNA cytochrome b gene). Their conclusion was that the subspecies in question is a valid taxon, and is genetically distinct from neighboring subspecies. In contrast, Ramey *et al.* (2005) sampled extensively across the distribution of the mouse species, favoring more localities over large numbers of individuals. Ramey *et al.* (2005) gathered genetic data for the mitochondrial control region and five microsatellite regions, in addition to morphological measurements, from 195 individuals for over 80 localities. Their conclusion was that the subspecies in question is not a valid taxon because of evidence of recent gene flow with a neighboring subspecies. Nested Clade Phylogeographic Analysis was not conducted for either study, but it would have been possible if the datasets for the mitochondrial control region were combined. The use of NCPA would have been a beneficial and statistically-based approach for examining the taxonomic status of the mouse subspecies. The inference procedure for the Nested Clade Analysis asks explicitly "is the species present between the sampled localities?" and if the species IS present, then the inference would be "inadequate geographic sampling" for clades showing regional genealogical exclusivity due to poor sampling design. The researchers would then be provided with areas that need to be sampled (geographic gaps) by using the inference

procedure (Hedin & Wood 2002; Paquin & Hedin 2004). We provided this example of the contradictory mouse conclusions to illustrate that project design (and hence, gene sampling and locality sampling) along with subsequent adjustments to project design are critical in elucidating evolutionary history, contemporary processes, and species' boundaries for conservation.

HOW TO CONDUCT A NESTED CLADE PHYLOGEOGRAPHIC ANALYSIS

Network Construction

In theory, any phylogenetic reconstruction of the history of the sampled haplotypes can be used for the NCPA. However, as we have argued above, at the population level, network approaches are often more useful. There are many different ways to construct a haplotype network, including clustering, hierarchy, distance, and least squares methods (reviewed in Posada & Crandall 2001), but in the case of NCPA, statistical parsimony is most often employed (although any other method could be used as well). A recent study by Cassens *et al.* (2005) found that the minimum spanning networks constructed by the program Arlequin resulted in poor genealogical estimates, while parsimony and median-joining methods performed well, particularly in cases with extinct or unsampled interior haplotypes. The program TCS (Clement *et al.* 2000; freely available at http://darwin.uvigo.es) constructs haplotype networks using the method of statistical parsimony (Templeton *et al.* 1992). The input format is a simple nexus file with aligned DNA sequences from every individual. Sequences of closely-related outgroups should be included in the input file to root the network. The output is the genealogical network depicting number of mutational steps between haplotypes.

Network Diagrams Illustrate Different Types of Information

The phylogeography of obligate cave crayfish in the genus *Orconectes* was examined using 485 base pairs of sequence data from the mitochondrial 16S gene (Buhay & Crandall 2005). These sequences were used to construct a statistical parsimony network (Figure 1) resulting in 69 unique haplotypes identified from 421 individuals sampled at 67 cave localities, thoroughly covering the entire distribution along the western escarpment of the Cumberland Plateau in the Southern Appalachians. This network shows the mutational steps between each haplotype (haplotypes are represented as circles with different numbers), including missing haplotypes (marked as small empty circles). Missing haplotypes may be extinct or unsampled haplotypes. A 95% confidence level is first calculated to decide whether we should connect two haplotypes. The 95% confidence level is the maximum number of mutational steps between two haplotypes under which we are 95% sure that no multiple mutations at the same site (overimposed changes) have occurred. The idea is that because we cannot see overimposed changes, we do not want to make those connections in which we can easily underestimate the actual number of differences between two haplotypes. The 95% confidence level for this network is nine steps, which means that there must be less than nine mutational differences for the method to directly connect two haplotypes. If the number of mutational steps between sampled haplotypes is greater than the 95% confidence level, multiple separate networks will result. In Figure 1, this was the case with haplotypes 1 through 16 (*O. australis packardi*), which form a distinct network separated from the rest of the haplotypes (17 through 69 which included *O. incomptus*, *O. sp. nov 1*, and *O. a. australis*) by ten steps. In each network, the putative ancestral

haplotype is the one with the highest outgroup probability (Castelloe & Templeton 1994) and is depicted as a rectangle, while the other haplotypes are drawn as circles. The *O. a. packardi* haplotype determined to be ancestral was 7, while the ancestral haplotype for the other species was 27 and they are both depicted as rectangles. Frequency (number of individuals) is indicated by the size of the circles. The outgroup probability for each haplotype are also provided by the TCS program and can be found by clicking on the haplotype in the network. The haplotypes with the greatest numbers of individuals in Figure 1 are haplotypes 7, 27, 40, 51, 61, and 65, which are represented by the largest symbols.

A rough estimation of the relative age of haplotypes is determined by both the frequency of the haplotypes in the sample (which is why it is important to include all the sequence data in the analysis and not just the unique sequences) and the number of connections (Castelloe & Templeton 1994). Neutral coalescent theory suggests that high frequency haplotypes are usually older than low frequency haplotypes and are typically found in more internal locations in the network. Rarer haplotypes are thought to have arisen recently, occupy tips, and often have fewer connections to other haplotypes (Crandall & Templeton 1993).

How to Resolve Loops and Ambiguous Connections

When there is homoplasy (due to parallel changes, reversals, or recombination) in the data, some haplotypes may be connected to several other haplotypes forming loops or reticulations, resulting in unresolved networks. Using predictions from coalescent theory and information about the sampling, loops can be broken to facilitate nesting through the

higher nesting levels, although rules exist for nesting with very simple ambiguous connections (Templeton & Sing 1993). Three different criteria can be used to resolve loops: frequency, network location, and geography (Crandall & Templeton 1993). First, haplotypes are most likely connected to higher frequency haplotypes, rather than to haplotypes representing a single individual. Second, haplotypes are most likely connected to interior haplotypes than to haplotypes on the tips of the network. And third, haplotypes are most likely connected to haplotypes from the same geographic area than to haplotypes found in distant areas.

An example of an unresolved network and how to resolve the connections can be found in Pfenninger & Posada (2002). In this study, 16S sequences were sampled in 204 land snails (*Candidula unifasciata)* from 37 localities. The initial network included 46 haplotypes and four major loops (Figure 2). Three of the loops (marked 1, 2, and 3) included connections between singletons (haplotypes represented by only a single sequence in the sample) and therefore were broken based on the frequency criterion (see arrows). The fourth loop was more complex (labeled 4A through 4D in Figure 2). The authors in fact explored all solutions of loop 4, and despite the differences in the resulting nesting designs, the NCAP inferences were essentially identical.

Building the Nesting Design

Once the haplotype network is resolved, the next step in the NCPA is to build the nesting design. Beginning at the tips, clades will include haplotypes connected by one mutational step while working toward the interior of the network (Templeton & Sing 1993; Crandall 1996). Using the example of Pfenninger & Posada (2002) (with loop 4 cut at 4C), haplotype 44 and 25 are joined into clade 1-5 (Figure 3). Missing haplotypes,

represented by open circles, are considered when making the nesting decisions in the same terms as sampled haplotypes. Haplotype 19 is connected to a missing haplotype to build clade 1-8. Two missing haplotypes can be grouped together in a clade. All haplotypes, sampled or unsampled, must be grouped. There will be cases where more than one haplotype is connected to another haplotype by one step. Haplotypes 9, 32, and 39 are joined into clade 1-7 and haplotypes 1, 4, 6, 7, 12, and 33 are joined to form clade 1-14. Each of the nesting groups is called a step clade. Haplotypes are the 0-step clades. The next step is to hierarchically join the 1-step clades into 2-step clades based on one mutational step and so on, using the same rules, until the entire network is grouped at the highest level, working from the tips to the interior of the network. Clade 1-1 and Clade 1-2 are joined to form Clade 2-1 (Figure 3). The nesting process is completed when all clades are nested together at the highest nesting level, which is the total cladogram. Clade 4-1 and Clade 4-2 comprise the total cladogram in Figure 3.

From Haplotype Network to Geographical Analysis Using GeoDis

GeoDis is a program that statistically tests the associations between the genetic and geographical distances (Posada *et al.* 2000). GeoDis is freely available at http://darwin.uvigo.es and can be run on a PC or Mac platform. The input file is a written description of the nested cladogram and the corresponding geographic information. The process of writing the input file can be exceptionally tedious for large datasets, and must be done carefully to prevent mistakes.

The first step in constructing this input file for GeoDis is to make a list of the haplotypes obtained from the TCS network (Figure 4) and all the sampled individuals

that are represented by each haplotype. Using a simple example of five haplotypes and 18 salamanders for illustration, Haplotype 1 = 4 individuals (2 from Arkansas, 2 from Missouri); Haplotype 2 = 7 individuals (1 from Illinois, 2 from Iowa, and 4 from Michigan); Haplotype 3 = 1 individual from Iowa, Haplotype 4 = 3 individuals from Alabama, and Haplotype 5 = 3 individuals (1 from Georgia, 2 from Tennessee).

The next step is to make a geographic description of the sampled localities. If the studied organism can move between localities through the shortest possible path (i.e., a straight line), this is done by specifying latitude-longitude coordinates (in Degree Minute Second: DMS or decimal degrees: DD format). Pairwise distances (km) in table format can also be specified by the user, particularly for cases where species are limited by habitat barriers, such as in the case of aquatic species and linear river systems (see discussion below on Fetzner & Crandall 2003).

A sample input file using the example salamander dataset above is given in Table 1. The input file only includes clades that have BOTH genetic and geographic variation, such as in Clade 1-2 on Figure 4. Clade 1-2 includes two haplotypes (haplotype 4 and 5) AND includes individuals from three locales: Alabama, Georgia, and Tennessee. Table 1 provides the line by line input on the left, with the explanation on the right for clarification. Out of the possible eleven nested clades (four 1-step clades, four 2-step clades, two 3-step clades, and 1 total cladogram), only four clades (Clades 1-2, 3-1, 3-2, and Total) contained both genetic and geographic variation for analyses. Once the program GeoDis is opened, input the data file, choose what format your geographic information is in (DMS or DD or pairwise table) and select "run." The program will output the statistical relationships between the genetic and geographic distances for each

clade based on the number of permutations chosen (default = 1000 resamples for a 5% level of significance). For each clade, the observed chi-square is given along with its probability of being observed under the null hypothesis of no association between geography and genetic variation (Table 2).

For each clade, the NCPA statistics are reported as 'within clade distance' (D_c) and 'nested clade distance' (D_n) (Table 3). When both interior and tip subclades exist, there is also a test for interior vs. tip clades, reported as 'I-T' distance. The D_c is calculated as the average distance of the individuals from the geographical center of the clade. The nested-clade distance D_n is calculated as the average distance of the clade individuals from the next higher-level clade's geographical center. Significantly small (reported as 'S') and large (reported as 'L') deviations at the 0.05 confidence level are key measures for making inferences with the key of Templeton (2005) found at http://darwin.uvigo.es/software/geodis.html.

Using the Inference Key to Uncover Evolutionary Processes and Patterns

The geographic results from GeoDis are often presented by researchers as a flowchart between step levels, as shown by Table 3 from Tarjuelo *et al.* (2004). Once the NCPA results are organized, the next step is to examine the results of each clade with significant genetic-geographic variation (significantly small and significantly large values) using a dichotomous inference key provided by Templeton (2005). This key is primarily used to translate the statistical output of GeoDis into biological inferences. Some of the inferences include "restricted gene flow / isolation by distance," “contiguous range expansion,” “allopatric fragmentation,” and “long distance colonization.” The key

is only used for the clades that show significantly large or small values for D_c, D_n, or I-T. When there are no significant distances within a clade, the null hypothesis of no geographical association of haplotypes cannot be rejected (Templeton 1998, 2001). The chain of inference is also usually reported for each outcome, such as the inference chain (1-2-3-4-9-NO) for the Total Cladogram in Table 3.

NCPA also provides information about evolutionary time, with lower level nesting processes occurring more recently than the processes that are significant at higher nesting levels. Based on the inferences of the older clades, researchers can gather information about past geographic and environmental events, such as the effects of Pleistocene glaciations or the uplift of mountain ranges on distribution patterns (Templeton 1998). Using younger clade groupings, we can elucidate the impacts of human activities on species' ranges, or show recent expansions by invasive species.

The NCPA inference key has recently come under debate because of the "subjective" interpretations that are made with respect to the data being examined (Knowles & Maddison 2002). Although the NCPA approach is indeed a statistical framework, the inference key of processes and patterns is largely flexible, lacking a "standard" method to test the inferences themselves. Knowles & Maddison (2002) argue that the inferences made are outside the realm of confidence, meaning that there is no way to test statistically between inferred events of long distance colonization, isolation by distance, migration, or past fragmentation. Indeed, model-based approaches (and model selection) for phylogeography are desirable, but if one has no a priori hypotheses to test, there are infinite models (through time and space) that one could use. Obviously, the NCPA approach is not a stand-alone answer to questions of evolutionary history of a

species, and although NCPA provides many insights and generates hypotheses when none exist, the use of other analytical methods implemented by programs such as Mesquite (Maddison & Maddison 2004), IM (Hey & Nielsen 2004), Arlequin (Schneider *et al.* 2000), and Mismatch distributions (Rogers & Harpending 1992), can be a tremendous asset in complementing and validating the inferences from the NCPA (Carstens *et al.* 2004). By incorporating results from multiple sources, a stronger case can be made for the phylogeographic patterns of the species that were elucidated by the NCPA.

A Special Case: Terrestrial Versus Riparian Species

As was previously stated, straight-line distances may not be always the best method of representing physical distances between populations. Aquatic species are restricted to the current paths of the waterways they inhabit, whereas, terrestrial species are not limited to a linear habitat. Therefore, geographic distances reflected by latitude and longitude coordinates may not be appropriate for riparian species (Figure 5). This idea was empirically tested using mitochondrial data from the 16S gene of a widespread freshwater crayfish species in the Ozarks (Fetzner & Crandall 2003). *Orconectes luteus* were collected from 35 stream sites mostly across Missouri. Geographic coordinates were recorded for comparison to river distances, which were measured using topographic maps. A distance matrix (in kilometers) between pairwise comparions for every site was used as input for GeoDis, and a separate input file was assembled using latitude-longitude coordinates for the collection sites. The objective of the study was to compare the two distance methods and their outcomes in inferring phylogeographic structure. It should be

noted that the NCPA statistics are not exactly the same in both cases, and differences are expected because of the variation in calculated geographic distances between locales and their genetic diversity. Results from the NCPA showed distinct differences between the significant clades, often inferring different processes from the two methods, particularly for lower nested clades (Table 4). Results from Fetzner & Crandall (2003), highlight the importance of using the appropriate (biologically-relevant) geographical distances when implementing the NCPA.

Inferring Biogeographic Patterns

Regional biogeographic patterns can be eluciduated by examining the population structure of species, and the inferences provided by NCPA reflect both historical and contemporary patterns of genetic variation. In the landsnail example (Pfenninger & Posada 2002), isolation by distance with long range dispersal was the inferred pattern for clade 4-2 (Figure 6). Northward contiguous range expansion was inferred for clades nested within clade 4-1, and included areas of secondary contact with clade 4-2. Because these inferred patterns relate to the highest nesting levels (which are the oldest groupings), they are possibly responses to historical environmental changes, such as glaciation events. Many phylogeographic studies have investigated responses of species to glacial advance and retreat cycles (Cooper *et al.* 1995; Comes & Abbott 1998; Turgeon & Bernatchez 2001; Branco *et al.* 2002; Hoffman & Blouin 2004). No other phylogeographic method incorporates both temporal and spatial structure in a statistically-testable framework, and many of these aforementioned studies also validated the inferences using other metapopulation analyses.

NCPA CAN BE USED TO DELIMIT SPECIES' BOUNDARIES

A contentious issue in conservation biology is the diagnosis of species and the methods employed to delimit species' boundaries (Sites & Crandall 1997; Sites & Marshall 2003, 2004; Agapow *et al.* 2004). There seems to be little agreement on the definition of "species" even though "species" are deemed by many to be the fundamental units of conservation biology. Species are certainly important entities regardless of one's concept, but conservation biology is also deeply concerned with intraspecific variation within and among populations (Sites & Crandall 1997; Crandall *et al.* 2000). It is at the population level that evolutionary forces operate to drive speciation processes, and in this regard, it is necessary to simultaneously recognize the importance of species and populations for conservation measures. NCPA can be used to diagnose species under the Genealogical Concordance Species Concept (Avise & Ball 1990) and the Cohesion Species Concept (Templeton 1989). Both approaches use concordance as criteria for species delimitation as well as "exclusivity".

In the case of genealogical concordance, multiple types of markers (such as mitochondrial and nuclear DNA, morphological, ecological, behavioral, habitat, etc.) must show consistent patterns. There is, however, no set "level" of concordance. In other words, the number of markers and degree of concordance among the markers is subjective (Hudson & Coyne 2002). This concept is largely based on coalescent theory (Hudson 1990), and that concordance is the result of long evolutionary separation, which will be reflected in the concordant gene genealogies (Baum & Shaw 1995). A *genealogical species* is a group of organisms whose members are more closely related to each other ("exclusivity") than to any other organisms outside the group (Baum & Shaw

1995). The boundaries of genealogical species can be defined using the testable null hypotheses of Templeton (1989) in the NCPA framework. The first null hypothesis is that the sampled group represents a single evolutionary lineage. If the first null hypothesis is rejected, for example through the inference of fragmentation, then a second null hypothesis is tested. The second null hypothesis is that there is no significant difference across lineages with respect to genetic and/or demographic adaptations. Genealogical concordant species are recognized after both null hypotheses are rejected.

The cohesion species concept is largely based on the ability to rigorously test a set of null hypotheses concerned with the association of geography and genotype/phenotypes (Templeton *et al.* 1995). The first null hypothesis is the same listed above for genealogical species diagnosis. If the organisms represent multiple lineages with 95% confidence, then the first null hypothesis is rejected. The second null hypothesis is that populations of different lineages are genetically and/or ecologically interchangeable among each other. The second null hypothesis is rejected when there is a significant association between geography and genetic and/or ecological variables, determined by NCPA. The ecological basis is that individuals can be "exchanged" or moved between populations because they occupy the same niche (see Rader *et al.* 2005 for a variety of approaches to test ecological exchangeability and see Finlay *et al.* 2006 for an application of the ecological exchangeability approach). The genetic basis is that individuals are "exchangeable" if there is extensive gene flow among populations.

Nested Clade Phylogeographic Analysis was recently applied to delimit species boundaries of a South American lizard complex *(Liolaemus elongatus-kriegi*; Morando *et al.* 2003) using the combined approach of Wiens & Penkrot (2002) and Templeton

(2001). The Wiens-Penkrot protocol complements species' delimitation studies that combine haplotype phylogenies and NCPA, but does not require "exclusivity." Morando *et al.* (2003) followed the protocol but with modifications: 1. multiple gene regions were used to test for genealogical concordance and 2. exclusivity was a criterion for species' boundaries. Because the authors did not have ecological data to address exchangeability, they only tested the first null hypothesis of Templeton's cohesion species criterion, with independent lineages arbitrarily defined as those outside the 95% confidence limit determined from their most variable mitochondrial gene region. The combined approach, using a priori defined criteria, supported the same clades and identified many more independent lineages than previously recognized under existing taxonomic names. Lineages that were supported by multiple lines of evidence were interpreted as "candidate species" because they met the criteria for genetic concordance, geographic concordance, exclusivity, and / or fragmentation / isolation by distance determined by NCPA.

SUMMARY

Nested Clade Phylogeographic Analysis provides a statistical framework to elucidate historical and contemporary evolutionary processes that have contributed to the present-day genetic variation of a species. Some practical applications of NCPA include inferences about species' responses to past environmental events, current routes of gene flow and expansion, and the diagnosis of species under the Cohesion Species Concept and Genealogical Concordance Species Concept. It is a powerful tool for understanding population-level and species-level patterns of variability both temporally and spatially.

Indeed, one of the primary goals of conservation biology is the protection of the evolutionary forces that naturally drive speciation and biodiversity.

ACKNOWLEDGEMENTS

We thank the editors for inviting us to participate in this publication and for their patience. Two anonymous reviewers and Jack Sites provided valuable comments which greatly improved this manuscript. JEB was supported by a graduate fellowship from the Graduate Studies Office at Brigham Young University and an NSF Dissertation Improvement Grant (DEB 0508580).

REFERENCES

Agapow, P.-M., Bininda-Emonds, O. R. P., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C. & Purvis, A. 2004 The impact of species concept on biodiversity studies. *Quarterly Review of Biology* **79**, 161-79.

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

Avise, J. C. & Ball, R. M. Jr 1990 Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* **7**, 45-67.

Baum, D. A. & Shaw, K. L. 1995 Genealogical perspectives on the species problem. In *Experimental and Molecular Approaches to Plant Biosystematics* (eds. P. C. Hoch & A. G. Stephenson), pp. 289-303. St. Louis Missouri: Missouri Botanical Gardens.

Branco, M., Monnerot, M., Ferrand, N. & Templeton, A.R. 2002 Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analysis of mitochondrial DNA genetic variation. *Evolution* **56**, 792–803.

Buhay, J. E., & Crandall, K. A. 2005 Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. *Molecular Ecology* **14**, 4259-73.

Carbone, I. & Kohn, L. 2004 Inferring process from patterns in fungal population genetics. *Applied Mycology and Biotechnology* **4**, 1-30.

Carstens, B. C., Stevenson, A. L., Degenhardt, J. D. & Sullivan, J. 2004 Tested nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology* **53**, 781-92.

Cassens, I., Mardulyn, P. & Milinkovitch, M. C. 2005 Evaluating intraspecific network construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? *Systematic Biology* **54**, 363-72.

Castelloe, J. & Templeton, A. R. 1994 Root probabilities for intra-specific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution* **3**, 102-13.

Clement, M., Posada, D., & Crandall, K. A. 2000 TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-59.

Comes, H. P. & Abbott, R. J. 1998 The relative importance of historical events and gene flow on the population structure of a Mediterranean ragwort,

Senecio gallicus (Asteraceae). *Evolution* **52**, 355-67.

Cooper, S. J. B., Ibrahim, K. M. & Hewitt, G. M. 1995 Postglacial expansion and genome subdivision in the European grasshopper *Chorthippus parallelus*. *Molecular Ecology* **4**, 49-60.

Crandall, K. A. 1996 Multiple interspecies transmissions of human and simian T-cell leukemia/lymphoma virus type I sequences. *Molecular Biology and Evolution* **13**, 115-31.

Crandall, K. A. & Templeton, A. R. 1993 Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**, 959-69.

Crandall, K. A., Templeton, A. R. & Sing, C. F. 1994 Intraspecific phylogenetics: problems and solutions. In: *Models in Phylogeny Reconstruction* (ed. R. W. Scotland, D. J. Siebert & D. M. Williams), pp. 273-97. Oxford: Claredon Press.

Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M. & Wayne, R. K. 2000 Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **15**, 290-5.

Fetzner, J. W. Jr & Crandall, K. A. 2003 Linear habitats and the nested clade analysis: an empirical evaluation of geographic versus river distances using an Ozark crayfish (Decapoda: Cambaridae). *Evolution* **57**, 2101-18.

Finlay, J. B., Buhay, J. E. & Crandall, K. A. 2006 Surface to subsurface freshwater connections: phylogeographic and habitat analyses of

Cambarus tenebrosus, a facultative cave-dwelling crayfish. *Animal Conservation*, in press.

Hedin, M. C. & Wood, D. 2002 Genealogical exclusivity in geographically proximate populations of *Hypochilus thorelli* Marx (Araneae, Hypochilidae) on the Cumberland Plateau of eastern North America. *Molecular Ecology* **11**, 1975-88.

Hey, J. & Nielsen, R. 2004 Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-60.

Hoffman, E. A. & Blouin, M. S. 2004 Evolutionary history of the Northern Leopard Frog: reconstruction of phylogeny, phylogeography, and historical changes in population demography from mitochondrial DNA. *Evolution* **58**, 145-59.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1-44.

Hudson, R. R. & Coyne, J. A. 2002 Mathematical consequences of the genealogical species concept. *Evolution* **56**, 1557-65.

King, T.L., Switzer, J. F., Morrison, C. L., Eackles, M. S., Young, C. C., Lubinski, B. A. & Cryan, P. 2006 Comprehensive genetic analyses reveal evolutionary distinction of a mouse (*Zapus hudsonius preblei*) proposed for delisting from the U.S. Endangered Species Act. *Molecular Ecology*, in press.

Knowles, L. L. & Maddison, W. P. 2002 Statistical phylogeography. *Molecular*

Ecology **11**, 2623-35.

Maddison, W. P. & Maddison, D. R. 2004 Mesquite: a modular system for evolutionary analysis. Version 1.05 http://mesquiteproject.org

Masta, S. E., Laurent, N. M. & Routman, E. J. 2003 Population genetic structure of the toad *Bufo woodhousii*: an empirical assessment of the effects of haplotype extinction on nested cladistic analysis. *Molecular Ecology* **12**, 1541-54.

Morando, M., Avila, L. J. & Sites, J. W. Jr 2003 Sampling strategies for delimiting species: genes, individuals, and populations in the *Liolaemus elongatus-kriegi* complex (Squamata: Liolaemidae) in Andean-Patagonian South America. *Systematic Biology* **52**, 159-85.

Paquin, P. & Hedin, M. 2004 The power and perils of 'molecular taxonomy': a case study of eyeless and endangered *Cicurina* (Araneae: Dictynidae) from Texas caves. *Molecular Ecology* **13**, 3239-55.

Pearse, D. E. & Crandall, K. A. 2004 Beyond Fst: Analysis of population genetic data for conservation. *Conservation Genetics* **5**, 585-602.

Pfenninger, M. & Posada, D. 2002 Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): fragmentation, corridor migration, and secondary contact. *Evolution* **56**, 1776-88.

Posada, D. & Crandall, K. A. 2001 Intraspecific gene genealogies: trees grafting into network. *Trends in Ecology and Evolution* **16**, 37-45.

Posada, D., Crandall, K. A. & Templeton, A. R. 2000 GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic

haplotypes. *Molecular Ecology* **9**, 487-88.

Rader, R. B., Belk, M. C., Shiozawa, D. K. & Crandall, K. A. 2005 Empirical tests for ecological exchangeability. *Animal Conservation* **8**, 1-9.

Ramey, R. R., Liu, H. P., Epps, C. W., Carpenter, L. M. & Wehausen, J. D. 2005 Genetic relatedness of the Preble's meadow jumping mouse (*Zapus hudsonius preblei*) to nearby subspecies of *Z. hudsonius* as inferred from variation in cranial morphology, mitochondrial DNA, and microsatellite DNA: implications for taxonomy and conservation. *Animal Conservation* **8**, 329-46.

Roe, K. J. & Lydeard, C. 1998 Species delineation and the identification of evolutionarily significant units: lessons from the freshwater mussel genus *Potamilus* (Bivalvia: Unionidae). *Journal of Shellfish Research* **17**, 1359-63.

Rogers, A. R. & Harpending, H. 1992 Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-69.

Schneider, S. D., Roessli, D. & Excoffier, L. 2000 Arlequin version 2.0: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sinclair, E. A., Losada, M. P. & Crandall, K. A. 2005 Molecular phylogenetics for conservation biology. In *Phylogenies and Conservation* (ed. A. Purvis, T. Brooks & J. Gittleman), pp. 19-56. Cambridge: Cambridge University Press.

Sites, J. W. Jr & Crandall, K. A. 1997 Testing species boundaries in

biodiversity studies. *Conservation Biology* **11**, 1289-97.

Sites, J. W. Jr & Marshall, J. C. 2003 Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution* **18**, 462-70.

Sites, J. W. Jr & Marshall, J.C. 2004 Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics* **35**, 199-227.

Tarjuelo, I., Posada, D., Crandall, K. A., Pascuel, M. & Turon, X. 2004 Phylogeography and speciation of colour morphs in the colonial ascidian *Pseudodistoma crucigaster*. *Molecular Ecology* **13**, 3125-36.

Templeton, A. R. 1989 The meaning of species and speciation. In: *Speciation and Its Consequences* (ed. D. Otte & J. A. Endler), pp. 3-27. Sunderland, MA: Sinauer Associates.

Templeton, A. R. 1998 Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**, 381-97.

Templeton, A. R. 2001 Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology* **10**, 779-91.

Templeton, A. R. 2002 Out of Africa again and again. *Nature* **416**, 45-51.

Templeton, A. R. 2004 Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**, 789-809.

Templeton, A. R. & Sing, C. F. 1993 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analysis with cladogram uncertainty and

recombination. *Genetics* **134**, 659-69.

Templeton, A. R., Crandall, K. A. & Sing, C. F. 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. III. Cladogram estimation. *Genetics* **132**, 619-33.

Templeton, A. R., Routman, E. & Phillips, C. A. 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**, 767-82.

Turgeon, J. & Bernatchez, L. 2001 Mitochondrial DNA phylogeography of lake cisco (*Coregonus artedi*): evidence supporting extensive secondary contacts between two glacial areas. *Molecular Ecology* **10**, 987-1001.

Wiens, J. J. & Penkrot, T.A. 2002 Delimiting species using DNA and morphological variation and discordant species limits in Spiny Lizards (*Sceloporus*). *Systematic Biology* **51**, 69-91.

Table 1. Sample input file for GeoDis on the left with the line by line explanation of the input shown on the right. Data corresponds to Figure 4.

Salamander mtDNA CO1	project name
8	number of sampled localities
1 Missouri	first locality
2 38 25 33 N 93 17 08 W	# individuals and geog. coords.
2 Michigan	second locality
4 42 16 00 N 83 38 36 W	# individuals and geog. coords.
3 Ilinois	third locality
1 37 58 36 N 88 30 11 W	# individuals and geog. coords.
4 Iowa	fourth locality
3 43 03 04 N 91 46 54 W	# individuals and geog. coords.
5 Arkansas	fifth locality
2 36 24 12 N 92 58 30 W	# individuals and geog. coords.
6 Georgia	sixth locality
1 32 31 12 N 84 38 56 W	# individuals and geog. coords.
7 Alabama	seventh locality
3 34 00 53 N 85 52 27 W	# individuals and geog. coords.
8 Tennessee	eighth locality
2 36 12 32 N 85 30 23 W	# individuals and geog. coords.
4	# nested clades w/ genetic & geog. info
Clade 1-2	name of first nested clade
2	# of haps in clade 1-2
4 5	name of haps in clade 1-2
1 1	position of haps: both are tips (1)
3	# of localities within clade 1-2
7 6 8	locality #s from list (AL, GA, TN)
3 0 0	Hap 4 = 3 AL, 0 GA, 0 TN
0 1 2	Hap 5 = 0 AL, 1 GA, 2 TN
Clade 3-1	name of second nested clade
2	# of 2-step subclades in clade 3-1
1-1 1-2	names of subclades in clade 3-1
0 1	position: 1-1 is interior (0), 1-2 = tip
5	# of localities within 3-1
5 1 7 6 8	locality #s (Ark, MO, AL, GA, TN)
2 2 0 0 0	2 Ark, 2 MO, 0 AL, 0 GA, 0 TN in 1-1
0 0 3 1 2	0 Ark, 0 MO, 3 AL, 1 GA, 2 TN in 1-2
Clade 3-2	name of third nested clade
2	# of 2-step subclades in clade 3-2
1-3 1-4	names of subclades in clade 3-2
0 1	position: 1-3 is interior (0), 1-4 = tip
3	# of localities within 3-2
3 4 2	localities #s (IL, Iowa, Mich)
1 2 4	1 IL, 2 Iowa, 4 Mich in 1-3
0 1 0	0 IL, 1 Iowa, 0 Mich in 1-4
Total Cladogram	Final grouping of highest clades
2	# of 3-step clades within Total
3-1 3-2	names of subclades within Total
1 1	position: both are tips
8	# of localities in Total
5 1 7 6 8 3 4 2	Ark, MO, AL, GA, TN, IL, Iowa, Mich
2 2 3 1 5 0 0 0	
0 0 0 0 0 1 3 4	
END	

Table 2. Nested contingency results based on 9999 permutations for clades with genetic and geographic associations. Probability (P) is the probability of obtaining a chi-square statistic larger than or equal to the observed statistic. Clades with P values less than 0.05 suggest significant geographic structure (Pfenninger & Posada 2002).

Clades nested with	Permutational χ^2 statistic	*P*
Clade 1-1	7.00	0.046
Clade 1-14	66.93	0.044
Clade 1-17	459.69	0.000
Clade 2-1	7.47	0.024
Clade 2-9	64.65	0.020
Clade 3-1	17.00	0.016
Clade 3-2	21.43	0.000
Clade 3-4	22.42	0.022
Clade 3-5	124.90	0.044
Clade 4-1	97.99	0.000
Clade 4-2	83.70	0.000
Entire cladogram	187.65	0.000

Table 3. Results of the nested geographical analysis for *Pseudodistoma crucigaster*. Column Name is the name of the clade, *Dc* is the clade distance and *Dn* is the nested clade distance at each one of the levels of the analysis (haplotype, one-step, and two-step levels). The row *I-T* indicates the average difference between interior and tip clades. Superscript S means that the statistic was significantly small and superscript L that the statistic was significantly large (both at the 5% level). The lines in bold describe the steps followed in the inference key and the conclusion reached by this method: NS (not significant), Past Frag (past fragmentation), RE (range expansion), RGF (restricted gene flow) and CRE (contiguous range expansion) (Tarjuelo *et al.* 2004).

Haplotype			**1-step clade**			**2-step clade**		
Name	***Dc***	***Dn***	***Name***	***Dc***	***Dn***	***Name***	***Dc***	***Dn***
IV	183.15	176.46						
XI	0	102.31	**NS**			**(Yellow-Grey Morphotypes)**		
I-T	183.15	74.14	1-1	176.57	201.04^{L}			
X	0	0	1-2	0	181.51			
V	0	0						
VI	0	0						
VII	0	0						
VIII	0	0						
IX	0	0	1-3	0^{S}	176.44	**1-2-3-4-9-No: Past. Frag.**		
			I-T	176.57^{L}	24.21	2-1	191.75^{L}	172.26^{L}
I	80.24^{S}	135.56^{L}				**(Orange Morphotype)**		
II	22.90^{S}	142.97^{L}	**1-2-3-5-6-'too few clades' RE/RGF**					
I-T	57.33^{L}	-7.408^{S}	1-4	140.08^{S}	139.44			
III	0	0	1-5	0	130.25	**1-2-11-12-No: CRE**		
			I-T	140.08	9.19^{S}	2-2	140.78^{S}	149.47^{S}

Total cladogram	**1-2-3-4-9-No: Past Frag**

Table 4. Comparison of inferences drawn from the geographic and linear river distance methods for geographically significant clades. At lower nesting levels, the use of linear river distances made a drastic difference in the inferences made about contemporary patterns (Fetzner & Crandall 2003).

Clade	Geographic distances	River distances
1-1	restricted gene flow with isolation by distance (rgf/ibd)	past fragmentation (pf)
1-18	panmixia	inconclusive outcome
2-1	panmixia	restricted gene flow with isolation by distance (rgf/ibd)
2-5	restricted gene flow with some long-distance dispersal (rgf/ldd)	past fragmentation (pf)
2-6	past fragmentation (pf)	panmixia
2-8	contiguous range expansion (cre)	panmixia
2-13	restricted gene flow with isolation by distance (rgf/ibd)	restricted gene flow/dispersal with some long-distance dispersal (rgf/dispersal ldd)
3-1	restricted gene flow with isolation by distance (rgf/ibd)	long-distance colonization (ldc)
3-2	past fragmentation (pf)	past fragmentation (pf)
3-4	past fragmentation (pf)	more sampling needed
4-1	past fragmentation (pf)	past fragmentation (pf)
4-2	past fragmentation (pf)	past fragmentation (pf)
4-3	allopatric fragmentation (af)	allopatric fragmentation (af)
5-1	long-distance colonization (ldc)	long-distance colonization (ldc)

Figure 1. *O. a. packardi* (haplotypes 1-16) was outside the 95% confidence limit (nine steps) while *O. incomptus* (haplotypes 17-20), *O. sp. nov*. (haplotypes 21-25), and *O. a.australis* (haplotypes 26-69) were connected within the 95% confidence level. Empty circles in the network represent unsampled, possibly extinct haplotypes. The outgroups *Cambarus gentryi* and *C. graysoni* were outside the 95% limit and connected to haplotype 2 of *O. a. packardi.* (Buhay & Crandall 2005).

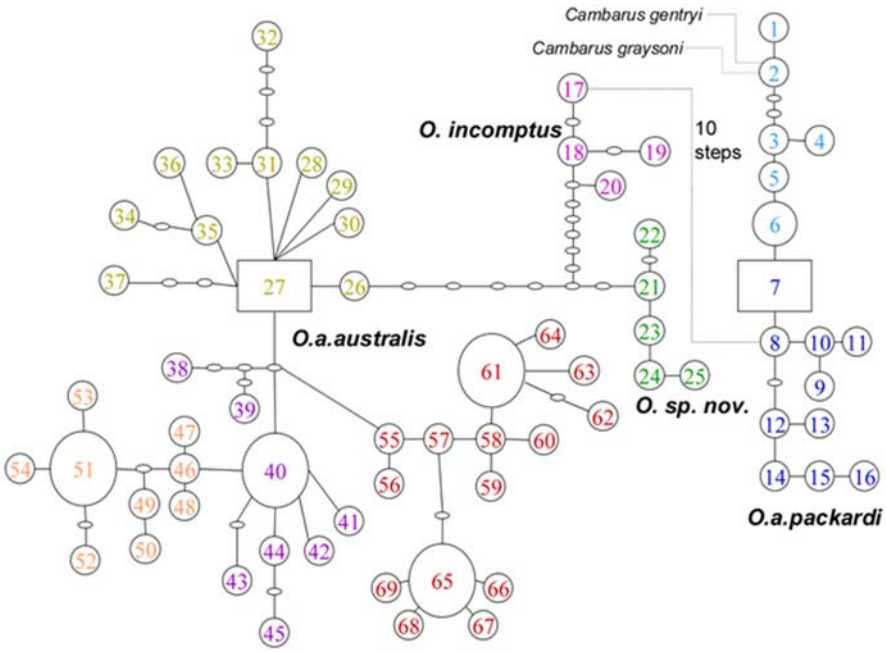


Figure 2. Unresolved network constructed using TCS showing different break possibilities (marked by numbers and arrows) for resolution. Numbered arrows 1, 2, and 3 were broken between singletons, while loop 4 had four different break options (4A through 4D) but was broken at arrow 4C (along with the unsampled haplotype between 4A and haplotype 11) based on geographical criteria (Pfenninger & Posada 2002).

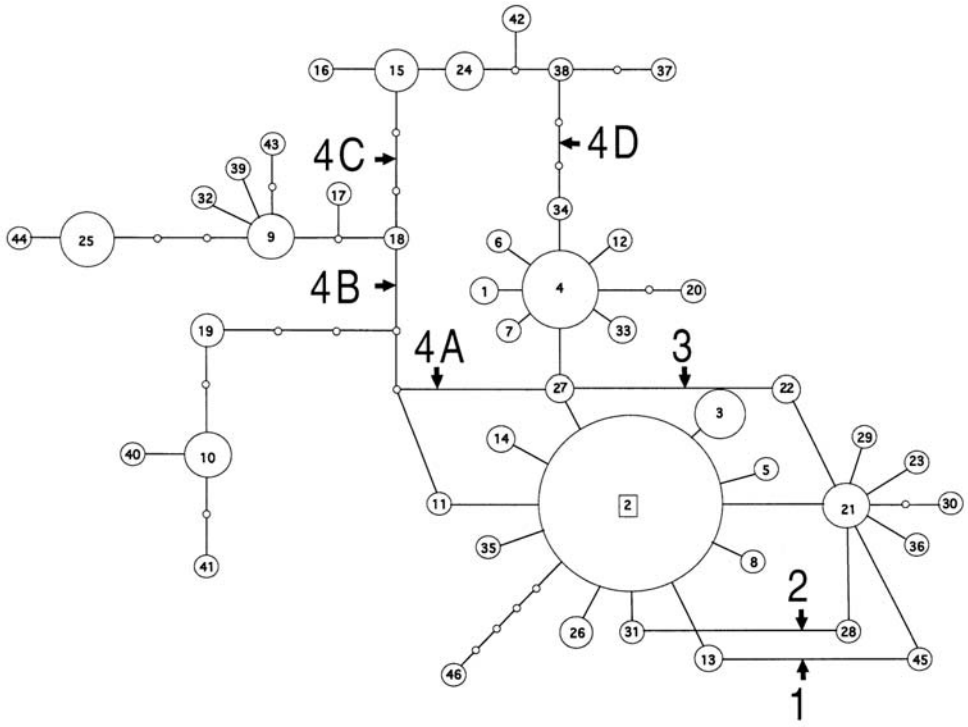


Figure 3. Nesting levels shown as hierarchical clade groupings of the haplotype network. The total cladogram is comprised of two four-step clades: 4-1 and 4-2 (Pfenninger & Posada 2002).

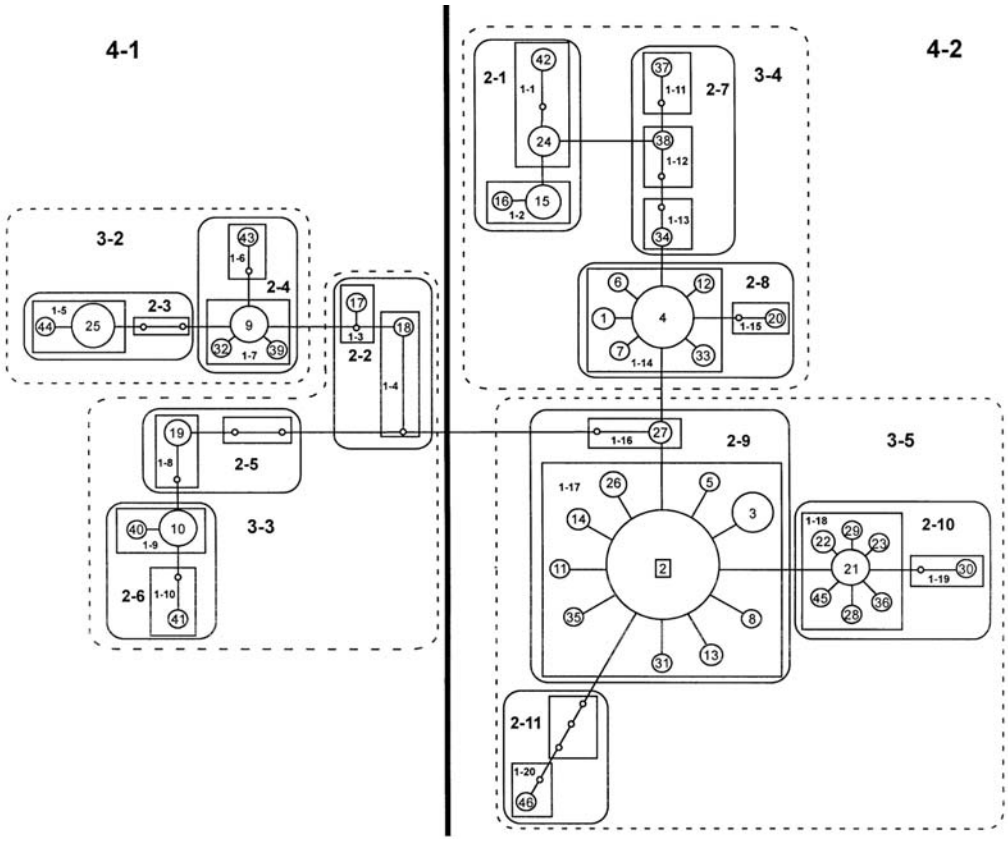


Figure 4. Example network using five haplotypes and the nesting levels. This network was then used to write the input file for GeoDis shown in Table 1.

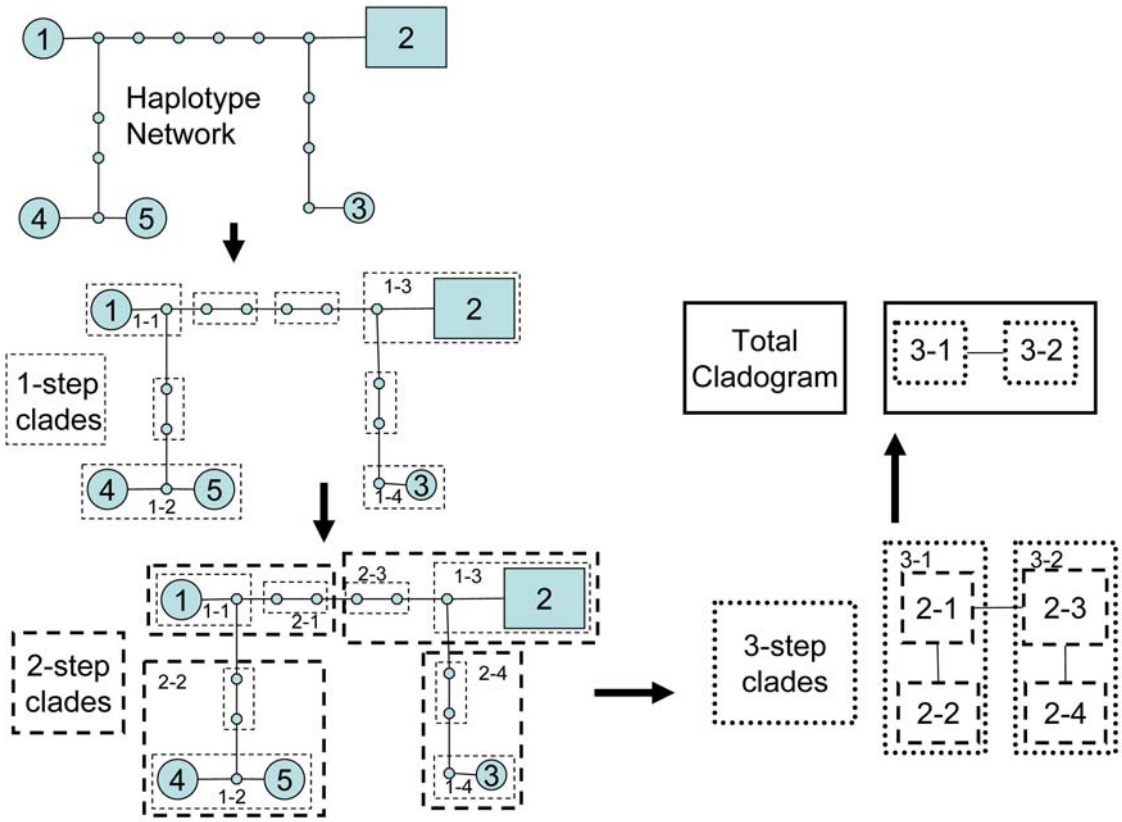
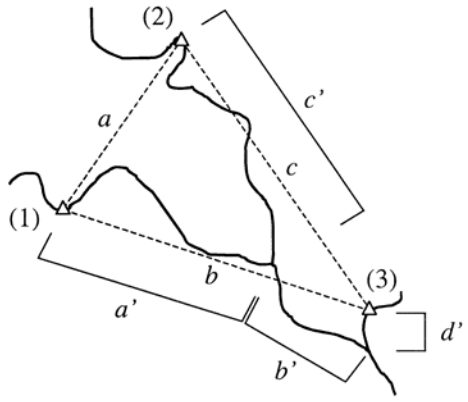
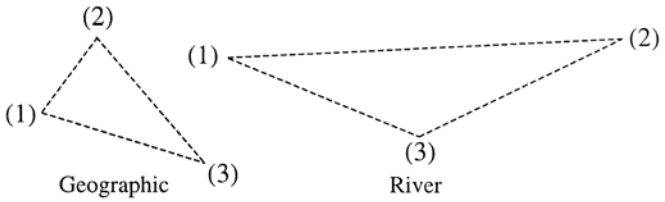


Figure 5. Diagram of the distance differences between geographic and river approaches for three sites labeled (1), (2), and (3). (A) Great circle geographic distances are labeled a, b, and c, while river/linear distances are labeled a', b', c' and d'. (B) Illustration of the differences in the geographic and river distances between sites. (C) How the calculations are done between sites (Fetzner & Crandall 2003).

A.



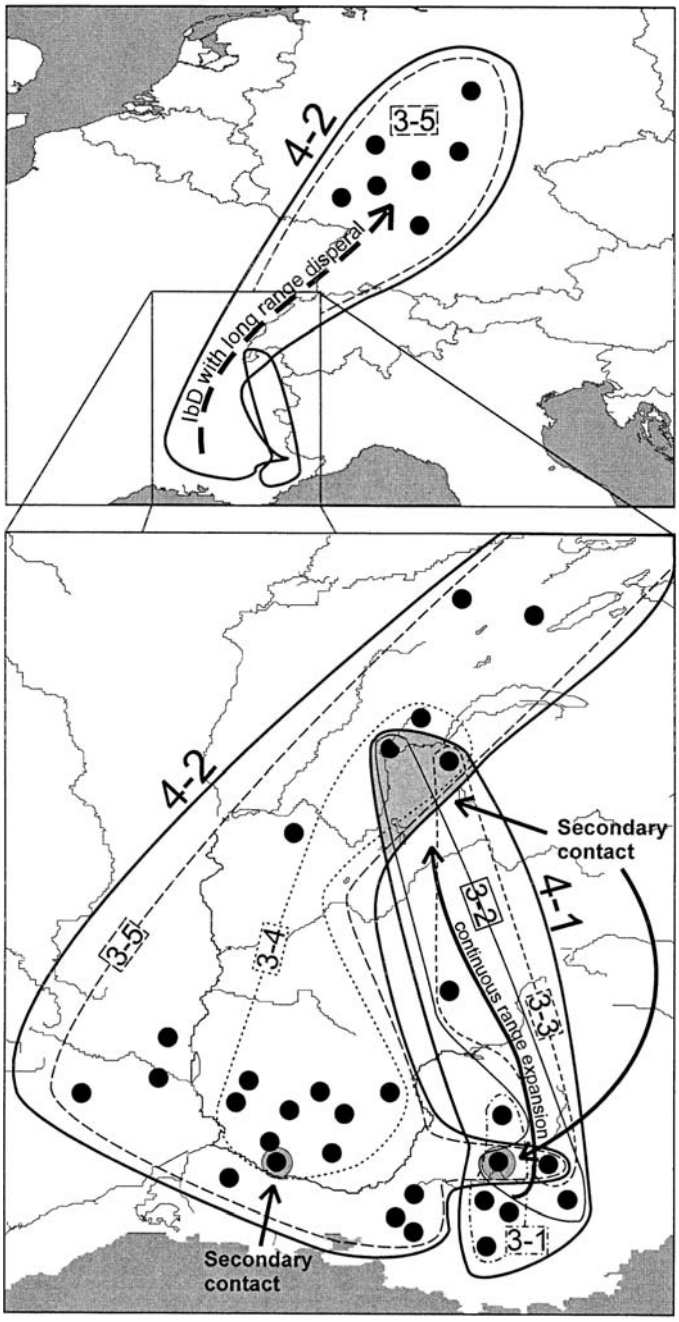
B.



C.

Population Distances	Geographic	River
(1) – (2)	*a*	*a'+ c'*
(1) – (3)	*b*	*a' + b' + d'*
(2) – (3)	*c*	*c' + b' + d'*

Figure 6. Geographic distribution of three and four step clades with inferred events determined with the inference key of Templeton (2005). Areas of secondary contact between clades are shaded (Pfenninger & Posada 2002).



CHAPTER 2

SUBTERRANEAN PHYLOGEOGRAPHY OF FRESHWATER CRAYFISHES SHOWS EXTENSIVE GENE FLOW AND SURPRISINGLY LARGE POPULATION SIZES*

ABSTRACT

Subterranean animals are currently viewed as highly imperiled, precariously avoiding extinction in an extreme environment of darkness. This assumption is based on a hypothesis that the reduction in visual systems and morphology common in cave faunas reflects a genetic inability to adapt and persist coupled with the perception of a habitat that is limited, disconnected, and fragile. Accordingly, 95% of cave fauna in the United States are presumed endangered due to surface environmental degradation and limited geographic distributions. Our study explores the subterranean phylogeography of stygobitic crayfishes in the southeastern United States, a global hotspot of groundwater biodiversity, using extensive geographic sampling and molecular data. Despite their endangered status, our results show that subterranean crayfish species have attained moderate to high levels of genetic diversity over their evolutionary histories with large population sizes and extensive gene flow among karst systems. We then compare the subterranean population histories to those of common surface stream-dwelling crayfishes. Our results show recent drastic declines in genetic variability in the surface crayfish and suggest that these species also warrant conservation attention.

---

* This chapter was published as: *Buhay, J. E.* and K. A. Crandall. 2005. Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. Molecular Ecology 14:4259-4273.

INTRODUCTION

According to the Nature Conservancy, 95% of subterranean fauna in North America is considered vulnerable or imperiled using criteria similar to the IUCN-World Conservation Union Red List (Master 1991; Culver *et al.* 2000). The listings are based mostly on surface threats to groundwater systems (Danielopol *et al.* 2003), small geographic ranges (Culver *et al*. 2000), and habitat destruction, not in-depth species-specific biologic studies. In fact, current scientific information on subterranean fauna is scarce, leaving the field of biospeleology and the unique biome in the dark. The convergent nature of cave life obscures species' relationships and geographic boundaries, while the inaccessibility of the underground microhabitat makes physical counts of census sizes almost impossible to confidently assess. Molecular genetic approaches are best employed in these situations to accurately estimate biodiversity and critically evaluate the conservation status of elusive organisms (DeSalle & Amato 2004).

Two hypotheses (as reviewed by Kane 1982) have been proposed concerning the genetic diversity and hence, the conservation status and extinction risk (Spielman *et al.* 2004), of subterranean fauna. Barr (1968) suggested that a genetic bottleneck initially occurs during the separation of the surface ancestor from its obligate cave-dwelling descendent. Barr suggested that this bottleneck is short in duration and that cave populations recover from the break in gene flow by range expansion and population growth into new uninhabited subterranean areas. In contrast, Poulson & White (1969) proposed that older fauna show low genetic variability due to the long isolation and adaptation to the stable underground environment. They also suggested that the decrease

in phenotypic variance in visual structures and morphological traits reflects a decreased genetic variability. Poulson & White (1969) also stressed the probable relationship between reduced genetic variability with the reduction of population size, reduced rate of population growth, longer maturation times, and longer life spans. Previous studies (Avise & Selander 1972; Swofford *et al.* 1980; Koppelman & Figg 1995) on aquatic obligate cave species (stygobites) were consistent with the Poulson and White hypothesis, but each of the studies had sparse sampling across small geographic areas within the species' ranges and these studies were conducted using allozymes, which can underestimate genetic diversity. Our study tests these two alternative hypotheses for the first time using exceptional sampling and high resolution genetic data from a group of subterranean crayfishes. We also compare our cave crayfish findings to those of two common surface stream-dwelling crayfish species for broader understanding of subsurface and surface freshwater habitats and conservation.

MATERIALS AND METHODS

Study Organisms

One of the largest animals in caves are blind crayfish, which are found in all kinds of subterranean aquatic areas, including deep rivers and lakes, small seeps, rimstone pools, and mudholes. A group of stygobitic crayfishes in the genus *Orconectes* inhabits the karst groundwaters of the western escarpment of the Cumberland Plateau, ranging from eastern Kentucky south to northern Alabama (Hobbs & Barr 1972; Hobbs *et al.* 1977). As currently recognized, there are three obligate cave-dwelling *Orconectes* species along

the Plateau: *O. incomptus*, *O. australis* (with two subspecies, *australis* and *packardi*), and *O. sheltae,* which was only known from one Mississippian Age cave in Alabama (Cooper 1975; Cooper & Cooper 1997) and is currently presumed extinct, with the last sighting by Hobbs & Bagley (1989). *O. incomptus* is found only in Ordovician Age limestone in an area just west of the escarpment. *O. australis* is found in Mississippian Age limestone along the escarpment, which was formed by the recession and erosion of the Cumberland Plateau in an eastward direction, allowing for cave development on the western side. The conservation categories for these species are: *Orconectes australis australis* (IUCN stable), *O. a. packardi* (IUCN vulnerable), *O. incomptus* (IUCN vulnerable), *O. sheltae* (unlisted).

To thoroughly investigate the genetic diversity and phylogeographic patterning of this unique assemblage, we collected mostly tissue samples (a claw or leg which are regenerated) from 421 individuals from 67 caves spanning the entire geographic range (Table 1). Non-destructive sampling involved returning the captured individual to the capture site immediately after removal of claw or leg. In a few cases, one or two voucher male specimens (preserved in 90% ethanol at the Monte L. Bean Museum at Brigham Young University) were taken from caves discovered after Hobbs *et al.* (1977)'s distribution list of cave crayfish localities to serve as voucher specimens for these caves.

For comparison to surface species, we chose two common surface stream-dwelling *Orconectes* species for which we have substantial molecular data and thoroughly-sampled distributions as part of other research investigations. *Orconectes luteus* is a wide-ranging surface species throughout Missouri, while *O. juvenilis* has a

restricted range in the Upper Cumberland River and Kentucky River basins of Kentucky. Both *O. luteus* and *O. juvenilis* are assigned to the subgenus Procericambarus of the genus *Orconectes* and are IUCN stable species.

Data Collection

Genomic DNA was extracted using standard methods and the 16S mtDNA gene was amplified during PCR with primers 16sf-cray: GACCGTGCKAAGGTAGCATAATC and 16s-1492r: GGTTACCTTGTTACGACTT (Crandall & Fitzpatrick 1996). The 16S mtDNA is the most variable gene for freshwater crayfishes (Crandall 1997; Fetzner & Crandall 2003). Cycle-sequencing reactions were run with purified PCR products and the Big Dye Ready-Reaction kit on a Perkin Elmer Thermocycler. Reactions were cleaned using Millipore plates and then sequenced using an ABI377 automated DNA sequencer. Sequences were edited and aligned by eye using BioEdit (Hall 1999). GenBank (http://www.ncbi.nlm.nih.gov) accession numbers of the 16S mtDNA haplotypes used for this study are: *Orconectes a. packardi* AY853595-AY853610; *O. incomptus* AY853611-AY853614; *O. sp. nov* AY853615-AY853619; *O. a. australis* AY853620-AY853663; *Cambarus gentryi* AY853664; and *Cambarus graysoni* AY853665. R. Ziemba collected samples of *Orconectes juvenilis* (n=100 individuals), which we sequenced for 16S (unpublished data, available on request from R. Ziemba). The *Orconectes luteus* (n=393 individuals) aligned 16S dataset (Fetzner & Crandall 2003; GenBank AF376483-AF376521) was provided by J. Fetzner. Both surface species were amplified in PCR and sequenced using primers 16s-1492r and 16s-17sub: ATASRGTCTRACCTGCCC (Fetzner & Crandall 2003).

Phylogenetic Analyses

Phylogenetic analyses included 69 unique haplotypes (485 base pairs) from the 421 cave individuals and two outgroup sequences from the closest relatives *Cambarus gentryi* and *C. graysoni* (Sinclair *et al.* 2004; Buhay *et al.*, unpublished data). The Bayesian Analysis (Ronquist & Huelsenbeck 2003) was run for 10 million generations using four chains, sampling 1/1000 trees with parameters nst=6 and rates=adgamma. We discarded the burnin (first 1001 trees of 10001 total determined by Tracer (http://evolve.zoo.ox.ac.uk/software.html), checked for convergence using Tracer, and constructed a 50% majority rule consensus tree. Five independent runs of the same dataset with random start trees resulted in nearly-identical results. Posterior probabilities (PP) greater than 95% are considered significant support for a clade (Huelsenbeck & Ronquist 2001). The maximum likelihood analysis was run in PAUP* (Swofford 2001) by heuristic search (fast-stepwise addition with random seed) with 500 replicates using the TrN+I+G model of evolution selected by ModelTest (Posada & Crandall 1998). Nodal support was assessed using 100 bootstrap (BS) replicates (Felsenstein 1985) with strong clade support of 70% (Hillis & Bull 1993).

Genetic Diversity and Effective Population Sizes

To address current and recent historical levels of variation, genetic diversity and effective population sizes within each surface and cave lineage were determined using several methods. We used different estimators of the parameter $\theta = 2N_{e(f)}\mu$ for maternally-inherited mitochondrial DNA, to determine effective population size (N_e) with a mutation

rate μ (2.2 x 10^{-8} substitutions per site per year; based on Cunningham *et al.* 1992 estimate for crabs) with generation times of two years for surface-dwelling species (Hobbs 1991) and ten years for stygobitic species (Cooper 1975), and an equal sex ratio (Cooper 1975).

Current genetic diversity (θ_π; Nei 1987 equations 10.5 or 10.6, and the standard error, equation 10.7) was assessed using DnaSP 4.0 (Rozas *et al.* 2003). Watterson's (1975) historical genetic diversity estimates (θ_W) were determined using LAMARC (http://evolution.genetics.washington.edu/lamarc.html; Kuhner *et al.* 2004). Current genetic diversity estimates (θ_π) are based on pairwise differences between sequences, while historical diversity estimates (θ_W) are based on the number of segregating sites among the sequences. These two methods used together provide insight into population dynamics over recent evolutionary history (Templeton 1993; Crandall *et al.* 1999; Pearse & Crandall 2004). Differences between current diversity and recent historical diversity are indicative of recent bottlenecks (if $\theta_\pi < \theta_W$) or recent population growth (if $\theta_\pi > \theta_W$) (Templeton 1993; Sinclair *et al*. 2002; Roman & Palumbi 2003; Yu *et al.* 2003).

Pairwise comparisons were used for genealogical estimates of diversity (θ_1, θ_2, $\theta_{Ancestor}$) and divergence times using the program IM (Isolation-Migration Model: Nielsen & Wakeley 2001; Hey 2005; Won and Hey 2005; http://lifesci.rutgers.edu/~heylab/heylabsoftware.htm#IM). The HKY model with an inheritance scalar of 0.25 for mitochondrial DNA was used with a random seed to initiate the run. A burnin of 200000 steps was discarded before recording genealogical steps, and each comparison was run until the effective sample sizes (ESS) were larger than 1000,

and in most cases, over 1 million. Multiple independent runs with random start seeds were done to ensure values were converging on similar estimates. Maximum likelihood estimates of diversity were used to determine bottleneck (<1) or growth trends (>1) between descendent pairs and their ancestors (Descendents:Ancestor ratio) to test the two competing hypotheses about subterranean genetic diversity (Poulson & White 1969 and Barr 1968). Descendent:Ancestor ratios were computed by $(\theta_1 + \theta_2) / \theta_{Ancestor}$ for each pair.

Phylogeographic Analyses

Nested Clade Analysis (NCA: Templeton *et al.* 1995; Templeton 1998) was used to test the null hypothesis of no genetic differentiation between sampled sites and provide insight into historical processes. The program TCS (Clement *et al.* 2000) was used to construct the haplotype network and GeoDIS (Posada *et al.* 2000) was used to test for significant associations between geographic cave locations and genetic distances over 5000 random permutations. Latitude and longitude coordinates of cave localities (at the entrance) were used for the geographic analysis. Haplotypes with the most connections and the highest frequencies are thought to be older, while haplotypes on the tips are more recently evolved. Clade distances (Dc) represent geographic ranges of the clades at each step-level. Nested clade distances (Dn) represent the average distances of samples with a particular haplotype with respect to the geographic center of the clade. Inferences about the historical processes that gave rise to the current genetic patterns were made using the 2004 inference key from A. R. Templeton (http://darwin.uvigo/es/software/geodis.html).

RESULTS

Phylogenetic Analysis of 16S mtDNA haplotypes

There are several operational methods available to delineate species boundaries using statistically testable frameworks, as reviewed by Sites & Crandall (1997) and Sites & Marshall (2003). The Genealogical Concordance Species Concept (Avise & Ball 1990; Baum & Shaw 1995) is a lineage-based extension of the phylogenetic species concept, in which there is concordance among multiple characters (genetic, environmental, geographic, etc.). A genealogical species is a group of organisms whose members are more closely related to each other ("exclusivity") than to any other organisms outside the group (Baum and Shaw 1995).

We determined the phylogenetic relationships among the two extant species (*Orconectes incomptus* and *O. australis*) using sequence data from the mitochondrial 16S gene (485 base pairs) and identified four distinct lineages: *O. a. packardi*, *O. incomptus*, *O. a. australis*, and *O. sp. nov.* (Fig. 1 and Table 1), each with significant posterior probability support. The cave-dwelling *Orconectes* members are most closely related to burrowing members of the genus *Cambarus* (Fetzner 1996; Crandall & Fitzpatrick 1996; Sinclair *et al*. 2004), rather than to the surface-dwelling members of *Orconectes,* as was previously thought based on similar (convergent) male morphology (Hobbs & Barr 1972), and accordingly, *Cambarus gentryi* and *C. graysoni* were used as the closest outgroup taxa (Sinclair *et al*. 2004; Buhay & Crandall, unpublished data).

The most basal member, *O. a. packardi,* was represented by 16 unique mtDNA 16S haplotypes from 13 Mississippian Age caves and 93 individuals, and is distributed

from Rockcastle County, Kentucky south to Pulaski County, Kentucky (Fig. 1: range shown as blue circles, haplotypes 1-16). *O. incomptus* was represented by four unique haplotypes from three Ordovician Age caves in Jackson County, Tennessee (Fig. 1: range shown as pink triangles, haplotypes 17-20). A new species, *O. sp. nov*., found along the Kentucky-Tennessee border (Wayne and Clinton Counties, Kentucky south to northern Fentress County, Tennessee), included five unique haplotypes from eight Mississippian Age caves and 40 individuals (Fig. 1: range shown as green pentagons, haplotypes 21-25). *O. a. australis* was represented by 321 individuals from southern Fentress County, Tennessee south to Madison County, Alabama and included 44 unique haplotypes from 43 Mississippian Age caves (Fig. 1: range shown as orange squares, haplotypes 26-69). Genetic data was acquired from type locality specimens: *O. a. packardi* (Sloans Valley Cave, Pulaski County, Kentucky), *O incomptus* (Cherry Cave, Jackson County, Tennessee) and *O. a. australis* (Shelta Cave, Madison County, Alabama), and this information was used to clarify species boundaries and their geographic distributions.

Each of these lineages will be considered distinct species based on genetic and geographic concordance ((Avise & Ball 1990; Baum & Shaw 1995). Rather than two species (*O. australis* and *O. incomptus*), there are five stygobitic cave *Orconectes* species on the Cumberland Plateau, including the unsampled, possibly extinct *O. sheltae*.

Genetic Variation, Effective Population Sizes, and Divergence Times

Estimates of current (θ_π) and historical (θ_W) genetic diversity were moderate to high (Nei 1987) for the cave-dwellers, with the exception of *O. sp. nov* (Table 2). Similarly, current effective population sizes (N_e) were also higher than expected,

suggesting the occurrence of a vast groundwater network unknown to humans, but as accessible habitat to the stygobitic crayfish. Surprisingly, current (θ_π) and historical (θ_W) estimates for the stygobites were similar (Table 2, with exception of *O. a. australis* which exhibited decline), whereas both surface species estimates show serious recent declines ($\theta_\pi < \theta_W$).

We used a coalescent-based method (Nielsen & Wakeley 2001) to determine genetic diversity over the genealogical histories of each cave species to test the two competing hypotheses regarding genetic diversity of ancestors versus descendents. Using pairwise species comparisons, we determined genealogical diversity (θ_1 and θ_2) for each crayfish species and $\theta_{Ancestor}$ for their common ancestor, along with their times since divergence (Table 3). These results show a growth trend (descendents/ancestor ratio > 1) after the initial split from the ancestors in cave species comparisons (Fig. 2).

The estimated divergence times for the cave crayfish species are much older than previous speculation (Hobbs *et al.* 1977). Given the broad credibility intervals (90% highest posterior probability densities; HPD) for the *O. a. packardi* - *O. incomptus* and *O. incomptus* - *O. sp. nov.* comparisons, it appears that more loci are needed to resolve divergence times for these species. It is also possible that more individuals of *O. incomptus* are needed for the IM pairwise analyses, since only eight individuals from three caves of the ten known sites were sampled for this study. *O. incomptus* is listed in Tennessee as a "management concern species" and as a "vulnerable species" by the International Conservation Union (IUCN) which required that sampling restrictions be placed on the collection permit. Interestingly, the split between *O. sp. nov.* and *O. a.*

australis was estimated to be 110 million years ago (90% HPD interval: 105-116 MYA), in the mid-Cretaceous, which was speculated to be the beginnings of cave invasion for the genus *Cambarus* (Hobbs and Barr 1960). The lower bounds of the 90% HPD intervals for the other two comparisons (*O. a. packardi – O. incomptus* at 125 MYA; *O. incomptus – O. sp. nov.* at 102 MYA) are similar to that of the *O. sp. nov. - O. a. australis* split. Such calculations necessarily make a number of simplifying assumptions and the resulting dates should be taken with caution, however, as outlined below, these divergence times nicely correspond to geological events that might cause such divergences.

Nested Clade Analysis of Cave Crayfish

To explain how the cave species attained high levels of genetic variation, we used Nested Clade Analysis to uncover the major historical processes and patterns (Templeton 2001). A statistical parsimony network was constructed using a 95% confidence interval which resulted in 69 unique haplotypes, 34 1-step clades, 14 2-step clades, eight 3-step clades, and three 4-step clades in the total cladogram (Table 4, Fig. 3). The statistical parsimony analysis revealed two haplotypes as ancestral, *O. a. packardi* haplotype 7 and *O. a. australis* haplotype 27, and these are shown as rectangles on Fig. 4. *O. a. packardi* haplotype 8 is connected to *O. incomptus* haplotype 17 by ten mutational steps (the significant 95% level was nine steps). *Cambarus gentryi* and *C. graysoni* were outside the 95% level, at 21 and 25 mutational steps, respectively, from haplotype 2 of *O. a. packardi*.

To geographically illustrate the historical speciation routes, we used the eight 3-step clades because they mostly resulted in significant inferences of "contiguous range expansion" or "isolation by distance" and they show “big picture” historical biogeographical patterns (Table 5). On Figure 4, *O. a. packardi* is shown as clades 3-1 (light blue) and clade 3-2 (dark blue) in the network and as circles on the corresponding map, and *O. incomptus* is clade 3-3 (pink) and is represented on the map as pink triangles. Clade 3-4 (green) is *O. sp. nov.* and is marked as green pentagons on the map, while four 3-step clades (3-5 through 3-8) comprise *O. a. australis* (marked as squares on the map of Figure 4). The 3-step clades of *O. a. australis* geographically overlap extensively in central Tennessee, with several *australis* caves containing haplotypes from different 3-step clades (Table 1).

DISCUSSION

It was hypothesized that the surface ancestor to the cave *Orconectes* originally expanded in a northeast direction from the Mississippi embayment, spawning obligate cave-dwelling species along the Cumberland Plateau en route to the northern Appalachian Mountains (Hobbs & Barr 1972). On the contrary, our phylogenetic and NCA results show that *Orconectes a. packardi*, which is distributed across the northern end of the Cumberland Plateau, is the most basal member of the cave assemblage. This suggests that the surface ancestor (a member of the burrowing genus *Cambarus*) ranged somewhere in eastern Kentucky and gave rise to the stygobitic species *O. a. packardi.* The other stygobitic species then diverged from a common ancestor with *O. a. packardi.*

The southern limit of the cave *Orconectes* distribution is the area just north of the Fall Line in Alabama, the pre-historic Atlantic Ocean coastline.

Our estimates of divergence times, though based on one mtDNA region, place the oldest cave *Orconectes* species on the Plateau present during the Cretaceous Period, which was the suggested time period for cave invasion by surface members of the genus *Cambarus* (Hobbs & Barr 1960). This timeframe also correlates with the age estimates of the oldest passages in Plateau caves and the beginnings of the eastward recession of the Cumberland Plateau (Barr 1961). It appears that the long evolutionary histories of crayfishes in the stable underground environment have allowed them to persist and accumulate genetic diversity, despite environmental changes on the surface, long generation times, and isolation over the past millions of years. Poulson & White (1969) speculated that older cave species would show low levels of diversity due to the long period of isolation underground, but it appears that levels of diversity for the cave crayfish species are not related to their estimated old divergence times.

One of the arguments made by Culver *et al.* (2000) for the endangered status of cave fauna was restricted geographic ranges, as most United States cave-adapted fauna (61%) are limited to caves in a single county. Although this is a common and practical approach for identifying possible conservation concerns for endemics and rare species as well as habitat types, species-specific information, particularly thorough geographic surveys (van Jaarsveld *et al.* 1998) and demographic and genetic studies (Lande 1988) are critical pieces of information in assessing the requirements needed for species survival. In this study, *O. a. australis* with the largest range of the stygobitic *Orconectes*,

is now currently known from eleven counties and has the highest genetic diversity of the cave crayfish species, but *O. incomptus*, with the smallest geographic range, and currently only known from nine caves in Jackson County and one cave in Putnam County, Tennessee, has the second highest diversity of the assemblage. *O. a. packardi* is currently known from three Kentucky counties, and *O. sp. nov.* is distributed across four counties in Kentucky and Tennessee, with moderate and low levels of genetic diversity, respectively. In our case, geographic range is not reflective of genetic diversity or conservation status for these cave species. Rather, the decline in genetic diversity over recent history ($\theta_\pi < \theta_W$; Templeton 1993; Sinclair *et al.* 2002; Roman & Palumbi 2003) is a better indicator for conservation concern with *O. a. australis* (currently 0.00894 from historically 0.01593), along with the low levels of diversity for *O. sp. nov.* (currently and historically, 0.00238 and 0.00242). It is interesting that the IUCN stable cave crayfish species, *O. a. australis,* shows a recent loss of diversity, whereas, the two IUCN vulnerable cave species, *O. incomptus* and *O. a. packardi*, show little difference between historical and current diversity estimates.

We show in Figure 4 a series of colonizations beginning in Kentucky with *O. a. packardi* and progressing down the Cumberland Plateau in a leading-edge small-stepwise manner, following the flow of pre-historic waters. This colonization pattern is consistent for animal groups limited by mountain landscapes and by dispersal ability, particularly in response to glacial advance and retreat cycles (Hewitt 1996; Hewitt 2000). Stygobitic crayfishes are severely limited in dispersal abilities by both subterranean and surface barriers, except during high water levels when they can migrate (or wash) out of caves

into a limestone-based surface stream across short distances, and into a nearby underground system via a spring resurgence or cave entrance. These findings suggest that pre-historic groundwater levels were much higher, and allowed for subterranean fauna to disperse over the surface landscape in small distances. Phreatic caves form below the water table, and as karst dissolves and creates voids, the water table lowers to fill in the spaces, which increases groundwater habitat for stygobites (White 1988). Although the major surface rivers along the Cumberland Plateau historically and currently flow in a southern direction, ongoing cave development and subsequent groundwater lowering have probably lead to "isolation by distance" and the prevention of further stepwise range expansion of the species and clades.

Contiguous range expansion followed by periods of isolation appears to be the main mechanism for the increased variation within the cave crayfish species. A similar trend has been reported for invasive and introduced species (Tsutsui *et al.* 2000; Kolbe *et al.* 2004), in which genetic diversity and population size accumulates and recovers, rather than resulting in a series of bottlenecks leading to lower diversity and extirpation. One example (Sbordoni 1982) has also been documented for a troglobitic beetle species in Italy, in which 50 individuals were introduced into an isolated cave with no beetles. After thirty years, the estimated census size was 15,000 individuals with a greater genetic diversity than the original "founder" population. Clearly, pre-adaptation and continued expansion into suitable habitat of the subterranean environment allowed cave crayfish to successfully and repeatedly colonize new areas, regardless of population size or genetic diversity of the founder populations.

O. a. packardi, *O. incomptus,* and *O. sp. nov*. are currently distributed across small geographic ranges (four counties or less), possibly due to the hydrologic impacts of the pre-historic watercourses of the Cumberland River. Caves in the path of the Cumberland River during its formation would have been completely submerged by surface waters. The missing haplotypes in the parsimony network may be evidence of past drainage evolution events between the ancestors of *O. a. packardi* and *O. incomptus*, and *O. incomptus* and *O. sp. nov*. leading to local extirpations, range restrictions, and lower diversity in those species compared to *O. a. australis*.

Orconectes luteus and *O. juvenilis* are currently listed as IUCN stable species in conservation status based on the fact that they are widespread throughout their ranges (Taylor *et al.* 1996), but it appears that they are in need of some protection and study (based on the large discrepancy between θ_π and θ_W for both common surface-dwellers). The stable underground environment may provide enough suitable "habitat pockets and hideouts" to buffer the subterranean biota from the direct impacts of ongoing surface pressures, but it appears that the surface species are not so fortunate. It is surprising that species considered to be common stream inhabitants show a reduction in population sizes whereas most of the cave species show consistent population sizes over evolutionary time.

We also hope that these findings shed light on the conservation status of other subterranean taxa and propel biospeleogists to test their assumptions concerning biodiversity. We suggest that management strategies be redirected toward molecular genetic assessments of effective population sizes and diversity (Thorpe *et al*. 1995) for

cave species and other elusive fauna considered to be on the brink of extinction because of a lack of scientific information (Holmes 2001). Current cave conservation activities focus on general efforts to protect subterranean habitat by purchasing karstlands, avoiding pollution catastrophes, and gating highly-visible entrances. Although these are important defenses for the protection of the biome, the ultimate goal of cave conservation is the sustainability of each unique obligate cave-dwelling species. Stochastic factors are well-known causes of biodiversity losses, yet, current research shows that the genetic factors, specifically loss of heterozygosity and inbreeding, can play major roles in driving endangered and threatened species to extinction (Brook *et al.* 2002; Spielman *et al.* 2004). Hopefully, this research will turn the efforts of conservation agencies toward protecting gene flow routes and areas of connectivity to prevent future imperilment of the amazing fauna under our feet and the common inhabitants in our backyards.

ACKNOWLEDGEMENTS

Many dedicated cavers assisted with the fieldwork, especially N. Mann, G. Moni, M. Niemiller, W. Simpson, K. Toepke, W. Walden, and D. Withers, as well as the local caving groups of the National Speleological Society: Dayton Underground, Central Ohio, Greater Cincinnati, Spencer Mountain, Nashville, Golden Pond, Huntsville, and the Tennessee Cave Survey. We are also grateful to the Tennessee Department of Conservation for landowner reconnaissance and transportation and to Alabama, Tennessee, and Kentucky wildlife agencies for collecting permits. Special thanks to the Elmore, Evans, and Gulley households for part-time residence. T. Barr and J. Cooper

provided valuable insight and continual support, and M. Achtman reviewed an earlier draft. J. Fetzner Jr. aided with DNA extractions and provided *O. luteus* data. R. Ziemba provided *O. juvenilis* data. This project was funded by the Cave Research Foundation, National Speleological Society, University of Alabama Graduate School and Biology Department, Brigham Young University Graduate Studies and Integrative Biology Department, and United States Fish and Wildlife to J.E.B, and the National Science Foundation (DDIG DEB 0508580) to K.A.C and J.E.B.

REFERENCES

Avise JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology,* **7**, 45-67.

Avise JC, Selander RK (1972) Genetics of cave-dwelling fishes of the genus *Astyanax*. *Evolution*, **26**,1-19.

Barr TC Jr (1961) The Caves of Tennessee. *Tennessee Division of Geology Bulletin*, **64**,1-567.

Barr TC Jr (1968) Cave ecology and the evolution of troglobites. *Evolutionary Biology, Vol. 2*. (eds Dobzhansky T, Hecht M, Steere WC), pp. 35-102. Appleton Century Crofts, New York.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to PlantBiosystematics* (eds Hoch PC, Stephenson AG), pp. 289–303.Missouri Botanical Garden, St. Louis.

Brook BW, Tonykn DW, O'Grady J, Frankham R (2002) Contribution of inbreeding to

extinction risk in threatened species. *Conservation Ecology*, **6**, 16.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1659.

Cooper JE (1975) Ecological and behavioral studies in Shelta Cave, Alabama with emphasis on decapod crustaceans. PhD dissertation, University of Kentucky.

Cooper JE, Cooper MR (1997) New troglobitic crayfish of the genus *Orconectes*, subgenus *Orconectes* (Decapoda: Cambaridae), endemic to Shelta Cave, Huntsville, Alabama. *Journal of Cave and Karst Studies*, **59**, 119-127.

Crandall KA (1997) Genetic variation within and among crayfish species. *Freshwater Crayfish*, **11**, 135-145.

Crandall KA, Fitzpatrick JF Jr (1996) Crayfish molecular systematics: using a combination of procedures to estimate phylogeny. *Systematic Biology,* **45**, 1-26.

Crandall KA, Posada D, Vasco D (1999) Effective population sizes: missing measures and missing concepts. *Animal Conservation*, **2**, 317-319.

Culver DC, Master LL, Christman MC, Hobbs HH, III (2000) Obligate cave fauna of the 48 contiguous United States. *Conservation Biology*, **14**, 386-401.

Cunningham CW, Blackstone NW, Buss LW (1992) Evolution of king crabs from hermit crabs ancestors. *Nature*, **355**, 539-542.

Danielopol DL, Griebler C, Gunatilaka A, Notenboom J (2003) Present state and future prospects for groundwater ecosystems. *Environmental Conservation*, **30**, 104-130.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews in*

Genetics, **5**, 702-712.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783-791.

Fetzner, JW Jr (1996) Biochemical systematics and evolution of the crayfish genus *Orconectes* (Decapoda: Cambaridae). *Journal of Crustacean Biology*, **16**, 111-141.

Fetzner JW Jr., Crandall KA (2003) Linear habitats and the nested clade analysis: an empirical evaluation of geographic vs. river distances using an Ozark crayfish (Decapoda: Cambaridae). *Evolution*, **57**, 2101-2118.

Hall TA (1999) BioEdit:a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**, 182-192

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLOS Biology*, **3**, e193.

Hobbs HH Jr, Barr TC Jr (1960) The origins and affinities of the troglobitic crayfishes of North America (Decapoda, Astacidae). I. The genus *Cambarus*. *American Midland Naturalist*, **64**,12-33.

Hobbs HH Jr, Barr TC Jr (1972) Origins and affinities of the troglobitic crayfishes of North America (Decapoda: Astacidae). II. Genus *Orconectes*. *Smithsonian Contributions to Zoology*, **105**, 1-84.

Hobbs HH Jr, Hobbs HH III, Daniel MA (1977) A review of the troglobitic decapod crustaceans of the Americas. *Smithsonian Contributions to Zoology*, **244**, 1-176.

Hobbs HH III, Bagley FM (1989) Shelta Cave Management Plan, Biology Subcommittee of the Shelta Cave Management Committee. 78 p.

Hobbs HH III (1991) Decapoda. In: *Ecology and Classification of North American Freshwater Invertebrates* (eds Thorp JH, Covich AP), pp. 823-858. Academic Press, San Diego, California.

Holmes EE (2001) Estimating risks in declining populations with poor data. *Proceedings of the National Academy of Sciences USA*, **98**, 5072-5077.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754-755.

Kane TC (1982) Genetic patterns and population structure in cave animals. In: *Environmental Adaptation and Evolution* (eds Mossakowski D, Roth G), pp. 131-149. Gustav Fisher & Stuttgart, New York.

Kolbe JJ, Glor RE, Schettino LR, Lara AD, Larson A (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177-181.

Koppelman JB, Figg DE (1995) Genetic estimates of variability and relatedness for conservation of an Ozark cave crayfish species complex. *Conservation Biology*, **9**, 1288-1294.

Kuhner MK, Yamato J, Beerli P (2004) LAMARC v.1.2.1, University of Washington, http://evolution.gs.washington.edu/lamarc.html

Lande R (1993) Genetics and demography in biological conservation. *Science*, **241**, 1455-1460.

Master LL (1991) Assessing threats and setting priorities for conservation. *Conservation Biology*, **5**, 559-563.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Pearse DE, Crandall KA (2004) Beyond FST: Analysis of population genetic data for conservation. *Conservation Genetics*, **5**, 585-602

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487-488.

Poulson TL, White WB (1969) The cave environment. *Science*, **165**, 971-981.

Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science,* **301**, 508-510.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19,** 1572-1574.

Rozas J, Sánchez-DelBarrio JC, Messegyer X, Rozas R (2003) DnaSP, DNA

polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496-2497.

Sbordoni V (1982) Advances in speciation of cave animals. In: *Mechanisms of Speciation* (ed Barigoozi C), pp. 219-240. Alan R. Liss, New York.

Sinclair EA, Fetzner JF Jr, Buhay JE, Crandall KA (2004) Proposal to complete a phylogenetic, taxonomy, and systematic revision for freshwater crayfish (Astacidae). *Freshwater Crayfish*, **14**, 21-29.

Sinclair EA, Costello B, Courtenay JM, Crandall KA (2002) Detecting a genetic bottleneck in Gilbert's Potoroo (Potorous gilbertii) (Marsupialia: Potoroidae), inferred from microsatellite and mitochondrial DNA sequence data. *Conservation Genetics*, **3**, 191-196.

Sites JW, Jr, Crandall KA (1997) Testing species boundaries in biodiversity studies. *Conservation Biology*, **11**, 1289-1297.

Sites JW, Jr, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, **18**, 462-470.

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences USA*, **101**, 15261-15264.

Swofford DL (2001) *PAUP**: *Phylogenetic Analysis Using Parsimony* (and Other Methods).* Version 4.0b7 beta. Sinauer Associates, Sunderland, Massachusetts.

Swofford DL, Branson BA, Sievert GA (1980) Genetic differentiation of cavefish populations. *Isozyme Bulletin*, **13,** 109-110.

Taylor CA, Warren ML Jr, Fitzpatrick JF Jr *et al.* (1996) Conservation status of crayfishes of the United States and Canada. *Fisheries*, **21**, 25-38.

Templeton AR (1993) The "Eve" Hypothesis: a genetic critique and reanalysis. *American Anthropologist,* **95**, 51-72.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397

Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, **10**, 779-791.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographic distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma Tigrinum*. *Genetics*, **140**, 767–782.

Thorpe JP, Smartt J, Allcock AL *et al*. (1995) Genetic diversity as a component of biodiversity. In: *Global Biodiversity Assessment* (ed Heywood VH), pp. 57-87. Cambridge University Press, New York.

Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences USA*, **97**, 5948-5953.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theories in Population Biology*, **7**, 256-276.

White WB (1988) *Geomorphology and Hydrology of Karst Terrains.* Oxford Press, New York.

Won Y-J, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology & Evolution*, **22**, 297-307.

Van Jaarsveld AS, Freitag S, Chown SL *et al*. (1998) Biodiversity assessment and conservation strategies. *Science*, **279**, 2106-2108.

Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li WH (2003) Low nucleotide diversity in chimpanzees and bonobos. *Genetics*, **164**, 1511-1518.

This study represents part of Jen Buhay's doctoral research on the evolutionary history of cave crayfishes with Keith Crandall. Buhay's research interests focus on population genetics and phylogeography of surface and subterranean freshwater fauna. Keith Crandall is a Professor in the Integrative Biology and Molecular Biology Departments at Brigham Young University. Crandall works on a wide variety of organisms in an evolutionary context.

Table 1. List of cave *Orconectes* taxa, sampled caves, mtDNA 16S haplotype with number of individuals sequenced in parentheses, 3-step nested clade groupings, geographic information, and geologic age of cave sites used in this study. * represents a known introduced population from a nearby cave. ** represents type locality.

Species	Cave Name	16S Haplotype (# of individuals)	3-step Clade	State:County	Geologic Age
incomptus	Cherry**	19(2)	3-3	TN: Jackson	Ordovician
incomptus	Flynn Creek	17(1)	3-3	TN: Jackson	Ordovician
incomptus	North Fork	18(2), 20(3)	3-3	TN: Jackson	Ordovician
a. packardi	Teamers	1(1), 2(2)	3-1	KY: Rockcastle	Mississippian
a. packardi	Duvalts	2(1)	3-1	KY: Rockcastle	Mississippian
a. packardi	Pine Hill	2(1)	3-1	KY: Rockcastle	Mississippian
a. packardi	Fletcher Spring	7(2)	3-2	KY: Rockcastle	Mississippian
a. packardi	Cedar Creek	7(14)	3-2	KY: Pulaski	Mississippian
a. packardi	Dykes Bridge	7(3)	3-2	KY: Pulaski	Mississippian
a. packardi	Dave's	6(8), 7(2)	3-1, 3-2	KY: Pulaski	Mississippian
a. packardi	Big Sink	7(20)	3-2	KY: Pulaski	Mississippian
a. packardi	Hail	3(4), 4(3), 5(1)	3-1	KY: Pulaski	Mississippian
a. packardi	Wells	6(3), 7(1)	3-1, 3-2	KY: Pulaski	Mississippian
a. packardi	Jugornot	8(3),12(13), 13(1), 14(2), 15(1), 16(2)	3-2	KY: Pulaski	Mississippian
a. packardi	Coral	3(1)	3-1	KY: Pulaski	Mississippian
a. packardi	Sloans Valley**	9(1), 10(2), 11(1)	3-2	KY: Pulaski	Mississippian
sp. nov	Redmond Creek	24(9)	3-4	KY: Wayne	Mississippian
sp. nov	Grayson Gunner	23(1)	3-4	KY: Wayne	Mississippian
sp. nov	Stream	24(2), 25(2)	3-4	KY: Wayne	Mississippian
sp. nov	Tonya's	23(7)	3-4	KY: Wayne	Mississippian
sp. nov	Buffalo Saltpeter	23(3)	3-4	KY: Clinton	Mississippian
sp. nov	Clinton	21(5), 22(1)	3-4	TN: Pickett	Mississippian
sp. nov	Cornstarch	21(9)	3-4	TN: Fentress	Mississippian
sp. nov	Redbud	21(1)	3-4	TN: Fentress	Mississippian
a. australis	Fallen Entrance	27(6)	3-6	TN: Fentress	Mississippian
a. australis	Skillmans Mark	27(3), 30(1)	3-6	TN: Fentress	Mississippian
a. australis	Mountain Eye	27(4)	3-6	TN: Fentress	Mississippian
a. australis	Mill Hollow	27(16), 28(1), 50(1), 51(3)	3-6, 3-8	TN: Overton	Mississippian
a. australis	Raven Bluff	37(1)	3-6	TN: Overton	Mississippian
a. australis	Bailey's Webb	27(5)	3-6	TN: Overton	Mississippian
a. australis	Capshaw	27(12), 29(1)	3-6	TN: Putnam	Mississippian
a. australis	Knieps Spring	27(4)	3-6	TN: Putnam	Mississippian
a. australis	Blindfish	26(1), 27(2), 31(3), 32(1), 33(1)	3-6	TN: Putnam	Mississippian
a. australis	Virgin Falls	40(4)	3-7	TN: White	Mississippian
a. australis	Merrybranch	34(1), 35(7), 36(1), 40(22), 41(1), 42(1), 43(1), 44(4), 45(1)	3-6, 3-7	TN: White	Mississippian
a. australis	Lost Creek Resurgence	40(1)	3-7	TN: White	Mississippian
a. australis	Rumbling Falls	40(6)	3-7	TN: VanBuren	Mississippian
a. australis	Winching Hollow Water	35(9), 40(3)	3-6, 3-7	TN: VanBuren	Mississippian
a. australis	Glencora Spring	27(1), 40(4)	3-6, 3-7	TN: VanBuren	Mississippian
a. australis	Waterfall Hollow	54(7)	3-8	TN: VanBuren	Mississippian
a. australis	Lost Cove	51(10), 53(1)	3-8	TN: VanBuren	Mississippian
a. australis	Camps Gulf	40(2), 54(1)	3-7, 3-8	TN: VanBuren	Mississippian
a. australis	Laurel Creek	40(1), 51(17)	3-7, 3-8	TN: VanBuren	Mississippian
a. australis	Lower Norton Spring	49(1), 51(3)	3-8	TN: VanBuren	Mississippian
a. australis	Rocky River	46(5), 47(2)	3-8	TN: Warren	Mississippian
a. australis	Jaco Spring	48(4)	3-8	TN: Warren	Mississippian
a. australis	Cumberland Caverns*	46(1), 51(4)	3-8	TN: Warren	Mississippian
a. australis	Blowing	38(5)	3-7	TN: Warren	Mississippian
a. australis	Woodlee	39(1)	3-7	TN: Grundy	Mississippian
a. australis	Dry	39(1)	3-5	TN: Grundy	Mississippian
a. australis	Red Trillium	61(2)	3-5	TN: Grundy	Mississippian
a. australis	Big Mouth	61(4)	3-5	TN: Grundy	Mississippian
a. australis	Crystal	61(5)	3-5	TN: Grundy	Mississippian
a. australis	Smith Hollow NR1	61(4), 63(1)	3-5	TN: Grundy	Mississippian
a. australis	Lusk	51(1), 61(7), 64(1)	3-5, 3-8	TN: Coffee	Mississippian
a. australis	Pearson	61(26), 62(1)	3-5	TN: Franklin	Mississippian
a. australis	Wet	61(2)	3-5	TN: Franklin	Mississippian
a. australis	Dripping Spring	59(1)	3-5	TN: Franklin	Mississippian
a. australis	Witherspoon	51(7)	3-8	TN: Franklin	Mississippian
a. australis	Floorless	51(1), 52(1)	3-8	TN: Franklin	Mississippian
a. australis	Larkin Spring	65(2)	3-5	AL: Jackson	Mississippian
a. australis	Limrock Blowing	65(28), 67(1), 69(1)	3-5	AL: Jackson	Mississippian
a. australis	Doug Green	56(1)	3-5	AL: Jackson	Mississippian
a. australis	Langston	55(1)	3-5	AL: Jackson	Mississippian
a. australis	Scott	65(3)	3-5	AL: Madison	Mississippian
a. australis	Hering	57(1), 65(12), 66(1)	3-5	AL: Madison	Mississippian
a. australis	Shelta**	58(4), 60(1), 65(1), 68(1)	3-5	AL: Madison	Mississippian

Table 2. Current ($\theta_\pi \pm$ SE) and Historical-based (θ_W) estimates of genetic diversity and corresponding effective population sizes for obligate cave-dwelling *Orconectes* species and surface-dwelling *Orconectes* species. μ= 2.2*10^-8 substitutions per site per year. Surface-dweller generation time = 2 years, cave-dweller generation time = 10 years.

	Current		Historical	
Cave Species	θ_π	N_e	θ_W	N_e
O.a.packardi	0.00455 ± 0.00043	41364	0.00606	55082
O.incomptus	0.00508 ± 0.00092	46182	0.00477	43375
O.sp.nov.	0.00238 ± 0.00027	21636	0.00242	22034
O.a.australis	0.00894 ± 0.00020	81273	0.01593	144777
Surface Species				
O.juvenilis	0.00394 ± 0.00024	179091	0.03179	1445182
O.luteus	0.02501 ± 0.00015	1136818	0.06076	2761955

Table 3. Genealogical estimates of genetic diversity, descendents/ancestor ratio, and divergence time of four stygobitic *Orconectes* species and the ancestral species for each pairwise comparison estimated by IM. Upper values are the maximum likelihood estimates and the lower values represent the confidence interval range for the 90% highest posterior density. Descendents/Ancestor Ratio = $(\Theta_1 + \Theta_2) / \Theta_{Ancestor}$. A mutation rate of 2.2% per million years was used to determine time since divergence.

	Θ	$\Theta_{Ancestor}$	Descendents/Ancestor Ratio	Time Since Divergence (in millions of years)
Cave Species				
O.a.packardi	14.9498 8.82 - 25.51	16.4765 0.06 - 108.21	1.43	282.5 125.5 - 454.5
O.incomptus	8.5882 2.10 - 27.29			
O.incomptus	8.8723 1.94 - 34.39	12.4212 0.08 - 146.94	1.03	356.1 102.7 - 454.4
O.sp.nov.	3.9714 0.92 - 10.05			
O.sp.nov.	7.3871 5.56 - 9.74	11.5073 5.56 - 52.71	2.76	110.2 105.5 - 116.4
O.a.australis	24.3979 21.87 - 27.11			

Table 4. Results of the nested clade analysis of *Orconectes* 16S mtDNA haplotypes based on 5000 permutations in GeoDIS. Clade (Dc) and nested clade (Dn) distances are given. An 'S' indicates the distance is significantly small at the 5% level and an 'L' indicates the distance is significantly large. In clades with both tip and interior nested clades, the average distance I-T is given. Shaded regions indicate interior groupings.

0-step clades			1-step clades			2-step clades			3-step clades			4-step clades		
Haplotype	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn
1	0	5.6	1-1			2-1	3.7	21.1	3-1	19.2L	18.9	4-1	16.2S	148.7L
2	3.3	3.5												
I-T	3.3	-2.1												
3	3.4	3.8S	1-2	3.8L	3.9L	2-2	3.7S	17.7S						
4	0	4.1												
I-T	3.4	-0.3S												
5	0	4.0	1-3	3.5	3.5S									
6	3.5	3.5												
						I-T	-0.1	-3.4						
7	9.9	9.8	1-4	10.1S	10.3S	2-3	12.5	12.7	3-2	12.6S	13.6S			
8	0	18.8L												
9			1-5	0.0S	22.5L									
10														
11														
			I-T	10.1	-12.2S									
12			1-6			2-4	0.0S	12.1						
13														
14														
15			1-7											
16														
						I-T	12.5L	0.6	I-T	-6.6S	-5.3S			
17			1-8	0.0	10.7	2-5			3-3	9.0S	43.5L	4-2	33.9S	50.4S
19			1-9	0.0	8.3									
18			1-10	0.0S	8.1S									
20														
			I-T	0.0S	-1.0									
22			1-11	0.0	13.6	2-6			3-4	8.4S	34.3			
21	5.6S	10.1	1-12	9.4L	9.5L									
23	4.7S	8.8												
24	1.4	1.6	1-13	1.6S	4.8S									
25	0.0	1.6												
I-T	1.4	0.0	I-T	7.9L	4.1L									
26	0.0	23.1	1-19	21.5	21.6	2-10	20.7S	21.1S	3-6	23.9S	30.7S			
27	18.7	18.7												
28	0.0	0.5												
29	0.0	23.4												
30	0.0	25.4												
I-T	18.4	2.4												
31			1-20	0.0S	27.4									
33														
37			1-24	0.0	24.2									
			I-T	20.1S	-2.2									
32			1-21			2-11	0.0	18.0						
35	5.2	5.4S	1-22	5.6S	5.8	2-12	6.0S	50.5L						
36	0.0	13.0												
I-T	5.2	-7.6S												
34			1-23	0.0	12.7L									
						I-T	15.1L	-27.7S	I-T	9.7	-11.7S			
38			1-32	0.0S	6.2L	2-9	4.2	26.9L	3-7	19.2S	49.1	4-3	47.8S	86.3
39			1-33	0.12	3.1S									

40	6.2	6.2	1-25	6.2	6.2	2-13	6.2S	15.0S							
41	0.0	5.0													
42	0.0	5.0													
44	0.0	5.0													
I-T	6.1	1.2													
43			1-26	0.0	4.9										
45			1-27	0.0	4.9										
			I-T	6.2	1.3										
46	3.5	5.1	1-29	4.6S	4.6S	2-14	6.1S	17.9S	3-8	33.6S	48.2				
47	0.0	3.7													
48	0.0	4.5													
I-T	3.5	0.9													
49	0.0	11.1S	1-28	18.6	17.0										
50	0.0	58.0													
I-T	0.0	-46.9													
51	37.8	35.19	1-31	34.2S	36.5S	2-15	38.6L	36.5L							
53	0.0	22.8													
54	0.8S	31.0													
I-T	37.1L	5.1													
52			1-30	0.0	66.2										
			I-T	34.2S	-29.7S	I-T	-32.5S	-18.5S							
61	8.1	8.1	1-14	8.1S	13.6S	2-7	20.5S	35.4	3-5	35.7S	46.5				
63	0.0	2.8													
64	0.0	15.2													
I-T	8.1	-0.9													
62			1-15	0.0	3.7										
58	0.0	31.2	1-16	26.0	48.5L										
59	0.0	22.2S													
60	0.0	31.2													
I-T	0.0	4.5L	I-T	9.9S	13.5										
57			1-34	0.0	19.1L	2-8	14.1S	36.1							
55	0.0S	5.5	1-18	5.5	12.1										
56	0.0S	5.5													
I-T	0.0	0.0													
65	13.7	13.9	1-17	14.5	15.0										
66	0.0	14.5L													
67	0.0	12.4													
68	0.0	30.0L													
69	0.0	12.4													
I-T	13.8	-3.4S	I-T	-10.8	-0.65	I-T	-6.4	0.7	I-T	-15.7S	2.0	I-T	-5.1	-53.3S	

Table 5. Nested Contingency Results based on 5000 permutations in GeoDIS. * indicates significance at the $p < 0.05$ level. Inferences were made using with Templeton's (2004) revised key. Abbreviations for the inferences are: CRE, contiguous range expansion; RGF, restricted gene flow; IBD, isolation by distance; LDD, long distance dispersal.

Clade	Chi-Square	Probability	Inference Chain	Inferred Pattern
1-2	0.6857	1.0	1-2-11-17-NO	Inconclusive Outcome
1-4	45.00	0.0002*	1-19-20-2-11-17-4-NO	RGF w/IBD
1-12	26.00	0.0*	1-19-20-2-11-12-YES	CRE
1-16	6.00	0.33	1-2-11-17-4-NO	RGF w/IBD
1-17	28.6037	0.17	1-2-11-17-NO	Inconclusive Outcome
1-18	2.000	1.0	1-19-20-NO	Inadequate Geographic Sampling
1-22	1.1953	0.46	1-2-11-17-NO	Inconclusive Outcome
1-28	2.00	1.0	1-19-20-2-11-17-NO	Inconclusive Outcome
1-31	58.9	0.04*	1-2-3-4-NO	RGF w/IBD
2-2	16.354	0.0002*	1-2-11-17-4-NO	RGF w/IBD
2-3	49.0	0.0*	1-19-20-2-11-12-NO	CRE
2-5	16.0	0.0066*	1-19-20-NO	Inadequate Geographic Sampling
2-6	45.385	0.030*	1-2-3-4-NO	RGF w/IBD
2-7	60.093	0.0178*	1-2-3-4-NO	RGF w/IBD
2-8	55.801	0.0136*	1-19-20-2-11-17-4-NO	RGF w/IBD
2-9	7.00	0.05*	1-19-20-NO	Inadequate Sampling
2-10	52.60	0.0174*	1-2-3-4-NO	RGF w/IBD
2-12	1.0588	1.0	1-2-11-17-NO	Inconclusive Outcome
2-14	14.00	0.0238*	1-19-20-2-11-17-4-NO	RGF w/IBD
2-15	27.4909	0.0736	1-2-11-12-NO	CRE
3-1	25.00	0.0*	1-19-20-2-11-12-NO	CRE
3-2	55.13	0.0*	1-2-3-4-NO	RGF w/IBD
3-5	106.27	0.0*	1-2-11-12-NO	CRE
3-6	89.98	0.012*	1-2-3-5-6-7-YES	RGF w/some LDD
3-7	57.00	0.0*	1-19-20-2-11-17-4-NO	RGF w/IBD
3-8	55.63	0.0*	1-2-11-12-NO	CRE
4-1	81.044	0.0*	1-19-20-2-11-12-NO	CRE
4-2	257.00	0.0*	1-19-20-2-11-12-NO	CRE
4-3	442.20	0.0*	1-2-11-12-NO	CRE
Total	851.25	0.0*	1-19-20-2-11-12-NO	CRE

Fig. 1. Geographical distribution (on right) represented by sampled localities for *Orconectes a. packardi* (blue circles), *O. sp. nov.* (green pentagons), and *O. a. australis* (orange squares) along the western escarpment (dark gray shading) of the Cumberland Plateau in Mississippian Age caves at elevations between 180-450 meters. *O. incomptus* (pink triangles) is found in the area just west of the escarpment in Ordovician Age caves at 150-180m in elevation. Phylogenetic relationships (on left) are based on 69 haplotypes of 16S mtDNA sequence data using similar results from Maximum Likelihood and Bayesian methods. Colors marked on tree match cave species colors from distribution map. *Cambarus graysoni* and *C. gentryi* were used as outgroup taxa. Numbers below branches indicate bootstrap support and numbers above branches indicate posterior probabilities.

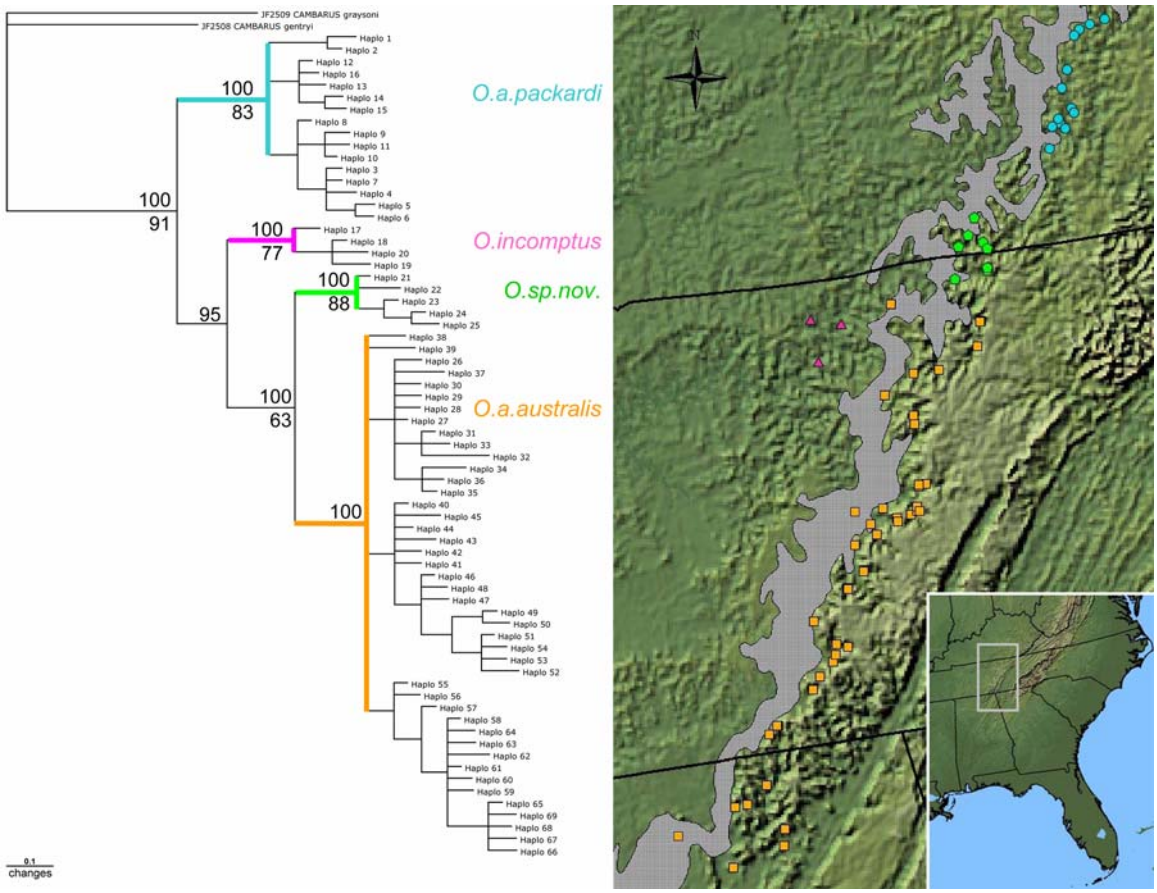


Fig. 2. The marginal posterior probability distributions for the IM model parameter of cave genetic diversity scaled by the neutral mutation rate. Curves are shown for the pairwise analyses of a) *O. a. packardi* (in blue) vs. *O. incomptus* (in pink), b) *O. incomptus* (in pink) vs. *O. sp. nov.* (in green), *and* c) *O. sp. nov.* (in green) vs. *O. a. australis* (in orange) with their corresponding ancestral (in black) diversities.

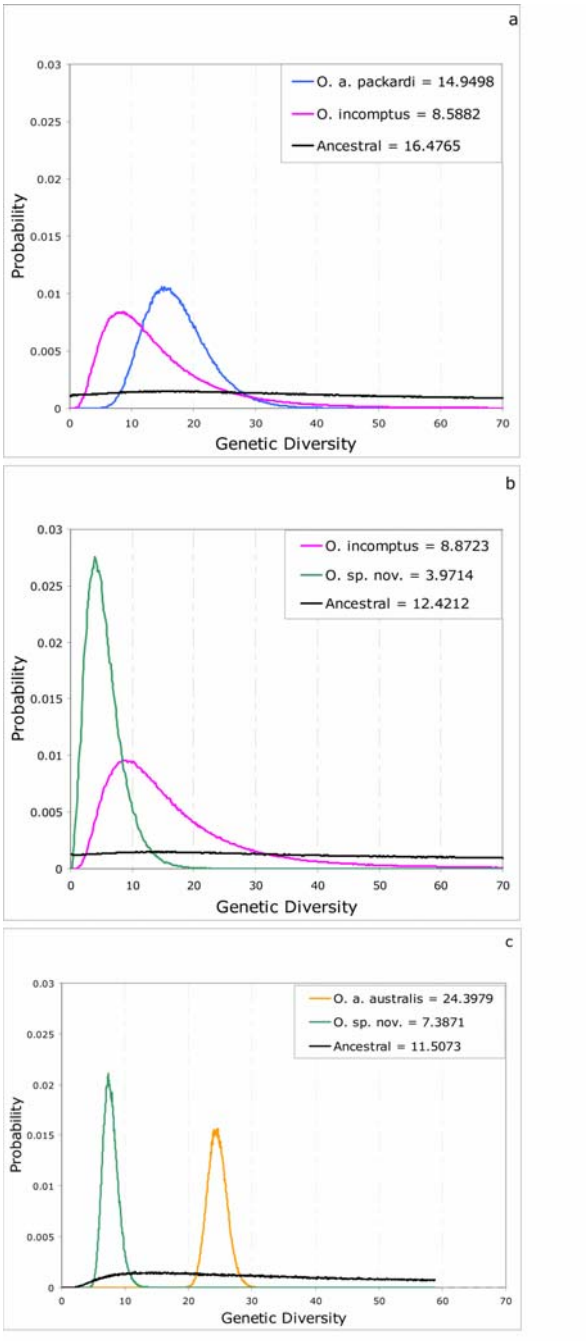
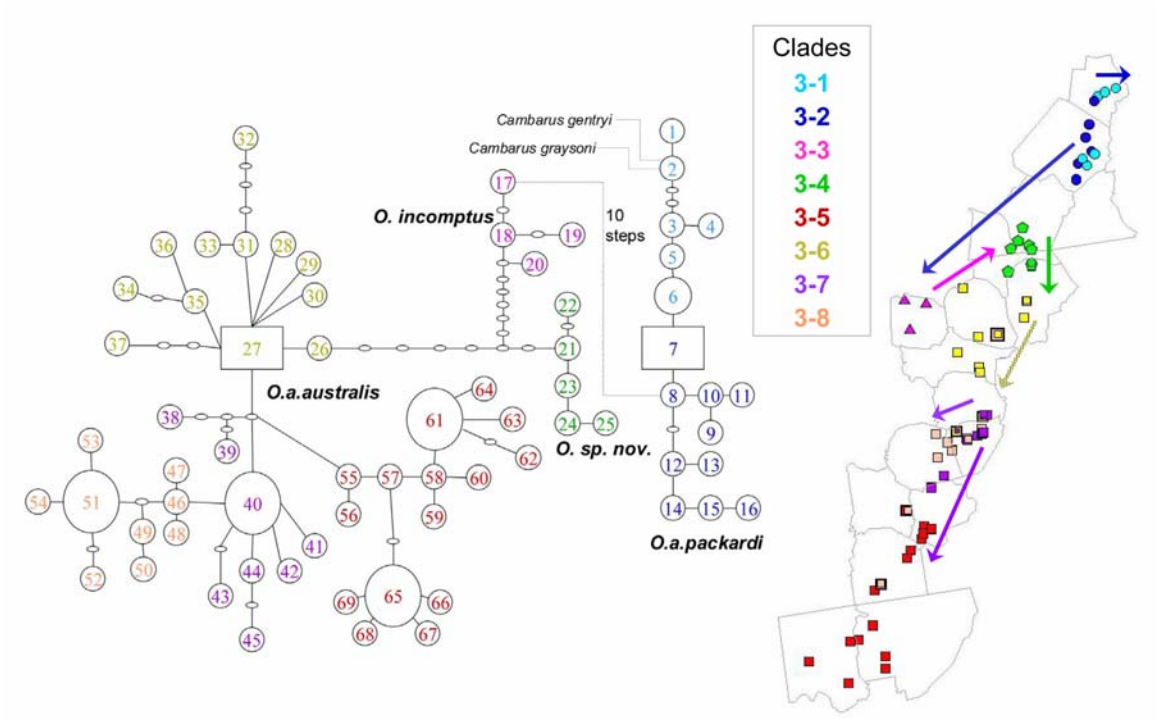


Fig. 3. Haplotype network showing the nesting levels used to infer historical processes. Haplotype circles are colored to represent four distinct lineages: *O. a. packardi* (blue), *O. incomptus* (pink), *O. sp nov.* (green), and *O. a .australis* (orange). Empty circles in the network represent unsampled, possibly extinct haplotypes. The total cladogram includes clades 4-1, 4-2, and 4-3.

Fig. 4. Haplotype network on left is geographically illustrated using the eight 3-step nested clades, which are corresponding marked by the same colors on the map with gray county outlines to the right. *O.a.packardi* (haplotypes 1-16 in network; circles on map) was outside the 95% confidence limit, while *O. incomptus* (haplotypes 17-20 in network; triangles on map), *O. sp. nov*. (haplotypes 21-25 in network; pentagons on map), and *O.a.australis* (haplotypes 26-69; squares on map) were connected within the 95% confidence level. Colored arrows on the dot map of sampled caves show routes of contiguous range expansion by the leading-edge expanding clade. Empty circles in the network represent unsampled, possibly extinct haplotypes. The outgroups *Cambarus gentryi* and *C. graysoni* were outside the 95% limit and connected to haplotype 2 of *O. a. packardi.*



CHAPTER 3

SURFACE TO SUBSURFACE FRESHWATER CONNECTIONS: PHYLOGEOGRAPHIC AND HABITAT ANALYSES OF *CAMBARUS TENEBROSUS*, A FACULTATIVE CAVE-DWELLING CRAYFISH*

ABSTRACT

This study examined the phylogeography and population demographics of *Cambarus tenebrosus,* which has an unusually large distribution for a freshwater crayfish species, encompassing the Interior Lowlands and Cumberland Plateau of the eastern United States. This facultative cave-dweller provides a unique perspective on the biologic connections between surface and subsurface freshwater ecosystems, which are considered to be highly imperiled due to pollution and habitat degradation. The 16S mitochondrial gene was sequenced for 233 individuals from 84 cave and 20 surface locations throughout the range, with most sampling concentrated around the Cumberland Plateau of the Southern Appalachians, to assess conservation status of this species and examine the extent of gene flow between the two habitat types. Cave and surface populations formed a single monophyletic group relative to *Cambarus striatus,* and clades showed strong geographical associations, but lacked habitat structuring. Occupation of subterranean environments does not appear to be a recent event in the evolutionary history of the species. The large amount of genetic diversity within the species, coupled with its ability to inhabit surface and subsurface environments, suggests that this species may pose a threat as a possible invasive species in other karst-dominated landscapes.

***** This study was published as: Finlay, J. B., J. E. Buhay, and K. A. Crandall. 2006. Surface to subsurface freshwater connections: phylogeographic and habitat analyses of *Cambarus tenebrosus,* a facultative cave-dwelling crayfish. Animal Conservation xx, xx-xx.

INTRODUCTION

The number of faunal extinctions occurring in North American freshwater environments has been steadily increasing (Master 1990; Williams *et al.* 1993; Taylor *et al.* 1996; Schuster 1997) and it has been estimated that the number of freshwater species in North America is decreasing at a rate of 4% per decade, which rivals extinction rates in tropical rain forests (Ricciardi & Rasmussen 1999). Elevated extinction rates of freshwater fauna are typically associated with habitat destruction, organic pollution, stream regulation by dams, and habitat fragmentation (Neves *et al.* 1997; Ricciardi *et al.* 1998), yet current research also suggests that genetic factors play important roles in driving threatened and endangered species to extinction (Spielman *et al.* 2004). Thus, it is important that the protection of freshwater environments be approached not only by reducing the impact of humans on the aquatic environment, but also by investigating the population structures and connectivity of its inhabitants using molecular assessments of conservation status.

The Nature Conservancy considers 95% of subterranean species in North America to be endangered or imperiled (Master 1991; Culver *et al.* 2000). There is little doubt that subsurface groundwater fauna are threatened by surface pollution and habitat deterioration (Danielopol *et al.* 2003), but studies of aquatic cave organisms are sparse and often inconclusive, further adding to the enigmatic nature of the subterranean environment. Furthermore, information about the biological connections between surface and subsurface environments is lacking, and this study is the first species-specific genealogical investigation of any North American stygophilic (aquatic facultative cave-dwelling) species.

Cambarus tenebrosus (Hay 1902) is unusual among freshwater crayfish species because it occupies both epigean (surface) and hypogean (subsurface) karst habitats. *C. tenebrosus* also

has a large range for a crayfish, extending from southcentral Indiana southward to northern Alabama (Fig. 1). Since it is found in subterranean habitats typically occupied by obligate cave dwellers (stygobites), it was originally thought that *C. tenebrosus* was a transient member of the cave environment, perhaps being washed into the cave by accident. Hay (1902) refuted this hypothesis based partially on morphological characteristics indicative of stygobitic crayfishes, including the presence of reduced eyes and elongated limbs, which *C. tenebrosus* possesses. These morphological characteristics, collectively referred to as troglomorphy, suggest that *C. tenebrosus* has partially adapted to subterranean life and therefore, is not a passing member of the underground environment. A previous morphological study of *C. tenebrosus* showed no difference between individuals collected from surface and subsurface sites, but reflected overall intra-specific phenotypic plasticity (Taylor 1997). This morphological plasticity might be caused by convergence due to similar environmental pressures in conjunction with active gene flow between the surface and cave habitats (Wiens *et al.* 2003).

The objectives of this project were to: (1) establish whether *Cambarus tenebrosus* shows intra-specific geographic structuring of genetic variation, (2) test if there is a significant genetic association with the two habitats the crayfish occupies (cave versus surface), and (3) provide molecular-based estimates of genetic diversity and effective population size for the species.

MATERIALS AND METHODS

Population Sampling

Samples were collected at 104 sites (84 cave and 20 surface) throughout the range of *Cambarus tenebrosus*, concentrating on areas of the Cumberland Plateau of the Southern Appalachians and the Interior Lowlands which range from southcentral Indiana to northern Alabama (Fig. 1; Table

1). A sample was considered subterranean or "cave" if it was collected from an area not lit by natural light. Samples were included from the type locality at Mammoth Cave in Kentucky. In most cases, a non-destructive method of sampling was used which involved collecting a leg from each individual, then returning the individual to the place of capture. Crayfish have the ability to regenerate lost limbs and therefore, removing a limb during capture is not detrimental to the animal's survival (crayfish often lose their limbs in territorial battles). Tissue samples were stored in 95% ethanol, and each sample was given a unique identification number. Latitude and longitude coordinates were taken by a GPS (Global Positioning System) device at each sample site, including entrances to sampled caves. In few cases, voucher specimens were taken and deposited at the Monte L. Bean Museum (BYU) and the North Carolina State Museum of Natural Sciences. Additionally, *Cambarus striatus* (Hay, 1902), a closely-related species (Buhay & Crandall, unpublished data), was used as the outgroup to root the phylogenetic tree and haplotype network. The network analysis clearly shows C. striatus to be outside the 95% confidence interval for C. tenebrosus and phylogenetic analysis of the genus shows this species to be the sister taxon to C. tenebrosus (J. E. Buhay & K. A. Crandall, unpubl. data), making it an appropriate outgroup for this analysis.

DNA Sequencing

DNA was extracted from the samples using a cell lysis protocol (Crandall *et al.* 1999). The protocol called for 5-15 mg of vacuum-dried tissue to be placed in a tube with 800 µL of cell lysis solution (1.21 g Tris, 37.1 g EDTA, 20 g SDS per litre, pH 8.0). Nine µL of Proteinase K (20mg/mL) was added to this solution and the samples were incubated overnight at 55 ºC while mixed on a shaker for tissue digestion. After 180 µL of 5M NaCl was added, the mixture was

vortexed and centrifuged to pellet out the salt. The supernatant was transferred to a clean cryotube. Immediately after, 420 µL of ice cold isopropanol was added and this mixture was centrifuged at 13000 rpm for 10 minutes to pellet the DNA. After discarding the supernatant, the DNA pellet was washed with 500 µl of 70% ethanol using a cell rotator for one hour. The supernatant was removed and the DNA pellet was vacuum-dried for 15 minutes at 55 ºC, then the pellet was re-suspended in 200 µL of double distilled water.

The 16S mitochondrial gene was sequenced for all samples because it is highly variable and appropriate for population genetic or intraspecific studies (Fetzner & Crandall 2003; Buhay & Crandall 2005). The following reactants were used in each of the 50 µL reactions: 5µL 10x buffer, 8µL dNTP's, 8µL 25mM magnesium chloride, 5µL of each 10mM primer, 0.3 µL Taq Polymerase and 1.5 µL DNA with water added to total 50 µL. Primers used were: 16SF (5' GAC CGT GCK AAG GTA GCA TAA TC 3') and 1472 (5' AGA TAG AAA CCA ACC TGG 3') (Crandall & Fitzpatrick 1996). PCR was performed on a Peltier Thermal Cycler machine (AB9800, Foster City, CA, USA) or a GeneAmp PCR System 9700 (AB9700, Foster City, CA, USA) using the following program: 96 ºC for 3 minutes, followed by 45 cycles of 94 ºC for 1 minute, annealing between 45 and 47 ºC for 1 minute, and 72 ºC for 1 minute, followed by a final elongation at 72 ºC for 5 minutes. PCR products were examined on a 1.5% agarose gel using an ethidium bromide stain. The PCR products were purified using a Montage PCR_{96} plate (Millipore, Billerica, MA, USA). The PCR products were cycle-sequenced using the ABI Big-dye Ready-Reaction kit with 1/4 or 1/8 of the normal reaction size, and sequences were generated on an Applied Biosystems (Foster City, CA, USA) 3730 XL Automated Sequencer at the BYU DNA Sequencing Center. Resulting sequences were edited using Sequencher 4.2 OS X

(Gene Codes Corporation, Ann Arbor, MI, USA) and aligned by eye using MacClade 4.05 OS X (Madison & Madison 2000).

Delimiting Species

Although methods of diagnosing species remain a controversial issue in systematic biology (Sites & Marshall 2003; Sites & Marshall 2004), they are highly relevant to conservation studies (Sites & Crandall 1997) because the method of delimitation can have a significant impact on the number of species diagnosed (Agapow *et al*. 2004). We prefer a statistically-testable method developed for use with molecular data for our study. Templeton's Test of Cohesion (Templeton 1989) uses both historical and current processes to statistically delimit species boundaries through a suite of nested null hypotheses. The hypotheses are then used to determine correlations between genotype and geographic location, habitat, or other ecological variables (Nested Clade Analysis; Templeton *et al.* 1995). Under this definition, two organisms would be considered a single species if they are genetically and/or ecologically exchangeable (Templeton 2001; Rader *et al*. 2005).

Phylogenetic Analysis

The model of evolution that best fit the sequence data was determined using the program ModelTest 3.06 (Posada & Crandall 1998), with the unique 16S haplotypes determined by TCS 1.18 (Clement *et al.* 2000). A Bayesian phylogeny was obtained using MrBayes v3.0b4 (Huelsenbeck & Ronquist 2001; Huelsenbeck *et al.* 2001) with over 20 Markov chains run simultaneously using only unique haplotypes, with each chain initiating at a random tree and parameters nst=6 and rates=adgamma provided by ModelTest. This analysis was run for 20

million generations on 20 processors on a 64-node RackSaver computing cluster, taking samples from the chain every 1,000th tree, totaling 20,001 trees. Using the sampled trees minus the burn-in determined by Tracer (http://evolve.zoo.ox.ac.uk/software.html), a majority-rule consensus tree was constructed. A posterior probability of 95% or greater is considered to be strong Bayesian support for a node (Huelsenbeck & Ronquist 2001).

Genetic - Geographic Associations

Nested Clade Analysis (NCA) allows the partitioning of current population parameters (e.g., recent gene flow) from historical events (e.g., range expansion). NCA is a statistical approach that distinguishes among alternative hypotheses to explain contemporary and historical genetic patterns using haplotype diversity information coupled with geographic location information (Templeton *et al.* 1995; Templeton 1998; Avise 2000). Inferences about genetic patterns can be made by testing a null hypothesis of no association between the collecting locale and the genetic variability (Templeton *et al.* 1995).

To perform a NCA, a haplotype network was constructed using TCS 1.18 (Clement *et al.* 2000) set at a 95% confidence level. The original haplotype network contained several loops, which would be ambiguous in the NCA. These loops were broken using the protocol of Crandall & Templeton (1993) and Templeton & Sing (1993), where number of sequences in a haplotype and geographic location were most heavily considered. The network was then converted into a series of nesting groups (Templeton *et al.* 1987), with the haplotypes exhibiting the highest sequence frequency and most connections being ancestral to the others. According to coalescent theory, those haplotypes found at the tips are more recently evolved than those in the interior of the network (Crandall & Templeton 1993; Templeton 2004).

To test the null hypothesis of no geographical association, two measurements were calculated by the program GeoDis 2.2 (Posada *et al.* 2000). The first is "clade distance" (D_c), which measures the geographical range of a clade at each nested level. Distances were determined by GeoDis using the longitude and latitude coordinates taken at each sample site. Fetzner & Crandall (2003) suggested that for aquatic species, a "river" distance (measuring the distance between two points following only linear water bodies) rather than great circle distance (which uses latitude-longitude coordinates). This approach was not taken for this project because aquatic distances are not known for subterranean basins due to unknown and inaccessible connections. Although the approach used in this study could have some effect on the lower (newer) nesting levels, the higher (older) nesting levels would presumably remain unaffected (Fetzner & Crandall 2003). The second measurement calculated by GeoDis is "nested clade distance" (D_n), which estimates the evolutionary distance between two haplotypes or clades from the center (oldest) nested clade. The output of GeoDis was used to answer a series of dichotomous questions in the NCA inference key (Templeton 2004). These inferences help explain what type of event [such as contiguous range expansion (CRE) or restricted gene flow (RGF)] led to the current haplotype diversity of a species. The most recent version of the GeoDis inference key can be found at: http://darwin.uvigo.es/software/geodis.html.

Genetic - Habitat Associations

GeoDis was used to test for significant associations between genetic and habitat (cave or surface) patterns for clades that include both habitat types. This was done by reducing the number of "locations" in the GeoDis input file to two (cave and surface). These two new "locations" were assigned different coordinates and the test of habitat association (chi-square) was performed over

5000 permutations. This effectively results in a permutation chi-square test as described by Roff & Bentzen (1989).

To test the hypothesis that *C. tenebrosus* is a recent invader of the cave habitat versus a longstanding resident, we used the Fisher's Exact Test to identify significant associations between tip haplotypes (more recent events) and interior haplotypes (older events) for cave and surface habitats in clades with both habitats represented. If the species was a recent invader into the subsurface waters, a significant association would be expected between the cave habitats and the tip locations of the tree. Likewise, if the species was historically located in the cave, but recently invaded the surface waters, a significant association would be observed between the cave and interior clades (or surface and tip clades). If no significant association was found, this would provide evidence for long-term residence in both cave and surface waters.

Demographic Parameters

The current genetic diversity (θ_π) (Tajima 1983) and historical-based genetic diversity (θ_w) (Watterson 1975) were obtained using the computer program DnaSP 4.0 (Rozas *et al.* 2003). Current genetic diversity was computed by pairwise differences between sequences while the historical-based Watterson's θ was determined by the number of segregating sites. These two methods together provide a diversity comparison between current and recent historical diversity of a species for a conservation perspective (Templeton 1993; Yu *et al.* 2003; Buhay & Crandall 2005). Recent losses of diversity (e.g., through selective sweeps or population bottlenecks) would typically show $\theta_\pi < \theta_w$, while recent increases in genetic diversity (e.g., through population growth) would show $\theta_\pi > \theta_w$.

RESULTS

Phylogenetic Analyses

A total of 233 partial 16S (485 base pairs) mitochondrial DNA sequences from 106 collection sites were gathered for *C. tenebrosus*, which included 62 unique haplotypes (Table 1). These haplotypes are accessioned into GenBank as DQ087332-DQ087393. The Bayesian analysis (Fig. 2) revealed that *Cambarus tenebrosus* from both cave and surface habitats formed a monophyletic group relative to *C. striatus* (GenBank DQ087394). The cave and surface populations did not form separate monophyletic groups indicating that there is ongoing gene flow between these two habitats. Additionally, the same haplotype was found in both surface and subsurface habitat types in six instances.

Nested Clade Analysis

Haplotype connections ≤ nine substitutions for the 485 bp of the 16S mitochondrial gene were determined to be part of the 95% confidence set of network connections. All haplotypes were included in a single network created by TCS with exception of haplotypes 54 and 56-62 (Fig. 3). Although these haplotypes were determined to be outside the 95% confidence level (by 13 or fewer mutational steps for every haplotype except 60 which was 19 steps), they were still included in the analysis. Haplotypes 56 and 57 (Indiana cave sites) may have connected to the network had more sampling taken place in northwestern Kentucky and southern Indiana (Fig. 1). *Cambarus striatus* was also outside the 95% confidence level, being ten mutational steps from haplotype 8. The network mostly centered around a single ancestral haplotype (haplotype 1 in Fig. 3) which contained 88 sequences from 42 locations (both cave and surface) found throughout the range of *C. tenebrosus*, excluding Indiana. Nesting of the haplotype network

resulted in 35 one-step clades, 19 two-step clades, 12 three-step clades, four four-step clades and the total cladogram (Fig. 3). The NCA returned 23 significantly large and 27 significantly small associations between genetic variance and geographic location (Table 2).

The NCA revealed significant genetic associations of clades and sampling locations at all clade levels except level two (Table 3). The null hypothesis of no geographic association was rejected at two 1-step clades (1-1 and 1-11), three 3-step clades (3-1, 3-2, and 3-3), all 4-step clades (4-1, 4-2, and 4-3), and the total cladogram. Restricted gene flow (RGF) with isolation by distance (IBD) was inferred for four of the nine significant clades (at 3 and 4-step levels) (Table 3). An inference of restricted gene flow (RGF) with long distance dispersal (LDD) was determined for the total cladogram.

Habitat Association

The habitat association chi-square test revealed no significant association between current genetic patterning and habitat type (cave and surface) except for clades 1-26, 3-2, and 4-2 (Table 4). For the Fisher's Exact Test, we counted 23 cave haplotypes occurring on the tips with 12 interior and 13 surface tip haplotypes and 5 interior. This resulted in no significant association between habitat (cave and surface) and relative age (recent and historical) of the tested haplotypes ($P = 0.76$).

Demographic Parameters

Estimates of the current (θ_π) and recent historical (θ_w) genetic diversity for *C. tenebrosus* are extremely high (Nei 1987) and independent of habitat type (Table 5). These diversity estimates are proportional to the effective population sizes ($\theta = 2 N_{ef}\mu$), suggesting that the number of

breeding individuals is large in both the cave and surface populations. The estimate of effective population size should not be considered a census of the total population of the species, as it only estimates the number of breeding individuals contributing to the gene pool. Interestingly, the recent historical diversity estimates (θ_w) are almost double those of the current diversity estimates (θ_π) (Table 5), showing a sharp decline, nearly 50% loss, in the recent history of the species (Sinclair *et al.* 2002; Yu *et al.* 2003; Buhay & Crandall 2005).

DISCUSSION

The support values for most nodes in the Bayesian topology are markedly low. The polytomies in the tree are not a result of low overall genetic diversity, but rather they are caused by small mutations in the 16S gene that cannot be resolved at the intraspecific level using a phylogenetic approach. However, some deep structure exists in the tree, showing four well supported clades, which mainly cluster according to geography (Fig. 2). Clade I is localized near the eastern border between Alabama and Tennessee along the Cumberland Plateau. All of the haplotypes in Clade II are from the surface sites in central Tennessee. Clade III is localized to southcentral Indiana and could represent a distinct Evolutionarily Significant Unit (ESU), but more sampling is required in this area to support this conclusion. The haplotypes in Clade IV are concentrated along the border separating Alabama and Tennessee to the west of those haplotypes found in Clade I. Clade V is a mixture of both surface and cave populations and span the entire sampled distribution of *C. tenebrosus*, except for Indiana.

Restricted gene flow and contiguous range expansion were inferred for most of the significant phylogeographic patterns within *Cambarus tenebrosus*, particularly in clade 4-2 (Fig. 3), which includes six (of 12 total) of the 3-step clades. This may explain why *C. tenebrosus* is

found across such a large distribution for a freshwater crayfish species. The network was less informative at some nesting levels due to possible short isolation periods, insufficient geographic sampling, or panmixia.

Samples of *C. tenebrosus* from Indiana (haplotype 56 and 57) were separated by twelve steps and haplotype 60 from central Tennessee was 19 steps from the 95% network. With such extensive geographic overlap of the clades particularly in Tennessee and Alabama, it becomes difficult to define boundaries for ESU designation within *Cambarus tenebrosus*. Additional sampling in northern Kentucky and southern Indiana may support the recognition of ESUs or even distinct species which do not overlap geographically with other clades. Our outgroup species, *Cambarus striatus*, fell just outside the 95% confidence limit in the haplotype network at ten mutational steps. Further sampling of *C. striatus* and other closely related species may provide additional insight into phylogenetic relationships with *C. tenebrosus*.

Cambarus tenebrosus appears to have occupied both cave and surface habitats throughout its evolutionary history. This is supported by the presence of haplotypes from both cave and surface habitats situated in the interior of the network. Therefore, rather than an incipient cave species, it appears that *C. tenebrosus* is a long-term inhabitant of caves and associated streams, despite the morphological changes typically associated with the obligate cave-dwelling species.

Despite having a relatively abrupt decrease in genetic diversity in recent history, *C. tenebrosus* still maintains an extremely high level of diversity. This high level of diversity is not surprising considering its unusually large range, its ability to survive in above-ground and below-ground aquatic habitats, and a certain degree of population subdivision amongst the major clades. *Cambarus tenebrosus* is an opportunistic crayfish, occupying almost any freshwater

karst area, including subterranean areas with and without obligate cave-dwelling crayfish species. For the subterranean populations, open habitat increases as the limestone erodes which creates new subterranean spaces and corridors (i.e., connections between two previously separated karst areas). These newly formed groundwater connections provide new habitat over time as well as access to other neighboring gene pools.

Cambarus tenebrosus is a robust species of freshwater crayfish in that it has attained an extremely high level of genetic diversity because it can thrive in two very different yet connected habitats. Important factors in shaping the genetic patterns of aquatic species are climatic fluctuations and glacial events (Graham & Grimm 1990; Vrba 1992; Roy *et al.* 1996). It appears that the cave populations of *C. tenebrosus* have slightly higher historical genetic diversity than surface populations. A higher genetic diversity in the caves may suggest that the underground environment possibly acted as refugia during glacial/interglacial periods when surface waters were in flux between drought during glacial periods and flooding periods during interglacials of south-flowing meltwaters.

Personal observations regarding the troglomorphisms of *C. tenebrosus* indicate that populations in the northern portion of the species' distribution were notably less sensitive to artificial light (e.g., flashlights) in the caves, whereas *C. tenebrosus* in northern Alabama caves were often startled by light and retreated. Moreover, *C. tenebrosus* in the more northern areas were mostly grey or light brown in body color, whereas the southern populations possessed more coloration such as light orange, green, and pink. This might indicate that the crayfish expanded into the southern regions more recently, and have not had time to accumulate fixed troglomorphisms (such as loss of body pigmentation) in the southern populations.

Invasive species are acknowledged as a major economical threat as well as a threat to indigenous species (Vitousek *et al.* 1996; Pimentel *et al.* 2000; Mooney & Cleland 2001) throughout the world. Invasive species are typically genetically diverse (Lee 2002), thus providing a rich pool to draw from to adapt to new surroundings and to out-compete species that occupy a similar niche. *Cambarus tenebrosus* would certainly fall into this description of a potential invasive species due to its high levels of genetic variability and its capacity to thrive in cave and surface environments, particularly karst-dominated areas. Identifying possible invasive species is necessary to protect the overall biological diversity of freshwater systems (Lodge *et al.* 1998). By identifying potential invasive species, precautions can be taken to help avoid their introduction into new areas. Crayfishes are particularly troublesome because they are often used as fish bait and therefore, are easily transferred artificially from one location to another. If this form of unnatural range expansion were to happen with *C. tenebrosus,* it would be especially problematic in both the surface and cave environments.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support for this research from the National Science Foundation (DDIG DEB-0508580) (JEB, KAC), Brigham Young University Graduate School and Integrative Biology Department (JEB), United States Fish and Wildlife Service (JEB), Brigham Young University Undergraduate Honors Program (JBF, KAC), and Brigham Young University Office of Research and Creative Activities (JBF). Comments by Horton Hobbs III, Jack Sites, and an anonymous reviewer were very helpful. We also thank the many members of the National Speleological Society who assisted with the fieldwork, especially Gerald Moni, Nathaniel Mann, and Kevin Toepke.

REFERENCES

Agapow, P. M., Bininda-Emonds, O. R. P., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C. & Purvis, A. (2004). The impact of species concept on biodiversity. *Quarterly Review of Biology* **79**:161–179.

Avise, J. C. (2000). Phylogeography: The History and Formation of Species. Cambridge, MA: Harvard University Press.

Buhay, J. E. & Crandall, K. A. (2005). Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. *Molecular Ecology* **14**:4259-4273.

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

Crandall, K. A., Fetzner, J. W., Jr., Lawler, S. H., Kinnersley, M. & Austin, C. M. (1999). Phylogenetic relationships among the Australian and New Zealand genera of freshwater crayfishes. *Australian Journal of Zoology* **47**: 199-214.

Crandall, K. A. & Templeton, A. R. (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**: 959-969.

Crandall, K. A. & Fitzpatrick, J. F., Jr. (1996). Crayfish molecular systematics: using a combination of procedures to estimate phylogeny. Systematic Biology 45:1-26.

Culver, D. C., Master, L. L., Christman, M. C. & Hobbs, H. H., III. (2000). Obligate cave fauna of the 48 contiguous United States. *Conservation Biology* **14**: 386-401.

Danielopol, D. L., Griebler, C., Gunatilaka, A. & Notenboom, J. (2003). Present and future prospects for groundwater ecosystems. *Environmental Conservation* **30**: 104-130.

Fetzner, J. W., Jr. & Crandall, K. A. (2003). Linear habitats and the nested clade analysis: an empirical evaluation of geographic versus river distances using an Ozark crayfish (Decapoda: Cambaridae). *Evolution* **57**: 2101-2118.

Graham, R. W. & Grimm, E. C. (1990). Effects of global climate change on the patterns of terrestrial biological communities. *Trends in Ecology and Evolution* **5**: 289–292.

Hay, W. P. (1902). Observations on the crustacean fauna of the region about Mammoth Cave, Kentucky. *Proceedings of the United States National Museum* **25**: 223-236.

Huelsenbeck, J. P. & Ronquist, F. (2001). MrBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R. & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**: 2310-2314.

Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* **17**: 386-391.

Lodge, D. M., Stein, R. A., Brown, K. M., Covich, A. P., Bronmark, C., Garvey, J. E. & Klosiewski, S. P. (1998). Predicting impact of freshwater exotic species on native biodiversity: challenges in spatial scaling. *Autralian Journal of Ecology* **23**: 53-67.

Maddison, D. R. & Maddison, W. P. (2000). MacClade 4. Analysis of phylogeny and character evolution. Sunderland, MA: Sinauer Associates.

Master, L. (1990). The imperiled status of North American aquatic animals. *Biodiversity Network News* **3**:1-2, 7-8.

Master, L. (1991). Assessing threats and setting priorities for conservation. *Conservation Biology* **5**: 559-563.

Mooney, H. A. & Cleland, E. E. (2001). The evolutionary impact of invasive species.

Proceedings of the National Academy of Sciences **98**: 5446–5451.

Nei, M. (1987). Molecular Evolutionary Genetics. New York, NY: Columbia University Press.

Neves, R. J., Bogan, A. E., Williams, J. D., Ahlstedt, S. A. & Hartfield, P. W. (1997). Status of aquatic mollusks in the southeastern United States: A downward spiral of diversity. In *Aquatic Fauna in Peril: The Southeastern Perspective*: 43-86. Benz, G. W. & Collins, D. E. (Eds.) Special Publication 1, Southeast Aquatic Research Institute. Decatur, GA: Lenz Design and Communications.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2000). Environmental and economic costs of nonindigenous species in the United States. *Bioscience* **50**: 53-65.

Posada, D. & Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Posada, D., Crandall, K. A. & Templeton, A. R. (2000). GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology* **9**: 487-488.

Rader, R. B., Belk, M. C., Shiozawa, D. K. & Crandall, K.A. (2005). Empirical tests of ecological exchangeability. *Animal Conservation* **8**: 239-247.

Ricciardi, A., Neves, R. J. & Rasmussen, J. B. (1998). Impending extinctions of North American freshwater mussels (Unionoida) following the zebra mussel (*Dreissena polymorpha*) invasion. *Journal of Animal Ecology* **67**: 613-619.

Ricciardi, A. & Rasmussen, J. B. (1999). Extinction rates of North American freshwater fauna. *Conservation Biology* **13**: 1220–22.

Roff, D. A. & Bentzen, P. (1989) The statistical analysis of mitochondrial DNA polymorphisms: Chi-square and the problem of small samples. *Molecular Biology and Evolution* **6:** 539-

545.

Roy, K., Valentine, J. W., Jablonski, D. & Kidwell, S. M. (1996). Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. *Trends in Ecology & Evolution* **11**: 458–563.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

Schuster, G.A. (1997). Resource management of freshwater crustaceans in the Southeastern United States. In *Aquatic Fauna in Peril: The Southeastern Perspective*: 269-282. Benz, G. W. & Collins, D. E. (Eds.) Special Publication 1, Southeast Aquatic Research Institute. Decatur, GA: Lenz Design and Communications.

Sinclair, E. A., Costello, B., Courtenay, J. M. & Crandall, K. A. (2002). Detecting a genetic bottleneck in Gilbert's Potoroo (*Potorous gilbertii*) (Marsupialia: Potoroidae), inferred from microsatellite and mitochondrial DNA sequence data. *Conservation Genetics* **3**: 191-196.

Sites, J. W., Jr. & Crandall, K. A. (1997). Testing species boundaries in biodiversity studies. *Conservation Biology* **11**:1289-1297.

Sites, J. W., Jr. & Marshall, J. C. (2003). Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution* **18**: 462-470.

Sites, J. W., Jr. & Marshall, J. C. (2004). Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, & Systematics* **35**: 199-227.

Spielman, D., Brook, B. W. & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences*

101: 15261-15264.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**:437-460.

Taylor, C. A. (1997). Taxonomic status of the subgenus Erebicambarus, genus *Cambarus*, (Decapoda: Cambaridae) east of the Mississippi. *Journal of Crustacean Biology* **17**: 352-360.

Taylor, C. A., Warren, M. L., Jr., Fitzpatrick, J. F., Jr., Hobbs, H.H., III, Jezerinac, R.F., Pflieger, W.L. & Robison, H. W. (1996). Conservation status of crayfishes of the United States and Canada. *Fisheries* **21**: 25-38.

Templeton, A. R. (1989). The meaning of species and speciation: a genetic perspective. In *Speciation and its consequences*: 3-27. Otte, D. & Endler, J. A. (Eds.). Sunderland, MA: Sinauer Associates.

Templeton, A. R. (1993). The "Eve" Hypothesis: a genetic critique and reanalysis. *American Anthropologist* **95**: 51-72.

Templeton, A. R. (1998). Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**: 381-397.

Templeton, A. R. (2001). Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology* **10**: 779-791.

Templeton, A. R. (2004). Statistical phylogeography: Methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**: 789-809.

Templeton, A. R. & Sing, C. F. (1993). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**: 659–669.

Templeton, A. R., Boerwinkle, E. & Sing, C. F. (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., Routman, E. J. & Phillips, C. A. (1995). Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in *Ambystoma tigrinum*. *Genetics* **140**: 767-782.

Vitousek, P. M., D'Antonio, C. M., Loope, L. L. & Westbrooks, R. (1996). Biological invasions as global environmental change. *American Scientist* **84**: 468-478.

Vrba, E. S. (1992). Mammals as key to evolutionary theory. *Journal of Mammalogy* **73**: 1-28.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256-276

Wiens, J. J., Chippindale, P. T. & Hillis, D. M. (2003). When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. *Systematic Biology* **52**: 501–514.

Williams, J. D., Warren, M. L., Cummings, K. S., Harris, J. L. & Neves, R. J. (1993). Conservation status of freshwater mussels of the United States and Canada. *Fisheries* **18**: 6-22.

Yu, N., Jensen-Seaman, M. I., Chemnick, L., Kidd, J. R., Deinard, A. S., Ryder, O., Kidd, K. K. & Li, W. H. (2003). Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**: 1511-1518.

Table 1. Locations and sample sizes for all *Cambarus tenebrosus* used in this study.

SITE	LOCATION	HABITAT	STATE	COUNTY	SAMPLE SIZE	HAPLOTYPE (# of Individuals)
1	Arthur Singleton	Cave	KY	Rockcastle	2	1(2)
2	Bakers	Cave	TN	Robertson	1	1(1)
3	Bartlett	Cave	TN	Putnam	4	1(3), 35(1)
4	Beaver Creek	Surface	TN	Wayne	1	47(1)
5	Bellamy	Cave	TN	Montgomery	3	1(3)
6	Bible Springs	Cave	TN	Marion	2	31(1), 32(1)
7	Big Bush Creek	Surface	KY	Green	2	1(1), 13(1)
8	Big Sink	Cave	KY	Pulaski	2	1(2)
9	Big Sulphur	Cave	KY	Trigg	2	3(2)
10	Blackpatch Hollow	Cave	TN	Robertson	3	1(2), 4(1)
11	Blind Fish	Cave	TN	Putnam	1	37(1)
12	Bluehole Resurgence	Cave	KY	Rockcastle	4	1(4)
13	Bluff River	Cave	AL	Jackson	2	49(1), 51(1)
14	Boone Hollow	Cave	TN	Clay	2	1(2)
15	Browns Creek	Surface	TN	Davidson	2	21(1), 22(1)
16	Bunkum	Cave	TN	Pickett	2	9(1), 10(1)
17	Camps Gulf	Cave	TN	Van Buren	2	37(1), 40(1)
18	Capshaw	Cave	TN	Putnam	4	37(4)
19	Car Parts	Cave	KY	Rockcastle	2	1(1), 5(1)
20	Cedar Creek	Cave	KY	Pulaski	4	1(4)
21	Cherry	Cave	TN	Jackson	1	11(1)
22	Climax	Cave	KY	Rockcastle	1	14(1)
23	Clinton	Cave	TN	Pickett	4	1(4)
24	Cornstarch	Cave	TN	Fentress	1	5(1)
25	Cummings Cove	Surface	TN	Van Buren	3	37(2), 58(1)
26	Dave's	Cave	KY	Pulaski	1	5(1)
27	Dillions	Cave	IN	Orange	2	56(1), 57(1)
28	Doug Green	Cave	AL	Jackson	1	27(1)
29	Dripping Spring	Cave	TN	Franklin	2	62(2)
30	Dumpling	Cave	KY	Pulaski	2	1(1), 5(1)
31	Dunbar	Cave	TN	Montgomery	1	5(1)
32	Duvalts	Cave	KY	Rockcastle	1	6(1)
33	Edmonson Branch	Surface	TN	Davidson	3	1(2), 46(1)
34	England Cove	Surface	TN	White	5	17(2), 20(1), 37(1), 44(1)
35	Estill Fork	Surface	AL	Jackson	3	33(1), 34(2)
36	Fancher	Cave	TN	Overton	3	1(1), 38(2)
37	Fletcher Spring	Cave	KY	Rockcastle	4	1(2), 7(2)
38	Flynn Creek	Cave	TN	Jackson	1	1(1)
39	Gallatin Steam Plant	Cave	TN	Wilson	1	1(1)
40	Garner Spring	Cave	TN	Franklin	2	33(2)
41	Garretts Mill	Cave	TN	Overton	3	40(3)
42	Grayson Gunner	Cave	KY	Wayne	3	1(3)
43	Hail	Cave	KY	Pulaski	3	1(2), 12(1)
44	Herring	Cave	TN	Rutherford	4	1(3), 8(1)
45	Hester Creek	Surface	AL	Madison	1	59(1)
46	Jared Hollow	Cave	TN	Putnam	3	1(2), 24(1)
47	John Griffin	Cave	KY	Jackson	1	14(1)
48	Kuykendall	Cave	TN	Putnam	12	37(7), 43(5)
49	Larkin Fork	Surface	AL	Jackson	1	33(1)
50	Larkin Spring	Cave	AL	Jackson	3	49(3)
51	Laurel Creek	Cave	TN	Van Buren	1	45(1)
52	Lick Fork	Surface	AL	Jackson	1	61(1)
53	Limrock Blowing	Cave	AL	Jackson	2	49(1), 50(1)

54	Lost Cove	Cave	TN	Franklin	1	33(1)
55	Lost Cove	Cave	TN	Van Buren	1	40(1)
56	Lost Creek	Cave	TN	White	3	1(2), 9(1)
57	Lost River	Cave	KY	Warren	2	14(2)
58	Mammoth	Cave	KY	Edmonson	3	1(2), 3(1)
59	Manning Spring	Cave	TN	Cumberland	1	60(1)
60	Markham	Cave	TN	Clay	1	1(1)
61	Martin Creek	Surface	TN	Putnam	3	9(2), 36(1)
62	McBrides	Cave	AL	Jackson	5	33(3), 49(2)
63	McKinney Pit	Cave	AL	Colbert	1	28(1)
64	Merrybranch	Cave	TN	White	2	1(1), 9(1)
65	Miller	Cave	TN	Warren	3	29(1), 30(1), 52(1)
66	Moore's Spring	Cave	TN	Giles	1	27(1)
67	Mud River	Surface	KY	Logan	2	1(2)
68	Muddy Creek	Surface	KY	Logan	1	1(1)
69	Natural Bridge	Cave	TN	Pickett	1	1(1)
70	North Fork Creek	Surface	TN	Bedford	2	23(2)
71	Norton Spring	Cave	TN	Warren	2	33 (1), 53(1)
72	Pearson Spring	Cave	TN	Franklin	3	26(1), 33(2)
73	Pennywinkle Spring	Cave	TN	Van Buren	1	40(1)
74	Pitman Creek	Surface	KY	Pulaski	1	48(1)
75	Pless	Cave	IN	Lawrence	1	56(1)
76	Pond Cave	Cave	TN	Cannon	1	1(1)
77	Price Valley	Cave	KY	Pulaski	2	1(2)
78	Redmond Creek	Cave	KY	Wayne	3	1(2), 5(1)
79	Richland Creek	Surface	TN	Davidson	2	18(1), 19(1)
80	Roundstone Creek	Surface	KY	Rockcastle	3	1(1), 7(1), 15(1)
81	Rumbling Falls	Cave	TN	Van Buren	2	37(2)
82	Sauta	Cave	AL	Jackson	2	49(2)
83	Sheldon	Cave	AL	Jackson	2	49(2)
84	Short Creek	Cave	KY	Pulaski	2	7(2)
85	Sinking Fork	Surface	KY	Trigg	1	2(1)
86	Skillmans Mark	Cave	TN	Fentress	1	5(1)
87	Spring at Fahey	Cave	TN	Putnam	1	25(1)
88	Spring off Little Creek	Cave	TN	Putnam	3	37(1), 41(1), 42(1)
89	State Trooper	Cave	KY	Warren	3	14(3)
90	Steele Branch	Surface	KY	Trigg	8	1(8)
91	Stout	Cave	TN	Putnam	2	35(1), 39(1)
92	Stream	Cave	KY	Wayne	4	1(2), 5(1), 16(1)
93	Sump Jump	Cave	TN	Robertson	6	1(6)
94	Thorp	Cave	TN	Clay	2	1(2)
95	Tonyas	Cave	KY	Wayne	2	1(2)
96	Trammel Creek	Surface	KY	Allen	1	1(1)
97	Trick or Treat	Cave	TN	Putnam	2	1(2)
98	Turkeyscratch	Cave	TN	Warren	2	33(1), 40(1)
99	Turner	Cave	TN	Houston	1	1(1)
100	Twin Levels	Cave	KY	Christian	1	5(1)
101	Upper Sheep	Cave	TN	White	1	1(1)
102	Waterfall Hollow	Cave	TN	Van Buren	1	37(1)
103	West Cemetery	Cave	TN	Putnam	1	37(1)
104	Winching Hollow Water	Cave	TN	Van Buren	3	37(1), 43(1), 44(1)

Table 2. Results of the nested clade analysis of *Cambarus tenebrosus* 16S mtDNA haplotypes based on 5000 permutations. Clade (Dc) and nested clade (Dn) distances are given. An 'S' indicates the distance is significantly small at the 5% level and an 'L' indicates the distance is significantly large. In clades with both tip and interior nested clades, the average distance I-T is given. Shaded regions indicate interior groupings.

0-step clades			1-step clades			2-step clades			3-step clades			4-step clades		
Haplotype	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn
1	93	93.62	1-1	97.36L	97.24	2-1	95.67	96.23	3-3	98.37	100.14	4-1	95.81S	96.38S
2	0.00	177.66												
3	53.67	151.04L												
4	0.00	105.75												
7	11.84S	135.42												
9	36.95S	67.72												
10	0.00	47.63												
13	0.00	77.22												
I-T	69.18L	-16.79												
11	0.00	29.19S	1-2	43.73	45.88									
12	0.00	87.35L												
I-T	0.00	58.16L	I-T	53.63	51.36									
14	105.03L	105.04	1-3	103.93	105.09S	2-2	105.03	116.12						
15	0.00	90.61												
I-T	105.03L	14.44												
46			1-4	0.00	175.99									
			I-T	103.93	-70.90S									
40	23.27	23.36	1-9	24.23	23.90	2-5	23.56	27.81	3-1	24.62S	52.25S			
45	0.00	16.66												
39	0.00	46.28												
I-T	23.27	-8.11												
52			1-10	0.00	17.99									
			I-T	24.23	5.91									
41			1-12	0.00	23.04	2-6	22.41S	22.85S						
42														
36	0.00	37.36L	1-11	22.37	22.36									
37	21.35	21.40												
38	0.00	28.24												
43	23.51	24.68												
44	8.98	16.16												
I-T	18.22	-3.12	I-T	22.37	-0.68	I-T	1.15	4.96						
47			1-5	0.00	47.76	2-3	63.74	131.49	3-2	105.62	159.68L			
58			1-6	0.00	95.70									
48			1-7	0.00	250.57	2-4	78.58	99.56						
49	12.36	13.02	1-8	12.99S	46.72S									
50	0.00	6.90												
51	0.00	18.84												
I-T	12-36	0.15	I-T	-12.99L	203.84L	I-T	-14.83	31.93	I-T	74.63L	55.11L			
5	103.49	100.27	1-13	79.01	86.06	2-7			3-4	93.77S	102.43S	4-2	128.07L	125.60L
6	0.00	108.48												
I-T	103.49	-8.21												
35	0.64S	1.81	1-14	2.44S	61.05									
24	0.00	3.86L												
I-T	-0.64	2.05												
25	0.00	11.04S	1-15	17.68	74.72									
8	0.00	44.22												
16			1-22	0.00	37.37									
			I-T	64.91	28.66									
23			1-16	0.00	29.83	2-8	46.67	49.35	3-5	45.42S	90.31S			
22			1-17	0.00	48.65									
17			1-19	0.00	86.28									
			I-T	0.00	50.17									
18			1-18			2-9	0.00	35.86						
19														
20			1-20	0.00	85.24L	2-10	48.68	48.42						
21			1-21	0.00	34.06									
			I-T	0.00	51.18	I-T	22.33	7.21						
56	1.99	2.24L	1-34			2-18			3-8	2.24S	269.13			
57	0.00	2.24												
I-T	1.99	0.00L												
54	0.00	35.10	1-33			2-17			3-9	25.67S	158.72			
61	0.00	25.55												
62	0.00	16.37												
I-T	0.00	-13.96												
58			1-32			2-16			3-10	0.00	78.77			
59			1-31			2-15			3-11	0.00	166.96			
									I-T	64.98	-36.50			
26			1-25	0.00	37.69	2-11	58.30	57.31	3-6	54.68	53.87	4-3	54.40S	131.65L
27	37.65S	52.99S	1-24	60.53	60.59									
55	0.00S	75.61												
I-T	37.65	-22.61S	I-T	60.53	22.90									
28			1-23	0.00	147.89	2-12	50.12	53.47						
33	26.76	26.32	1-26	26.77	34.51									
34	0.00	30.51												
I-T	26.76	-4.19				I-T	-8.18	-3.84						
31			1-27			2-13	0.00	29.02	3-7	34.83	57.79			
32			1-28											
30			1-29			2-14	0.00	43.53						
29			1-30											
						I-T	0.00	-14.51	I-T	19.85	-3.92			
60			1-35			2-19			3-12			4-4	0.00	84.02
												I-T	37.47L	25.25L

Table 3. Nested Contingency Results based on 5000 permutations in GeoDis. A "*" indicates significance with a probability of 0.05 or less. Inferences were made using Templeton's (2004) revised key. Abbreviations are as follows: RGF/D=Restricted Gene Flow/Dispersal, IBD=Isolation by Distance, CRE=Contiguous Range Expansion, IS=Inadequate Sampling, IO=Inconclusive Outcome and LDD=Long Distance Dispersal.

Clade	Chi-Square	Probability	Inference Chain	Inferred Pattern
1-1	464.80	0.01*	1-2-3-5-6-7-8-NO	IS
1-2	2.00	1.00	1-19-20-2-11-17-4-NO	RGF with some IBD
1-3	8.00	0.37	1-2-3-4-NO	RGF with some IBD
1-8	11.82	0.45	Nothing Significant	NA
1-9	17.00	0.43	Nothing Significant	NA
1-11	86.92	0.007*	1-2-11-17-NO	IO
1-13	10.00	0.30	Nothing Significant	NA
1-14	3.00	1.00	1-2-3-4-NO	RGF with some IBD
1-15	2.00	1.00	1-19-20-2-11-17-4-NO	RGF with some IBD
1-24	3.00	1.00	1-19-20-2-11-12-NO	CRE
1-26	8.56	0.60	Nothing Significant	NA
1-33	8.00	0.16	Nothing Significant	NA
1-34	0.75	1.00	1-2-11-17-4-NO	RGF with some IBD
2-1	70.65	0.14	1-2-3-4-NO	RGF with some IBD
2-2	8.00	0.34	1-19-20-2-3-5-6-7-YES	RGF/D with LDD
2-3	2.00	1.00	Nothing Significant	NA
2-4	13.00	0.07	1-19-20-2-3-5-6-7-8-NO	IS
2-5	10.00	0.30	Nothing Significant	NA
2-6	22.63	0.10	Nothing Significant	NA
2-7	34.13	0.88	1-2-3-4-NO	RGF with some IBD
2-8	10.00	0.06	Nothing Significant	NA
2-10	2.00	1.00	1-2-11-17-4-NO	RGF with some IBD
2-11	4.00	1.00	Nothing Significant	NA
2-12	14.00	0.34	Nothing Significant	NA
3-1	42.11	0.00*	1-2-3-4-NO	RGF with some IBD
3-2	16.00	0.03*	Nothing Significant	NA
3-3	114.00	0.0002*	Nothing Significant	NA
3-5	10.65	0.17	Nothing Significant	NA
3-6	14.99	0.11	Nothing Significant	NA
3-7	4.00	0.35	Nothing Significant	NA
4-1	346.72	0.00*	1-2-3-4-NO	RGF with some IBD
4-2	170.00	0.00*	1-2-11-12-NO	CRE
4-3	23.00	0.02*	Nothing Significant	NA
Total	612.58	0.00*	1-2-3-5-6-7-YES	RGF/D with LDD

Table 4. Chi-square test of habitat association executed in GeoDis. This test includes only clades with both cave and surface locales. A "*" indicates significance with a probability of 0.05 or less.

Clade	Chi-Square	Probability
1-1	10.682	0.162
1-3	8.000	0.126
1-11	8.957	0.086
1-26	8.556	0.036*
1-33	4.000	0.497
2-1	0.498	1.000
2-2	3.938	0.226
2-4	14.000	0.074
2-6	0.354	1.000
2-12	0.268	1.000
3-1	1.607	0.332
3-2	9.905	0.026*
3-3	0.035	1.000
3-6	0.950	0.569
4-1	1.724	0.452
4-2	30.716	0.000*
4-3	0.726	0.617
Total	6.681	0.074

Table 5. Current (θ_π) and Historical-based (θ_W) estimates of genetic diversity and corresponding effective population size estimates for *Cambarus tenebrosus* (collectively and segregated based on habitat). Effective population sizes were determined using a substitution rate of 2.2% per million years with a generation time of five years (Buhay & Crandall 2005).

	Current		**Historical**	
Cambarus tenebrosus	θπ	Ne	θw	Ne
All samples (n = 233)	0.02359	428910	0.04394	798910
Cave (n = 187)	0.02142	389450	0.04007	728550
Surface (n = 46)	0.02677	486730	0.03501	636550

Figure 1. Outlined range (in orange, adapted from Taylor, 1997) of *Cambarus tenebrosus* extending from central Indiana to northern Alabama. Blue dots represent surface collection sites whereas yellow dots represent cave collection sites.

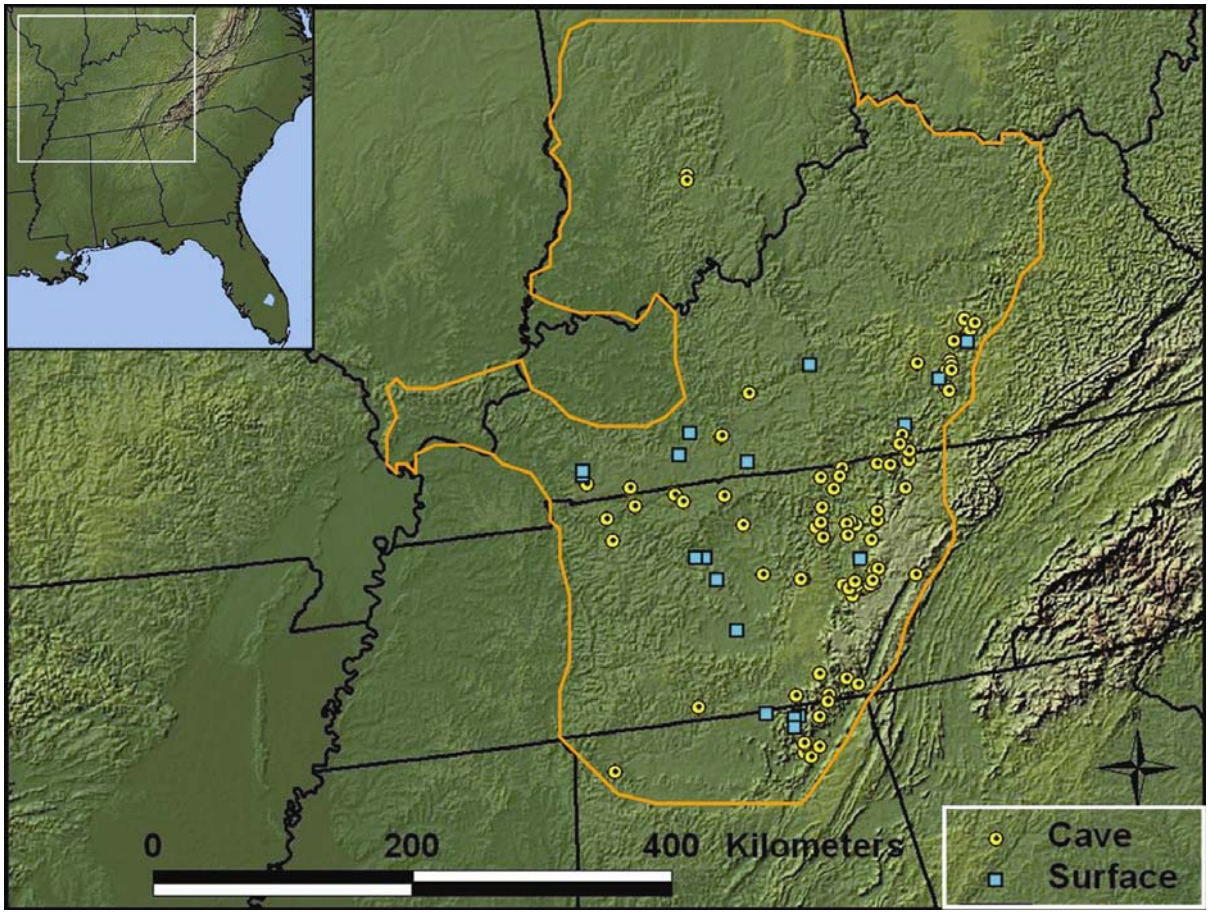


Figure 2. Phylogenetic relationships of 62 *Cambarus tenebrosus* haplotypes of 16S mtDNA sequences. The Bayesian analysis was run using the GTR+I+G (General Time Reversible plus proportional invariant plus gamma) model of evolution determined by ModelTest. The numbers above the branches indicate posterior probabilities. Haplotypes were colored according to habitat (red=cave, blue=surface, and green=both cave and surface). The five main geographical clades are labeled with roman numerals as I: Eastern Cumberland Plateau, II: central Tennessee, III: Indiana, IV: Western Cumberland Plateau, and V: entire sampled range except Indiana. *Cambarus striatus* was used as an outgroup.

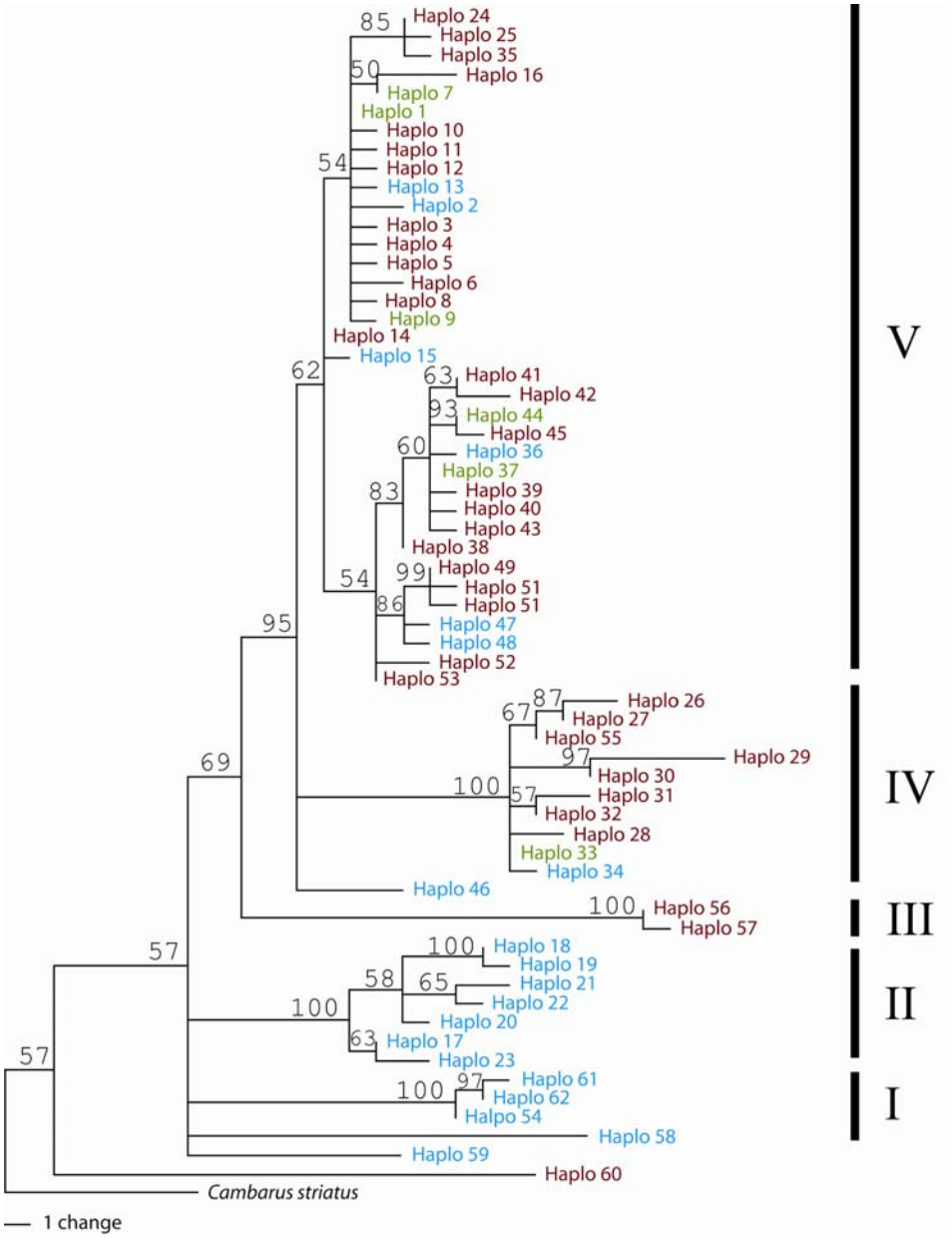
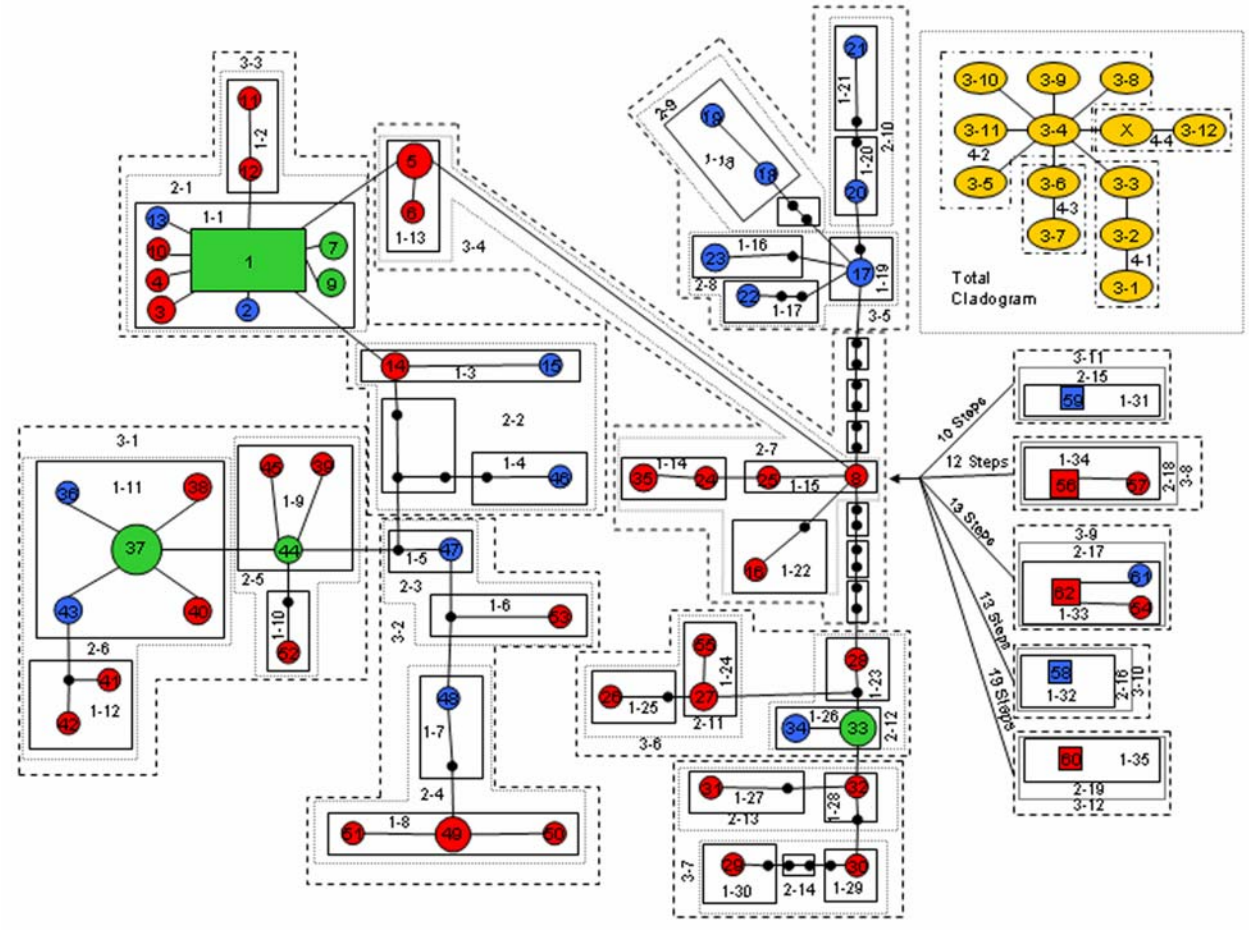


Figure 3. Haplotype network showing the nesting levels used to infer historical processes. Numbers indicate haplotypes (62 total) with black dots representing unsampled or possible extinct haplotypes. The rectangular shape designates the ancestral haplotype for that network. Haplotypes represented by larger numbers of individuals (frequency) are depicted as larger shapes, but the size is not proportional to frequency. Colors correspond to habitat (red=cave, blue=surface, and green=both cave and surface). The total cladogram is shown in orange.



CHAPTER 4

MOLECULAR TAXONOMY IN THE DARK: EVOLUTIONARY HISTORY, PHYLOGEOGRAPHY, AND DIVERSITY OF CAVE CRAYFISH IN THE SUBGENUS *AVITICAMBARUS*, GENUS *CAMBARUS**

ABSTRACT

Freshwater crayfish species in the subgenus *Aviticambarus* (Cambaridae: *Cambarus*) are restricted to caves along the Cumberland Plateau, the Sequatchie Valley, and the Highland Rim which extend along the Tennessee River in southcentral Tennessee and northern Alabama. Historically, three stygobitic species, *Cambarus jonesi, C. hamulatus*, and *C. veitchorum*, comprise this subgenus. We examine species' boundaries and phylogeographic structure of this imperiled Southern Appalachian assemblage to shed light on patterns of cave colonization. We also provide estimates of genetic diversity for conservation status assessment. Using geologic evidence, phylogeographic analyses, and sequence data from five gene regions (two nuclear: Histone H3 and GAPDH and three mitochondrial: 12S, 16S, and CO1 totaling almost 2700 base pairs), we identify two well-supported cryptic species in addition to the three currently recognized taxa. Four of these taxa exhibit low levels of genetic variation both currently and historically, which may indicate local extirpation events associated with geological and river basin changes. Our results also support other recent findings that pre-Pleistocene paleodrainages may best explain the current patterns of aquatic faunal biodiversity in the Southern Appalachians.

---

*This chapter is in press: Buhay, J. E., G. Moni, N. Mann, and K. A. Crandall. 2006. Molecular taxonomy in the dark: evolutionary history, phylogeography, and diversity of the cave crayfish subgenus *Aviticambarus*, genus *Cambarus*. Molecular Phylogenetics and Evolution.

INTRODUCTION

The freshwater crayfish genus *Cambarus* (Erichson, 1846) is one of the largest genera of crayfish in the world, with approximately 100 species (of the 600 global species) and a distribution across the eastern United States. This large genus is comprised of species with varying life history traits including inhabitation of streams, burrows, big rivers, lakes, and caves. Within *Cambarus*, only eleven species are restricted to caves, and these species are distributed in karst (limestone) areas of the Ozarks Plateau, the Greenbrier region of West Virginia, and Cumberland Plateau of the Southern Appalachians (Hobbs and Barr, 1960). In the Southern Appalachians, the subgenus *Aviticambarus* is currently comprised of three obligate cave-dwelling species (stygobites) with ranges restricted to southcentral Tennessee and northern Alabama. This subgenus is a monophyletic group within the genus *Cambarus* (Sinclair et al., 2003; Buhay and Crandall, unpublished data) and each of these groundwater species exhibits troglomorphisms, including albinistic morphology and reduced eyes without pigment.

Subterranean biomes are currently regarded as highly endangered ecosystems, with 95% of obligate cave-dwelling species (including aquatic and terrestrial) in the United States considered to be "vulnerable" or "imperiled" by the Nature Conservancy (Culver et al., 2000). Yet, none of the *Aviticambarus* species were previously evaluated for global extinction risk, and hence, remain unlisted on the IUCN Red List of Threatened Species (World Conservation Union, www.redlist.org). However, all three species are of "High Conservation Concern" according to the state of Alabama's Department of Conservation and Natural Resources, while Taylor et al. (1996) evaluated

national conservation status, considering only *Cambarus veitchorum* to be "Endangered," with *Cambarus hamulatus* and *Cambarus jonesi* regarded as "Currently Stable."

Cambarus hamulatus (Cope and Packard, 1881) inhabits caves in the Sequatchie Valley, Tennessee south to the headwaters of the Black Warrior River, Alabama. In the most recent survey, Hobbs et al. (1977) reported that *Cambarus hamulatus* is known from 22 caves, with the greatest concentration of localities in Jackson County, Alabama and Marion County, Tennessee. The type locale for the species is Nickajack Cave in Marion County, Tennessee which is now flooded due to dams and lakes built on the Tennessee River. *Cambarus hamulatus* has a somewhat patchy distribution, occurring in valleys not inhabited by *Orconectes australis australis*, another obligate cave-dwelling crayfish which ranges along the western escarpment of the Cumberland Plateau (Buhay and Crandall, 2005). The southern portion of the Cumberland Plateau and its western escarpment end in northern Alabama, which makes it difficult to identify whether *Orconectes australis australis* or *Cambarus hamulatus* occurs at a particular cave without capturing males for species-level diagnosis. Both species are known to inhabit caves in northern Alabama, but they have not been found to co-occur in the same cave. Prior to this study, surveys had not been conducted to determine which caves in the northern Alabama mountains are inhabited by *C. hamulatus* and large gaps occur between the 22 reported *C. hamulatus* sites from Hobbs et al. (1977).

Cambarus jonesi (Hobbs and Barr, 1960) was previously recorded from fourteen sites along the Highland Rim region of northern Alabama. This distribution encompasses six counties along both sides of the Tennessee River channel and overlaps with the ranges of other obligate cave crayfish species. Unlike *Cambarus hamulatus*, *C. jonesi* is known

to co-occur with other obligate cave-dwelling crayfish species. *Cambarus jonesi* co-occurs with *Orconectes australis australis* and *Orconectes sheltae* at Shelta Cave in Madison County, which is in the eastern part of the *C. jonesi* range. In the western part of its range, *C. jonesi* co-occurs with *Procambarus pecki.* Additionally, *C. jonesi* is found with *Cambarus veitchorum* (Cooper and Cooper, 1997) at White Spring Cave in Limestone County, which is the only currently known cave site for *C. veitchorum*. *Cambarus veitchorum* was last seen in 1968 and only seven individuals of the species have ever been seen and collected (Cooper and Cooper, 1997). Morphological differences separate *C. jonesi* and *C. veitchorum* at White Spring Cave. *Cambarus veitchorum* is a small species, with the maximum carapace length recorded as 16.7 mm, and the second through fifth tail segments have a spine. *Cambarus jonesi* is the larger species with a maximum carapace length of 28.9 mm and it lacks tail spines (Cooper and Cooper, 1997).

A recent study on the evolutionary history and phylogeography of obligate cave crayfish in the genus *Orconectes* along the western escarpment of the Cumberland Plateau found that current surface drainage patterns are not reflective of the species' boundaries between cave crayfish (Buhay and Crandall, 2005). Moreover, ancient drainage basin events appear to have played major roles in the speciation patterns of other cave animals in the Southern Appalachians as well (spiders: Hedin, 1997a; Hedin, 1997b; Hedin and Wood, 2002; beetles: Barr, 1969; amphipods: Holsinger, 1969), yet, the physical barriers (e.g., ridges or rivers) that once separated the species are no longer apparent or present on the surface (Kane et al., 1992). Therefore, determining species' boundaries and geographic limits for subterranean fauna must be approached using a

thorough sampling scheme across entire distributional ranges and incorporate high resolution genetic data because of the inherent difficulties of relying on a solely morphologically-based taxonomy (Marmonier et al., 1993; Wiens et al., 2003; Finston et al., 2004; Buhay and Crandall, 2005). Appropriately, Proudlove and Wood (2003) in their "Blind Leading the Blind" article called for "DNA taxonomy" to shed light on cryptic subterranean species, particularly for freshwater crustaceans, and to accurately assess biodiversity in the dark which is poorly understood and understudied. Thus, the objectives of our study were to: 1) determine species' boundaries within the cave crayfish subgenus *Aviticambarus* using phylogenetic and geologic information, and 2) assess the phylogeographic structure, genetic diversity, and conservation status of each lineage using inferences from Nested Clade Analysis and information about demographic and historical events.

MATERIALS AND METHODS

Tissue and Data Collection

We collected tissue samples (claw or leg which are regenerated) from 130 cave crayfish individuals from 27 caves spanning the entire previously known ranges of *Cambarus hamulatus* and *C. jonesi*, and we discovered new localities that extended the geographic range of the subgenus (Table 1, Fig. 1). *Cambarus veitchorum* was not found at White Springs Cave (type locality and only known locality for this species), but *C. jonesi* was collected at that locale. Individuals were captured by hand or by using small aquarium nets, and then returned to the place of capture immediately after removing the tissue sample which was placed in a vial containing 95% ethanol. In some cases, whole adult specimens were taken to serve as vouchers for caves added to the distribution after the

most recent listing by Hobbs et al. (1977) (Table 1). Voucher individuals were placed in 95% ethanol and are stored at the Monte L. Bean Museum at Brigham Young University, Provo, Utah.

Genomic DNA was extracted using standard methods and the 16S mtDNA gene, which shows variation within and between crayfish populations (Buhay and Crandall, 2005; Crandall and Fitzpatrick, 1996), was amplified for all sampled individuals (Table 2) during PCR. Two other mitochondrial genes, 12S (Mokady et al., 1999) and CO1 (Folmer et al., 1994) and two nuclear genes, Histone H3 (Colgar et al., 1998) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (M. Schultz, pers. comm.) were amplified for one individual from every sampled cave and three outgroup taxa (*Cambarus gentryi*, *C. brachydactylus*, and *C. friaufi*) (Table 3). Cycle-sequencing reactions were run with purified PCR products (Millipore Montage PCR_{96} plate cleanups) and the Big Dye Ready-Reaction kit on a Perkin Elmer Thermocycler. Reactions were sequenced using an Applied Biosystems 3730 XL automated DNA sequencer. Sequences were edited and aligned by eye using BioEdit (Hall, 1999) and were deposited into GenBank as accession numbers DQ411711-DQ411808 (Table 3). No indels were found in the protein-coding gene sequences.

Phylogenetic Analyses

Unique haplotypes of the 16S gene were analyzed using the Maximum Likelihood approach in PhyML (Guindon and Gascuel, 2003; http://atgc.lirmm.fr/phyml/) and the Bayesian approach in MrBayes v3.04b (Ronquist and Huelsenbeck, 2003) to determine monophyletic lineages and species' relationships. PhyML was run for 500 bootstrap replicates using the Modeltest 3.06 (Posada and Crandall, 1998) parameters: number of

substitution types (nst) = 2, invariable sites (I) = 0, transition/transversion ratio (Tratio) = 2.8493, model = HKY, and gamma distribution (G) = estimated at 0.035 (shape). The Bayesian analysis was run for 20 million generations over 10 chains (9 heated, 1 cold) with nst = 2 and rates = gamma as the starting parameters determined by ModelTest with 1/1000 trees sampled. Tracer (http://evolve.zoo.ox.ac.uk/software/tracer/) was used to determine the burnin and a consensus tree was estimated from the remaining trees. Multiple independent Bayesian and ML runs were performed to ensure convergence on similar results.

Data from the five gene regions were combined into one sequence (totaling 2686 base pairs) representing one individual from every cave sampled along with three outgroups. The Bayesian analysis was run for 20 million generations over 10 chains (9 heated, 1 cold) with nst = 2 and rates = gamma as the starting parameters determined by ModelTest. Every 1000^{th} tree was sampled and the burnin determined by Tracer was discarded. The remaining trees were used to make a consensus tree. PhyML was run for 500 bootstrap replicates using the parameters nst = 2, G = estimated, Tratio = estimated, I = 0.7355 with the HKY model determined by Modeltest and the initial tree determined by Neighbor-joining. Similar topologies and likelihood scores were found with repeated identical runs in both MrBayes and PhyML. The nuclear gene portions showed very little variation and were not analyzed separately, but the combined mtDNA sequences (12S+16S+CO1) were run separately to check for similar results. ModelTest parameters for the various datasets examined are available from the authors upon request.

We consider bootstrap support (BS) 70% and higher and Bayesian posterior probability (PP) 95% and higher to be significant support for a clade (Felsenstein, 1985; Hillis and Bull, 1993; Huelsenbeck and Ronquist, 2001; Wilcox et al., 2002).

Phylogeographic Analyses

We used Nested Clade Analysis (NCA: Templeton et al., 1995; Templeton, 1998; Templeton, 2001) to test for significant associations between geographic and genetic information to elucidate historical and contemporary evolutionary processes and patterns. We first used the program TCS (Clement et al., 2000) with our 16S data to build the haplotype network that illustrates mutational step distances between unique sequences. GEODIS (Posada et al., 2000) was then used to test for significant relationships between geographic locations (cave sites recorded as latitude-longitude coordinates at the entrance) and genetic distances for 5000 random permutations. Clade distances (D_c) represent geographic range for the respective clade level, while nested clade distances (D_n) represent the average distance of samples with a certain haplotype compared to the geographic center of the clade. The 2005 inference key, available from http://darwin.uvigo/es/software/geodis.html, was used to determine which historical processes might have lead to the current evolutionary patterns.

Genetic diversity, effective population sizes, and demographics

Current genetic diversity and recent historical diversity estimates were obtained from the program DNASP 4.0 (Rozas et al., 2003) using 16S sequence data. Current diversity estimates (θ_π; Nei, 1987 equations 10.5 or 10.6 and the standard error, equation 10.7) are based on pairwise base differences between sequences, while historical diversity

estimates (θ_W; Watterson, 1975) are based on the number of segregating sites among the sequences. These two estimates used together provide insight into recent declines or expansions in genetic diversity and effective population sizes ($\theta = 2Ne_{(f)}\mu$ for mitochondrial DNA where $Ne_{(f)}$ = effective population size for maternal lineages and μ = mutation rate) (Templeton, 1993, Yu et al., 2003, Buhay and Crandall, 2005). Using a rate of 2.2 x 10^{-8} substitutions per site per year (Cunningham et al., 1992), effective population sizes were calculated using ten year generation times for obligate cave-dwelling crayfish species and equal sex ratios (Cooper, 1975).

Sample sizes were low for *Cambarus jonesi* (n = 14 from 3 caves), *C. sp. nov. 1* (n = 5 from 1 cave), and *C. sp. nov. 2* (n = 8 from 3 caves) despite extensive fieldwork and range-wide coverage. Therefore, we only examined demographic history for *C. hamulatus* (n = 103 individuals from 20 caves). Tests for neutrality can be used to assess demographic history with significant negative *D* values of Tajima (1989) and *F** values of Fu and Li (1993) indicating population expansions. We also performed a mismatch analysis (which plots the distribution of the number of differences between pairs of haplotypes) for population growth (expansion) for *C. hamulatus* in DNASP (Rozas et al., 2003) using an initial $\theta = 0$, final $\theta = 1000$, with expansion parameter $\tau = 2\mu t = 3.803$. Population expansion would appear as a "wave" in the mismatch distribution, while stable population sizes produce ragged multi-modal distributions (Rogers and Harpending, 1992; Harpending, 1994). The probability of obtaining values of r (raggedness) less than the observed (P ($r_{expected} < r_{observed}$)) was calculated using the coalescent algorithm in DNASP over 1000 pseudoreplications with a random seed and no recombination.

RESULTS

Phylogenetic analyses

Phylogenetic relationships among the cave crayfish species of the subgenus *Aviticambarus* were determined using Bayesian and Maximum Likelihood approaches for both the 16S haplotype dataset and the combined five gene dataset representing one individual from each sampled cave. For each of the Bayesian analyses, the first 2000 trees were discarded as burnin and the consensus tree was estimated using the remaining 18000 trees.

Rather than two extant species (*Cambarus hamulatus* and *C. jonesi*), we found evidence of four distinct lineages (Figs. 1-3) in addition to the unsampled *C. veitchorum*. Using only 16S haplotypes (GenBank DQ411734-DQ411759), *Cambarus sp. nov. 2* (endemic to Marshall Co, Alabama) is sister to the other lineages with significant support (Fig. 2: 100% PP, 97% BS). *Cambarus hamulatus* diverged from a common ancestor with *C. jonesi* and *C. sp. nov. 1* with 100% BS and PP support for the node. The distinctiveness of *C. jonesi* and *C. sp. nov. 1* was highly supported (100% PP and 99% BS and 100/100%, respectively).

The combined five gene dataset was analyzed using both ML and Bayesian approaches and revealed some similar trends to the 16S haplotype analysis. *Cambarus sp. nov. 2* was recovered as basal to the other cave lineages with significant BS and PP support for the node (Fig. 3). In contrast to the haplotype analysis, the node separating *C. jonesi* from *C. hamulatus* and *C. sp. nov. 1* was significantly supported with both BS and PP, while the sister relationship between *C. sp. nov. 1* and *C. hamulatus* was not highly

supported. This same topology was recovered using only mtDNA combined sequence data of the three genes for each cave with similar nodal support values (not shown).

Nested Clade Analysis

The statistical parsimony network included 26 unique haplotypes within the subgenus *Aviticambarus* (Fig. 4). The Marshall County, Alabama (*Cambarus sp. nov. 2*) samples fell out as a separate network (higher-level clades are marked with an 'A' and the network is shaded orange in Fig. 4) connected to haplotype 4 of *Cambarus hamulatus* by 23 steps. Therefore, the main network contained *C. hamulatus* and *C. jonesi* connected by nine steps (= 95% confidence limit), but the *C. sp. nov. 1* haplotypes were outside the limit with twelve steps. These three lineages were grouped together for the phylogeographic analysis (main network) and a separate analysis was done on the Marshall County network (Fig. 4: *C. sp. nov. 2*, shaded in orange).

The main network contained twelve 1-step clades, seven 2-step clades, and four 3-steps clades in the total cladogram while the Marshall County network contained two 1-step clades and two 2-step clades in the total cladogram (Table 4). The outgroups *Cambarus gentryi* and *C. brachydactylus* were connected to Haplotype 24 by 31 steps and *C. friaufi* was connected by 28 mutational steps to Haplotype 9 (Fig. 4). Eight clades differed significantly from random distributions ($p < 0.05$; Table 5). All clades with significantly large or small distances were examined using Templeton's 2005 inference key to elucidate historical and current processes which contribute to the genetic structuring of each lineage. Contiguous range expansion, long-distance colonization, past fragmentation, restricted gene flow, and isolation by distance were the inferred patterns.

Species' Boundaries

Delineation of species' boundaries is a hotly-debated issue for systematists and conservation biologists, since species are the fundamental units of biodiversity (Sites and Crandall, 1997; Agapow et al., 2004). We chose to define species with criteria specified for the Genealogical Concordance Species concept (Avise and Ball, 1990; Baum and Shaw, 1995), since we are investigating multiple independent characters (genetic; geographic; geologic). This lineage-based concept necessitates concordance among different characters and defines a "genealogical species" as a group of organisms more closely related to each other ("exclusivity") than to organisms outside its group (Baum and Shaw, 1995).

Phylogenetic analyses of 16S haplotypes (Fig. 2) revealed significant support for the monophyly of *Cambarus hamulatus* (75% BS), *C. jonesi* (100/99% PP/BS), *C. sp. nov. 1* (100/100% PP/BS) and *C. sp. nov. 2* (100/100% PP/BS). The combined gene analyses also supported the distinction of *C. hamulatus*, *C. jonesi, C. sp. nov. 1,* and *C. sp. nov. 2* as separate lineages (Fig. 3), although the sister relationship of *C. sp. nov. 1* lacked significant support. Nested phylogeographic analyses supported the recognition of three separate lineages: *C. hamulatus/C. jonesi* which was grouped at the 95% confidence limit, *C. sp. nov. 1*, and *C. sp. nov. 2* (Fig. 4, Table 5).

Geologic evidence (province and district separations provided by the Geological Society of Alabama, S. McGregor) supported the recognition of each of the four lineages. *C. hamulatus* is restricted to the area around the Sequatchie Valley and the Jackson County, Alabama mountains district of the Cumberland Plateau province in Jackson County, while *C. jonesi* is endemic to the Tennessee Valley district of the Highland Rim

province (Fig. 1). *C. sp. nov. 1* is found in the Moulton Valley district of the Highland Rim province, while *C. sp. nov. 2* is restricted to the Jackson County Mountain district of the Cumberland Plateau in Marshall County, Alabama. Based on concordance of multiple characters and exclusivity, there appears to be no less than five distinct genealogical species in the subgenus *Aviticambarus*. Although we chose to employ the Genealogical Concordance Species Concept a priori, we also recognize that these results fit nicely with criteria of the Phylogenetic Species Concept (de Queiroz and Donoghue, 1990), including monophyly and "exclusivity" (Baum and Donoghue, 1995). Likewise, the species are supported by the Cohesion Concept using the exchangeability criteria of Templeton (2001).

Genetic Diversity and Demography

Estimates of genetic variability are reported in Table 6 and the four lineages show low to moderate levels of diversity. These low levels of diversity are common among species thought to have undergone a bottleneck, but both current and historical estimates of population size are similar. Examining deviations from neutrality can help clarify past demographic events, as significant negative *D* (Tajima, 1989) and *F** (Fu and Li, 1993) values are often associated with bottlenecks followed by range expansions. In the case of *C. hamulatus*, using 16S haplotypes (n=12), we found $D = -0.1814$ ($P > 0.10$) and $F^* = -0.52837$ ($P > 0.10$). A unimodal mismatch distribution is predicted for populations having undergone expansion (which is indicated by the expected curve in Fig. 5), but our observed distribution shows a slightly ragged bimodal distribution, typical of constant population sizes, not growth (raggedness = 0.036; $P (r_{expected} < r_{observed}) = 0.15$); Rogers and Harpending, 1992; Harpending, 1994).

Conservation Status Assessment

Using categories and criteria to evaluate species for endangerment, we suggest that all five *Aviticambarus* lineages be considered for conservation measures and listing on the IUCN Red List (version 3.1 criteria found on www.redlist.org). According to the Preamble of the 2001 Categories and Criteria, the IUCN affords protection to "species or lower taxonomic levels, including forms that are not yet formally described" (www.redlist.org), and therefore, we recommend conservation status for each species based on the information available.

Cambarus veitchorum should receive the highest protection, critically endangered (CR), as only a total of seven individuals (six adults, 1 juvenile) have ever been documented from White Spring Cave which is the only known locale for the species despite search efforts by many biologists for the past three decades since the last sighting in 1968. *C. veitchorum* meets the CR category with the criteria of decline in occurrence, extent of occurrence estimated to be less than a 100 sq. km. area, only known from a single location (area of occupancy), and inferred decline in number of mature individuals (IUCN A2c, B1a, B1bi-v, B2a, C2i, D).

Cambarus jonesi is currently known from only twelve locations along both sides of the Tennessee River basin in northern Alabama. We recommend this species receive vulnerable (VU) status, as it meets criteria of a geographic extent of occurrence less than 20,000 sq. km., severely fragmented range, area of occupancy less than 2,000 sq. km, and inferred decline in the quality of the cave habitat (IUCN B1a, B1biii, B2biii).

Cambarus sp. nov. 1 is currently only known from one cave locality in Alabama, with five individuals found, and we suggest that this species be afforded vulnerable (VU) status. The cave locality occurs on a major interstate highway, as the cave entrance was blasted open by the road construction. We feel that intensive field surveys might find a few new localities of this species, but the known caves in the vicinity have been extensively searched for cave crayfish for decades. This species meets the same vulnerable status criteria as *C. jonesi,* with additional criteria of a very small population size and a very restricted area of occupancy (IUCN B1a, B1biii, B2biii, D2).

Cambarus sp. nov. 2 is currently only known from four cave localities in Marshall County, Alabama. Extensive fieldwork was conducted in the direct vicinity of the four known cave locations for this species, and we feel additional surveys might result in only a few more locations at best. We suggest that this species be considered vulnerable (VU) because it meets the criteria of extent of occurrence less than 20,000 sq. km., known to exist at less than ten locales, area of occupancy less than 2,000 sq. km., and very small and restricted populations (IUCN B1a, B2a, D2).

Cambarus hamulatus is the most widespread species of the *Aviticambarus* assemblage, but the bulk of the known cave localities are clustered around the geographic center of its range. This species does not meet the criteria for vulnerable status, but we feel that its fragmented distribution coupled with a population size that does not appear to be expanding qualifies this species for "near threatened" (NT) status.

DISCUSSION

Phylogeographic studies on the faunal biodiversity of the Southern Appalachian Mountains, an area among the highest in species richness in North America, are increasing in the literature for various animal groups (e.g., salamanders: Rissler and Taylor, 2003; Crespi et al., 2003; Kozak et al., 2006; Jones et al., 2006, spiders: Hedin, 1997a, 1997b, Hedin and Wood, 2002, fish: Berendzen et al., 2003, insects: Schultheis et al., 2002, crayfish: Buhay and Crandall, 2005). The Cumberland Plateau and adjacent Highland Rim that surround the Nashville Basin, are disjunct topographic karst units off the main Southern Appalachian chain, and these areas also rank high for species richness in endemic fauna, particularly for freshwater mussels, snails, and crayfish. This area of the Southern Appalachians also lends itself to limestone (karst) cave development and ranks among the highest in cave density in the world with nearly 5000 caves located on the Cumberland Plateau province and 3500 caves on the Highland Rim (Aulenbach and Cressler, 1998). Following with this pattern of high cave density, Culver et al. (2000) found that the area of greatest diversity for terrestrial cave animals (troglobites) in the United States was the northeastern corner of Alabama, including Jackson, Madison, and Marshall Counties which is also the geographic center of the range of the cave crayfish subgenus *Aviticambarus*.

Over half of the animal species on the United States Natural Heritage List of imperiled and vulnerable taxa (http://www.natureserve.org) are terrestrial and aquatic subterranean species, which brings to light the need for science-based conservation assessments, status surveys, and "DNA taxonomy" to separate morphologically-cryptic taxa. Our study revealed two new cave crayfish lineages previously assigned to

Cambarus jonesi and found that all members of the subgenus *Aviticambarus* are indeed imperiled due to low genetic diversity and population size estimates, small geographic ranges, and few known locales. *Cambarus hamulatus*, a cave crayfish species that spans seven counties in two states, showed stable, not expanding, populations with only a moderate level of genetic diversity. These results are in stark contrast to another cave crayfish assemblage (genus *Orconectes*, subgenus *Orconectes*) on the western escarpment of the Cumberland Plateau, which showed moderate to high levels of genetic diversity, larger ranges, more known locales, and extensive gene flow (Buhay and Crandall, 2005).

The complex, dynamic geologic and hydrologic history of the Sequatchie Valley appears to have played major roles in the distribution and current population structure of *Cambarus hamulatus*. Run to the Mill Cave in Cumberland County, Tennessee (Fig.1: northernmost locale) is a massive groundwater system at the head of the Sequatchie Valley and Sequatchie River, which flows south directly into the Tennessee River. On the extreme southern end of the Sequatchie Valley is Rickwood Caverns in Blount County, Alabama (Fig. 1: southernmost locale), which currently drains south into the Black Warrior River of the Mobile Basin. But the next southernmost site was Graves Cave in Blount County, Alabama (Fig. 4: Haplotypes 1, 2, 3) which was colonized during a different migration episode than its southern neighbor Rickwood Caverns (Fig. 4: Haplotype 12). Perhaps, as the southern portion of the Sequatchie Valley was widened by erosion and geologic activity, separate colonization events (wash-outs or long-distance migrations) expanded the range of *C. hamulatus*. The Sequatchie Valley (anticline) was formed by a geological uplift which has since weathered and eroded down to the current

valley floor (Thomas, 1986), leaving a wide area of limestone (including Mississippian Limestone strata) exposed in the valley floor and slopes of the Sequatchie Valley. However, cave development along the Sequatchie Valley is limited due to the complex hydrogeologic history of the area, which may also have prohibited accumulation of genetic diversity or prevented population growth in *C. hamulatus*. It is interesting to note that while *Orconectes australis australis* migrated mainly southward in leading-edge expansion events along the Cumberland Plateau's western escarpment accumulating genetic diversity along the way (Buhay and Crandall, 2005), *C. hamulatus* originated in the center of its current range and expanded in both northward and southward directions from the area of the Alabama-Tennessee state line along the eroding Sequatchie Anticline.

Finally, this study supports previous findings that paleodrainages, specifically pre-Pleistocene water routes, played important roles in phylogeography and speciation processes of freshwater fauna in the southeastern United States. Ancient drainage changes may have lead to the extirpation of populations, which may be reflected as unsampled, possibly extinct haplotypes in the parsimony network. Current drainage basins do not reflect species' geographic boundaries or barriers between cave crayfish in the Southern Appalachians and even more importantly, biologists may be drastically underestimating biodiversity by relying on contemporary hydrologic delineations, physiographic boundaries, and convergent similar morphology. For example, a recent extensively-sampled study of the *Eurycea bislineata* complex in eastern North America identified 13 putative independent lineages rather than five taxonomically-recognized salamander taxa within the *bislineata* complex (Kozak et al., 2006). Our study supports

five distinct lineages rather than three within the subgenus *Aviticambarus*, and highlights the need for integration of genetic, phylogeographic, and environmental (hydrological and geological) analyses in well-sampled studies of freshwater fauna, particularly obligate cave-dwellers, to tease apart convergence and taxonomy and afford conservation and protection to these unique organisms.

ACKNOWLEDGEMENTS

Thank you to all the cavers of the Alabama and Tennessee grottos of the National Speleological Society, especially the Drake and Caudle families and Marion O. Smith and Jason Record, who provided new cave crayfish localities. Special thanks are due to Jim Godwin and Carl Scardina for permission to sample Rickwood Caverns State Park. This study would be impossible without fantastic field assistants, especially Heather Levy, Joe Douglas, Jonathan Griffith, Kristin Bobo, Kevin Toepke, Daniel Mann, Avis Moni, Lin Guy, Kenneth Rupil, Matt Niemiller, and some great landowners, including Mr. Wilson, the keeper of Run to the Mill Cave. Thank you to Stuart McGregor for help with geological provinces in Alabama, to Michael Gilbert for knowledge about Rickwood Caverns, and to Mark Schultz who provided primer sequences for the GAPDH gene. John Cooper, Tom Barr, Brice Noonan, and two anonymous reviewers kindly commented on earlier drafts which greatly improved this manuscript. This study was funded by NSF DDIG (DEB 0508580) to JEB and KAC and by the Cave Conservancy Foundation and Brigham Young University Graduate School and Integrative Biology Department to JEB.

REFERENCES

Agapow, P.-M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M.,

Marshall, J.C., Purvis, A., 2004 . The impact of species concept on biodiversity studies. Quarterly Rev. Biol. 79, 161-179.

Aulenbach, B.T., Cressler, A., 1998. Tennessee, Alabama, Georgia Cave Distribution. In: Schindel, G.M., Hickman, J.L. (Eds.), Journeys through TAG: 1998 NSS Convention Guidebook. 2 pp.

Avise, J.C., Ball, R.M., 1990. Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7, 43-67.

Barr, T.C., Jr., 1969. Evolution of the (Coleoptera) Carabidae in the Southern Appalachians. In: Holt, P.C., Hoffman, R.L., Hart, C.W. (Eds.), The Distributional History of the Biota of the Southern Appalachians. Part I: Invertebrates. Research Division Monograph. Blacksburg: Virginia Polytechnic Institute, pp. 67-92.

Baum, D.A., Donoghue, M.J., 1995. Choosing among alternative phylogenetic species concepts. Syst. Bot. 20, 560-573.

Baum, D.A., Shaw, K.L., 1995. Genealogical perspectives on the species problem. In: Hoch, P.C., Stephenson, A.G. (Eds.), Experimental and Molecular Approaches to Plant Biosystematics. St. Louis: Missouri Botanical Garden, pp. 289-303.

Berendzen, P.B., Simons, A. M., Wood, R.M., 2003. Phylogeography of the northern hogsucker, *Hypentelium nigrican*s (Teleostei: Cypriniformes): genetic evidence for the existence of the ancient Teays River. J. Biogeography 30, 1139-1152.

Buhay, J.E., Crandall, K.A., 2005. Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. Mol. Ecol. 14, 4259-4273.

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

Colgar, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D., Macaranas, J., Cassis, G., Gray, M.R., 1998. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. Aust. J. Zool. 46:419-437.

Cooper, J.E., 1975. Ecological and behavioral studies in Shelta Cave, Alabama with emphasis on decapod crustaceans. PhD dissertation, University of Kentucky.

Cooper, J.E., Cooper, M.R., 1997. A new species of troglobitic crayfish of the genus *Cambarus*, subgenus *Aviticambarus* (Decapoda: Cambaridae), endemic to White Spring Cave, Alabama. Proc. Biol. Soc. Washington 110, 608-616.

Cope, E.D., Packardi, A.S., 1881. The fauna of Nickajack Cave. Amer. Nat. 15, 881-882, 877-882.

Crandall, K.A., Fitzpatrick, J.F., Jr., 1996. Crayfish molecular systematics: using a combination of procedures to estimate phylogeny. Syst. Biol. 45, 1-26.

Crespi, E.J., Rissler, L.J., Browne, R.A., 2003. Testing Pleistocene refugia theory: Phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. Mol. Ecol. 12, 969-984.

Culver, D.C., Master, L.L., Christman, M.C., Hobbs, H.H., III, 2000. Obligate cave fauna of the 48 contiguous United States. Cons. Biol. 14, 386-401.

Cunningham, C.W., Blackstone, N.W., Buss, L.W., 1992. Evolution of king crabs from hermit crabs ancestors. Nature 355, 539-542.

de Queiroz, K., Donoghue, M.J., 1990. Phylogenetic systematics and species revisited. Cladistics 6, 83-90.

Erichson, W.F., 1846. Uebersicht der Arten der Gattung Astacus. Archiv fur Naturgeschichte (Berlin) 12, 86-103.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368-376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Finston, T.L., Johnson, M.S., 2004. Geographic patterns of genetic diversity in subterranean amphipods of the Pilbara, Western Australia. Marine and Freshwater Research 55, 619–628.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Marine Biol. Biotech. 3, 294-299.

Fu, X.-Y., Li, W.-H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693-709.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696-704.

Hall, T.A., 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95-98.

Harpending, H.C., 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. 66, 591-600.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160-174.

Hedin, M.C., 1997a. Molecular phylogenetics at the population / species interface in cave spiders of the southern Appalachians (Araneae: Nesticidae: *Nesticus*). Mol. Biol. Evol. 14, 309-324.

Hedin, M.C., 1997b. Speciational history in a diverse clade of habitat-specialized spiders (Araneae: Nesticidae: *Nesticus*): Inferences from geographic-based sampling. Evolution 51, 1927-1943.

Hedin, M.C., Wood, D.L., 2002. Genealogical exclusivity in geographically proximate populations of *Hypochilus thorelli* Marx (Araneae, Hypochilidae) on the Cumberland Plateau of North America. Mol. Ecol. 11,1975-1988.

Hillis, D.M., Bull J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42,182-192.

Hobbs, H.H., Jr., Barr, T.C., Jr., 1960. The origins and affinities of the troglobitic crayfishes of North America (Decapoda, Astacidea). I. The genus *Cambarus*. Am. Mid. Nat. 64, 12-33.

Hobbs, H.H., Jr, Hobbs, H.H., III, Daniel, M.A., 1977. A review of the troglobitic decapod crustaceans of the Americas. Smithsonian Contributions to Zoology 244, 1–176.

Holsinger, J.R., 1969. Biogeography of the freshwater amphipod crustaceans (Gammaridae) of the central and southern Appalachians. In: Holt, P.C., Hoffman, R.L., Hart, C.W. (Eds.), The Distributional History of the Biota of the Southern Appalachians. Part I: Invertebrates. Research Division Monograph. Blacksburg: Virginia Polytechnic Institute, pp. 19-50.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny.

Bioinformatics 17, 754-755.

Jones, M.T., Voss, S.R., Ptacek, M.B., Weisrock, D.W., Tonkyn, D.W., 2006. River drainages and phylogeography: an evolutionary significant lineage of shovel-nosed salamander (Desmonathus marmoratus) in the southern Appalachians. Mol. Phylogenet. Evol. 38, 280-287.

Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H. N. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, USA, pp. 21-132.

Kane, T.C, Barr, T.C., Jr., Badaracca, W.J., 1992. Cave beetle genetics: geology and gene flow. Heredity 68, 277-286.

Kozak, K.H., Blaine, R.A., Larson, A., 2006. Gene lineages and eastern North America palaeodrainage basins: phylogeography and speciation in salamanders of the *Eurycea bislineata* species complex. Mol. Ecol. 15, 191-207.

Marmonier, P., Vervier, P., Gibert, J., Dole-Olivier, M.-J., 1993. Biodiversity in groundwaters: a research field in progress. Trends Ecol. Evol. 8, 392–395.

Mokady, O., Loya, Y., Achituv, Y., Geffen, E., Graur, D., Rozenblatt, S., Brickner, I., 1999. Speciation versus phenotypic plasticity in coral inhabiting barnacles: Darwin's observations in an ecological context. J. Mol. Evol. 49, 367-375.

Nei, M. 1987. Molecular Evolutionary Genetics. 1st edn. Columbia University Press, New York.

Posada, D., Crandall, K.A., 1998. MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Posada, D., Crandall, K.A., Templeton, A.R., 2000. GEODIS: a program for the cladistic nested analysis of the geographic distribution of genetic haplotypes. Mol. Ecol. 9,

487-488.

Proudlove, G., Wood, P.J., 2003. The blind leading the blind: cryptic subterranean species and DNA taxonomy: Trends Ecol. Evol. 18, 272–273.

Rissler, L.J., Taylor, D.R., 2003. The phylogenetics of Desmognathine salamander populations across the southern Appalachians. Mol. Phylogenet. Evol. 27, 197-211.

Rogers, A.R., Harpending, H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9, 552-569.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DNASP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496-2497.

Schultheis, A.S., Weigt, L.A., Hendricks, A.C., 2002. Gene flow, dispersal, and nested clade analysis among populations of the stonefly *Peltoperla tarteri* in the southern Appalachians. Mol. Ecol. 11, 317-327.

Sinclair, E.A., Fetzner, J.W., Jr., Buhay, J.E., Crandall, K.A. 2004. Proposal to complete a phylogenetic taxonomy and systematic revision for freshwater crayfish (Astacidea). Freshwater Crayfish 14, 21-29.

Sites, J.W., Jr, Crandall K.A., 1997. Testing species boundaries in biodiversity studies. Cons. Biol. 11, 1289-1297.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-595.

Taylor, C.A., Warren, M.L., Jr., Fitzpatrick, J.F., Jr., Hobbs, H.H., III, Jezerinac, R.F., Pflieger, W.L., Robison, H.W., 1996. Conservation status of crayfishes of the United States and Canada. Fisheries 21, 25-38.

Templeton, A.R., 1993. The "Eve" hypothesis: a genetic critique and reanalysis. Amer. Anthropol. 95, 51-72.

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

Templeton, A.R., 2001. Using phylogeographic analysis of gene trees to test species status and processes. Mol. Ecol. 10, 779-791.

Templeton, A.R., Routman, E., Phillips, C.A., 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 149, 767-782.

Thomas, W.A., 1986. Sequatchie Anticline, the northwesternmost structure of the Appalachian fold-thrust belt in Alabama. In: Neathery, T. L. (Ed.), Centennial field guide, Southeastern section of the Geological Society of America. Boulder, CO: Geological Society of America, pp. 177-180.

Watterson, G.A., 1975. On the number of segregrating sites in genetical models without recombination. Theor. Pop. Biol. 7, 256-276.

Wiens, J.J., Chippindale, P.T., Hillis, D.M., 2003. When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. Syst. Biol. 52, 501–514.

Wilcox, T.P., Zwickl, D.J., Heath, T.A., Hillis, D.M., 2002. Phylogenetic relationships of

the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. 25, 361–371.

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Kidd, J.R., Deinard, A.S., Ryder, O., Kidd, K.K., Li, W.H., 2003. Low nucleotide diversity in chimpanzees and bonobos. Genetics 164, 1511-1518.

Table 1. *Cambarus* (subgenus *Aviticambarus*) taxa, cave names with locations and cave survey numbers, and abbreviated cave names.

Species	Cave	Sampled	State: County	Cave Survey #	Abbreviation
C. hamulatus	Aaron Tolletts[a]	yes	TN: Bledsoe	TBD1	AARON
C. hamulatus	Run To The Mill	yes	TN: Cumberland	TCD62	RUNTOTHEMILL
C. hamulatus	Cave Cove	no	TN: Franklin	TFR33	----------
C. hamulatus	Garner Spring	yes	TN: Franklin	TFR199	GARNER
C. hamulatus	Little Crow Creek	no	TN: Franklin	TFR15	----------
C. hamulatus	Payne Spring	yes	TN: Franklin	TFR358	PAYNE
C. hamulatus	Bible Spring	yes	TN: Marion	TMN91	BIBLE
C. hamulatus	Butterfly	no	TN: Marion	TMN160	----------
C. hamulatus	Druin Spring	yes	TN: Marion	TMN156	DRUIN
C. hamulatus	Gourdneck	no	TN: Marion	TMN14	----------
C. hamulatus	Honeycutt[a]	no	TN: Marion	TMN16	----------
C. hamulatus	Lost Pig[a]	no	TN: Marion	TMN20	----------
C. hamulatus	Nickajack[a] (type)	no**	TN: Marion	TMN26	----------
C. hamulatus	Owen[a]	yes	TN: Marion	TMN176	OWEN
C. hamulatus	Pryor Cave Spring	yes	TN: Marion	TMN129	PRYOR
C. hamulatus	Shakerag	yes	TN: Marion	TMN371	SHAKERAG
C. hamulatus	Ship[a]	no	TN: Marion	TMN39	----------
C. hamulatus	Signal Light Pit	yes	TN: Marion	TMN40	SIGNAL
C. hamulatus	Snake Well	no	TN: Marion	TMN262	----------
C. hamulatus	Speegle Saltpeter[a]	no	TN: Marion	TMN46	----------
C. hamulatus	Whiteside[a]	yes	TN: Marion	TMN48	WHITESIDE
C. hamulatus	Wine[a]	yes	TN: Marion	TMN141	WINE
C. hamulatus	Keyhole	yes	TN: Sequatchie	TSQ15	KEYHOLE
C. hamulatus	Wilmoth	yes	TN: Sequatchie	TSQ5	WILMOTH
C. hamulatus	Graves	yes	AL: Blount	ABA1200	GRAVES
C. hamulatus	Randolph[a]	no**	AL: Blount	ABA414	----------
C. hamulatus	Rickwood[a]	yes	AL: Blount	ABA236	RICKWOOD
C. hamulatus	Bluff River	yes	AL: Jackson	AJK2800	BLUFF
C. hamulatus	Crow Creek[a]	no	AL: Jackson	AJK1074	----------
C. hamulatus	Geiger[a]	yes	AL: Jackson	AJK459	GEIGER
C. hamulatus	Horseskull[a]	no	AL: Jackson	AJK613	----------
C. hamulatus	Jess Elliott[a]	no	AL: Jackson	AJK323	----------
C. hamulatus	Kyles	no	AL: Jackson	AJK289	----------
C. hamulatus	Russell[a]	no	AL: Jackson	AJK169	----------
C. hamulatus	Salt River[a]	yes	AL: Jackson	AJK221	SALT
C. hamulatus	Talley Ditch[a]	yes	AL: Jackson	AJK248	TALLEY
C. hamulatus	Tate	no	AL: Jackson	AJK324	----------
C. hamulatus	Tumbling Rock[a]	no	AL: Jackson	AJK171	----------
C. hamulatus	Buds[a]	no***	AL: Marshall	AMS1135	----------
C. hamulatus	King School[a]	no*	AL: Marshall	AMS39	----------
C. jonesi	McKinney Pit[a]	no*	AL: Colbert	ACE46	----------
C. jonesi	Key[a]	yes	AL: Lauderdale	ALD99	KEY
C. jonesi	Rockhouse[a]	no	AL: Limestone	ALM312	----------
C. jonesi	White Spring[a]	yes	AL: Limestone	ALM242	WHITE
C. jonesi	Arrowwood	no**	AL: Madison	AMD1908	----------
C. jonesi	Barclay	no	AL: Madison	AMD55	----------
C. jonesi	Bobcat[a]	no	AL: Madison	AMD1283	----------
C. jonesi	Byrd Spring[a]	no	AL: Madison	AMD606	----------
C. jonesi	Matthews[a]	no	AL: Madison	AMD23	----------
C. jonesi	Shelta[a]	no*	AL: Madison	AMD4	----------
C. jonesi	Cave Spring[a] (type)	yes	AL: Morgan	AMG53	CAVE
C. jonesi	Talucah[a]	no*	AL: Morgan	AMG47	----------
C. sp nov 1	Lacon Exit	yes	AL: Morgan	AMG3343	LACON
C. sp nov 2	Kellers[a]	yes	AL: Marshall	AMS326	KELLERS
C. sp nov 2	Porches Spring	yes	AL: Marshall	AMS693	PORCHES
C. sp nov 2	Cherry Hollow	yes	AL: Marshall	AMS1710	CHERRY
C. sp nov 2	Beech Spring[a]	no*	AL: Marshall	AMS347	----------
C. veitchorum	White Spring (type)	no*	AL: Limestone	ALM242	----------

[a] = records from Hobbs et al. 1977
* = cave visited but species not found
** = cave impounded by dam
*** = cave destroyed

Table 2. *Cambarus* individuals with voucher numbers, locality, and 16S haplotype.

Individual #	Species	Locality	16S haplotype
JC353	*C. hamulatus*	Aaron Tolletts Cave	11
JF2784	*C. hamulatus*	Bible Spring Cave	6
JF2785	*C. hamulatus*	Bible Spring Cave	6
JC752	*C. hamulatus*	Bluff River Cave	8
JC753	*C. hamulatus*	Bluff River Cave	6
JF2835	*C. hamulatus*	Bluff River Cave	6
JF2836	*C. hamulatus*	Bluff River Cave	6
JF2837	*C. hamulatus*	Bluff River Cave	6
JF2838	*C. hamulatus*	Bluff River Cave	6
JF2839	*C. hamulatus*	Bluff River Cave	6
JF2840	*C. hamulatus*	Bluff River Cave	6
JF2841	*C. hamulatus*	Bluff River Cave	6
JF2842	*C. hamulatus*	Bluff River Cave	6
JF2843	*C. hamulatus*	Bluff River Cave	6
JF2844	*C. hamulatus*	Bluff River Cave	6
JF2845	*C. hamulatus*	Bluff River Cave	6
JF2846	*C. hamulatus*	Bluff River Cave	6
JF2847	*C. hamulatus*	Bluff River Cave	6
JF2848	*C. hamulatus*	Bluff River Cave	6
JF2849	*C. hamulatus*	Bluff River Cave	6
JF2850	*C. hamulatus*	Bluff River Cave	6
JF2851	*C. hamulatus*	Bluff River Cave	6
JF2852	*C. hamulatus*	Bluff River Cave	6
JF2853	*C. hamulatus*	Bluff River Cave	6
JC844	*C. hamulatus*	Druin Spring Cave	10
JF2786	*C. hamulatus*	Druin Spring Cave	10
JF2787	*C. hamulatus*	Druin Spring Cave	10
JF2788	*C. hamulatus*	Druin Spring Cave	10
JF2789	*C. hamulatus*	Druin Spring Cave	10
JC1513	*C. hamulatus*	Garner Spring Cave	6
JC1515	*C. hamulatus*	Garner Spring Cave	7
JC1549	*C. hamulatus*	Garner Spring Cave	10
JC1550	*C. hamulatus*	Garner Spring Cave	10
JC1551	*C. hamulatus*	Garner Spring Cave	10
JC1552	*C. hamulatus*	Garner Spring Cave	10
JC1553	*C. hamulatus*	Garner Spring Cave	6
JC1554	*C. hamulatus*	Garner Spring Cave	6
JC1555	*C. hamulatus*	Garner Spring Cave	10
JC809	*C. hamulatus*	Geiger Cave	6
JF3338	*C. hamulatus*	Geiger Cave	6
JF3339	*C. hamulatus*	Geiger Cave	6
JC2415	*C. hamulatus*	Graves Cave	1
JC2416	*C. hamulatus*	Graves Cave	2
JC2417	*C. hamulatus*	Graves Cave	1
JC2418	*C. hamulatus*	Graves Cave	3
JC2419	*C. hamulatus*	Graves Cave	3
JC2245	*C. hamulatus*	Keyhole Cave	11
JC2246	*C. hamulatus*	Keyhole Cave	11
JC2247	*C. hamulatus*	Keyhole Cave	11
JC2248	*C. hamulatus*	Keyhole Cave	11
JC2249	*C. hamulatus*	Keyhole Cave	11
JC2250	*C. hamulatus*	Keyhole Cave	11
JC2251	*C. hamulatus*	Keyhole Cave	11
JC2252	*C. hamulatus*	Keyhole Cave	11
JF2790	*C. hamulatus*	Owen Spring Cave	10
JC1627	*C. hamulatus*	Payne Spring Cave	6
JC1628	*C. hamulatus*	Payne Spring Cave	10
JC1548	*C. hamulatus*	Pryor Cave Spring	10
JC2421	*C. hamulatus*	Rickwood Caverns	12
JC2422	*C. hamulatus*	Rickwood Caverns	12
JC2423	*C. hamulatus*	Rickwood Caverns	12
JC2424	*C. hamulatus*	Rickwood Caverns	12
JC2255	*C. hamulatus*	Run to the Mill Cave	9
JC2256	*C. hamulatus*	Run to the Mill Cave	9
JC2257	*C. hamulatus*	Run to the Mill Cave	9
JC2258	*C. hamulatus*	Run to the Mill Cave	9
JC2259	*C. hamulatus*	Run to the Mill Cave	9
JC2260	*C. hamulatus*	Run to the Mill Cave	9
JC2261	*C. hamulatus*	Run to the Mill Cave	9
JC2262	*C. hamulatus*	Run to the Mill Cave	9
JC2263	*C. hamulatus*	Run to the Mill Cave	9
JC2264	*C. hamulatus*	Run to the Mill Cave	9
JC2265	*C. hamulatus*	Run to the Mill Cave	9
JC2266	*C. hamulatus*	Run to the Mill Cave	9
JC2267	*C. hamulatus*	Run to the Mill Cave	9
JC2268	*C. hamulatus*	Run to the Mill Cave	9
JC2269	*C. hamulatus*	Run to the Mill Cave	9
JC2270	*C. hamulatus*	Run to the Mill Cave	9
JC2273	*C. hamulatus*	Run to the Mill Cave	9
KC713	*C. hamulatus*	Salt River Cave	6
JC2228	*C. hamulatus*	Shakerag Cave	6
JC2229	*C. hamulatus*	Shakerag Cave	10
JC1966	*C. hamulatus*	Signal Light Pit	4
JC1967	*C. hamulatus*	Signal Light Pit	4
JC1968	*C. hamulatus*	Signal Light Pit	4
JC2425	*C. hamulatus*	Talley Ditch Cave	6
JC2426	*C. hamulatus*	Talley Ditch Cave	6
JC2427	*C. hamulatus*	Talley Ditch Cave	6
JC2428	*C. hamulatus*	Talley Ditch Cave	6
JC2429	*C. hamulatus*	Talley Ditch Cave	6
JC2430	*C. hamulatus*	Talley Ditch Cave	6
JC2431	*C. hamulatus*	Talley Ditch Cave	6
JC2432	*C. hamulatus*	Talley Ditch Cave	6
JC2433	*C. hamulatus*	Talley Ditch Cave	6
JC2434	*C. hamulatus*	Talley Ditch Cave	6
JC2435	*C. hamulatus*	Talley Ditch Cave	6
JC2059	*C. hamulatus*	Whiteside Cave	5
JC2060	*C. hamulatus*	Whiteside Cave	4
JC2242	*C. hamulatus*	Wilmoth Cave	11
JC2243	*C. hamulatus*	Wilmoth Cave	11
JC2244	*C. hamulatus*	Wilmoth Cave	11
JC831	*C. hamulatus*	Wine Cave	10
JF2791	*C. hamulatus*	Wine Cave	10
KC1916	*C. jonesi*	Cave Spring Cave	20
JC1643	*C. jonesi*	Key Cave	19
JC778	*C. jonesi*	White Spring Cave	14
JC779	*C. jonesi*	White Spring Cave	14
JC781	*C. jonesi*	White Spring Cave	16
JC783	*C. jonesi*	White Spring Cave	14
JC785	*C. jonesi*	White Spring Cave	13
JC786	*C. jonesi*	White Spring Cave	13
JC787	*C. jonesi*	White Spring Cave	15
JC788	*C. jonesi*	White Spring Cave	18
JC789	*C. jonesi*	White Spring Cave	17
JC790	*C. jonesi*	White Spring Cave	13
JC791	*C. jonesi*	White Spring Cave	18
JC792	*C. jonesi*	White Spring Cave	18
JC2436	*C. sp nov 1*	Lacon Exit Cave	22
JC2437	*C. sp nov 1*	Lacon Exit Cave	22
JC2438	*C. sp nov 1*	Lacon Exit Cave	21
JC2439	*C. sp nov 1*	Lacon Exit Cave	23
JC2540	*C. sp nov 1*	Lacon Exit Cave	21
JC2440	*C. sp nov 2*	Kellers Cave	26
JC2441	*C. sp nov 2*	Kellers Cave	26
JC2442	*C. sp nov 2*	Kellers Cave	26
JC2287	*C. sp nov 2*	Kellers Cave	25
JC2288	*C. sp nov 2*	Kellers Cave	25
JC2289	*C. sp nov 2*	Kellers Cave	25
JC2227	*C. sp nov 2*	Porches Spring Cave	25
JC2412	*C. sp nov 2*	Cherry Hollow Cave	24
Outgroups:			
JF2508	*C. gentryi*	Williams Branch, Dickson Co. TN	
JF2543	*C. friaufi*	Salt Lick Ck, Monroe Co. KY	
JF2579	*C. brachydactylus*	Blue Ck., Humphreys Co. TN	

Table 3. *Cambarus* species, cave names, specimen voucher numbers, and GenBank accession numbers for gene sequences included in this study.

Species	Cave	Specimen #	16S[a]	12S[b]	CO1[c]	GAPDH[d]	Histone H3[e]
C. hamulatus	AARON	JC353	DQ411744	DQ411711	DQ411760	na	DQ411794
C. hamulatus	BIBLE	JF2785	DQ411739	DQ411712	DQ411761	DQ411786	DQ411794
C. hamulatus	BLUFF	JF2835	DQ411739	DQ411713	DQ411762	DQ411786	DQ411795
C. hamulatus	DRUIN	JC844	DQ411743	DQ411714	DQ411763	DQ411787	DQ411794
C. hamulatus	GARNER	JC1554	DQ411739	DQ411713	DQ411764	DQ411787	DQ411796
C. hamulatus	GEIGER	JC809	DQ411739	DQ411713	DQ411765	na	DQ411797
C. hamulatus	GRAVES	JC2418	DQ411736	DQ411715	DQ411766	DQ411788	DQ411794
C. hamulatus	KEYHOLE	JC2245	DQ411744	DQ411716	DQ411767	na	DQ411794
C. hamulatus	OWEN	JF2790	DQ411743	DQ411717	DQ411768	na	DQ411798
C. hamulatus	PAYNE	JC1627	DQ411739	DQ411713	DQ411764	DQ411787	DQ411794
C. hamulatus	PRYOR	JC1548	DQ411743	DQ411714	DQ411769	DQ411787	DQ411794
C. hamulatus	RICKWOOD	JC2424	DQ411745	DQ411718	DQ411770	DQ411786	DQ411794
C. hamulatus	RUNTOTHEMILL	JC2265	DQ411742	DQ411719	na	DQ411789	DQ411794
C. hamulatus	SALT	KC713	DQ411739	DQ411720	DQ411771	DQ411787	DQ411795
C. hamulatus	SHAKERAG	JC2229	DQ411743	DQ411717	DQ411772	DQ411786	DQ411794
C. hamulatus	SIGNAL	JC1967	DQ411737	DQ411721	DQ411773	DQ411786	DQ411795
C. hamulatus	TALLEY	JC2434	DQ411739	DQ411722	DQ411774	DQ411787	DQ411799
C. hamulatus	WHITESIDE	JC2059	DQ411738	DQ411723	DQ411775	na	DQ411794
C. hamulatus	WILMOTH	JC2242	DQ411744	DQ411714	DQ411776	DQ411786	DQ411794
C. hamulatus	WINE	JC831	DQ411743	DQ411717	DQ411768	DQ411786	DQ411800
C. jonesi	CAVE	KC1916	DQ411753	DQ411724	DQ411777	DQ411790	DQ411805
C. jonesi	KEY	JC1643	DQ411752	DQ411725	DQ411778	DQ411791	DQ411806
C. jonesi	WHITE	JC781	DQ411749	DQ411726	DQ411779	DQ411790	DQ411805
C. sp. nov. 2	CHERRY	JC2412	DQ411757	DQ411727	DQ411780	na	DQ411808
C. sp. nov. 2	KELLERS	JC2442	DQ411759	DQ411727	DQ411781	DQ411792	DQ411807
C. sp. nov. 2	PORCHES	JC2227	DQ411758	DQ411727	DQ411781	DQ411792	DQ411807
C. sp. nov. 1	LACON	JC2436	DQ411755	DQ411728	DQ411782	DQ411793	DQ411801
C. gentryi	na	JF2508	AY853664	DQ411731	DQ411783	na	DQ411804
C. friaufi	na	JF2543	DQ411733	DQ411730	DQ411784	na	DQ411803
C. brachydactylus	na	JF2579	DQ411732	DQ411729	DQ411785	na	DQ411802

[a]16S primers: 16sf-cray: 5' GACCGTGCKAAGGTAGCATAATC 3' and 16s-1492r: 5' GGTTACCTTGTTACGACTT 3'

[b]12S primers: 12sf: 5' GAAACCAGGATTAGATACCC 3' and 12sr: 5' TTTCCCGCGAGCGACGGGCG 3'

[c]CO1 primers: HCO2198: 5' TAAACTTCAGGGTGACCAAAAAATCA 3' and LCO1490: 5' GGTCAACAAATCATAAAGATATTG 3'

[d]GAPDH primers: G3PCq157F: 5' TGACCCCTTCATTGCTCTTGACTA 3' and G3PCq981R: 5' ATTACACGGGTAGAATAGCCAAACTC 3'

[e]Histone H3 primers: H3af: ATGGCTCGTACCAAGCAGACVGC 3' and H3ar: 5' ATATCCTTRGGCATRGTGAC 3'

Table 4. Results of the Nested Clade Analysis of *Aviticambarus* 16S haplotypes based on 5000 permutations.

0-step clades			1-step clades			2-step clades			3-step clades		
Haplotype	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn
1			1-1	0.00s	96.18L	2-1	64.06L	62.10L	3-1	30.70s	30.44s
2											
3											
4	1.66s	1.87s	1-2	1.87s	47.00s						
5	0.00	1.87									
I-T	1.66s	0.00s	I-T	1.87	-48.18s						
6	13.15	13.15	1-3			2-2	13.03s	16.09s			
7	0.00	6.43									
8	0.00	13.04									
I-T	13.15	3.41				I-T	-51.02s	-46.00s			
9			1-4	0.00s	72.72L	2-3	33.48s	31.09s	3-2	45.45	37.41
10	10.06s	20.74	1-5	26.46	28.90						
11	24.43	37.06									
I-T	-14.36	-16.31									
12			1-6			2-4	0.00s	182.61L			
13			1-7			2-5	0.00s	23.73	3-3	37.22	116.77L
14											
15			1-11	0.00s	28.48s	2-6	40.83	39.92			
16											
17											
18			1-8	0.00s	28.48s						
19			1-9	0.00	50.87L						
20			1-10	0.00	36.97						
			I-T	0.00	-6.17	I-T	-40.83s	-16.18			
21			1-12			2-7			3-4	0.00s	109.56
22											
23											
						I-T				15.59	14.99
24			1-2A			2-2A	0.00	7.46			
25	1.80	2.03	1-1A			2-1A	0.00	7.46s			
26	0	2.03s									
I-T	1.80	0.00L									

Note. Clade (Dc) and nested clade (Dn) distances are given. An "s" indicates that the distance is significantly small at the 5% confidence level and an "L" indicates that the distance is significantly large. In clades with both tip and interior groups, the average I-T distance is provided. Interior clades with genetic-geographic differences are shaded.

Table 5. Nested contingency results and inferred patterns**.**

Clade	Chi-Square	Probability	Inference Chain	Inferred Pattern
1-2	1.8750	0.4022	1-2-11-12-NO	CRE
1-3	11.3118	0.4276	NA	NA
1-5	27.0000	0.0000*	1-2-11-12-NO	CRE
2-1	10.0000	0.0076*	1-19-20-2-11-12-13-YES	LDC
2-3	45.0000	0.0000*	1-2-11-17-4-NO	RGF w/ IBD
2-6	16.0000	0.0342*	1-2-11-12-NO	CRE
3-1	54.0000	0.0000*	1-2-11-12-NO	CRE
3-2	49.0000	0.0000*	1-2-11-12-13-YES	LDC
3-3	1.7500	1.0000	1-2-11-12-NO	CRE
Total	350.6975	0.0000*	1-3-5-15-NO	PF and LDC
1-1A	0.8750	1.0000	1-2-11-17-4-NO	RGF
Total A	0.0000	0.0000*	1-2-11-17-4-NO	RGF w/ IBD

* indicates significant at the P<0.05 level. Inferences were made using Templeton's 2005 key. Abbreviations for the inferences are: CRE, contiguous range expansion; NA, not applicable; LDC, long-distance colonization; RGF, restricted gene flow; IBD, isolation by distance; PF, past fragmentation.

Table 6. Current ($\theta_\pi \pm$ SE) and historical-based (θ_W) estimates of genetic diversity and effective population sizes estimated using an equal sex ratio, ten-year generation time, and 2.2 X 10^{-8} substitution rate.

	Current		Historical	
Cave Species	θ_π	N_e	θ_W	N_e
C. hamulatus	0.00442 ± 0.00039	40182	0.00477	43364
C. jonesi	0.00452 ± 0.00085	41091	0.00585	53182
C. sp. nov. 1	0.00206 ± 0.00057	18727	0.00198	18000
C. sp. nov. 2	0.00369 ± 0.00183	33545	0.00478	43455

Figure 1. Distribution of each *Cambarus* (subgenus *Aviticambarus*) species. Dots in the middle of the symbols represent sampled caves and open symbols are not included in this study. The subgenus is currently known from 58 cave sites. Areas referred to in the text are labeled for geographic reference.

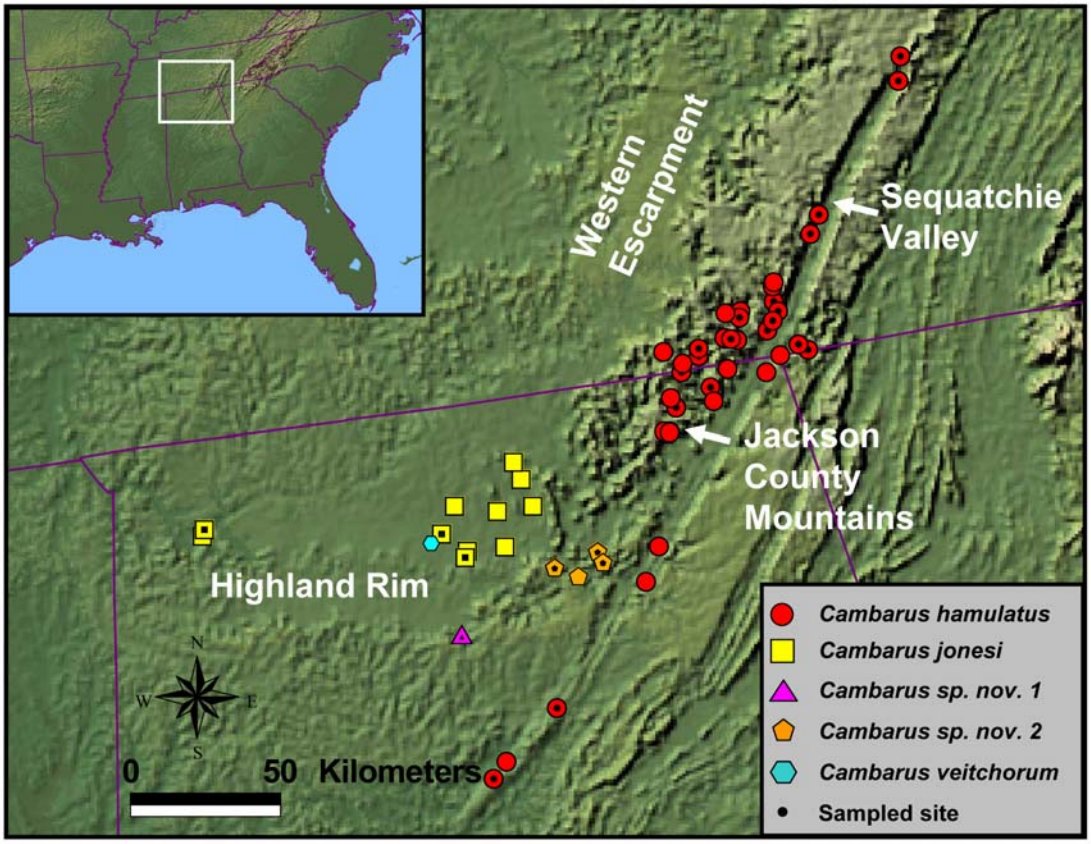


Figure 2. Phylogram of the relationships between 16S haplotypes for each of the species (Bayesian topology shown). Analyses done in PhyML are given below the nodes as bootstrap support (BS) percentages from 500 replicates (log likelihood = -1340.87). Bayesian support values are given above the nodes as posterior probability (PP) percentages (log likelihood = -1374.16). Support values are not shown for intra-specific groupings.

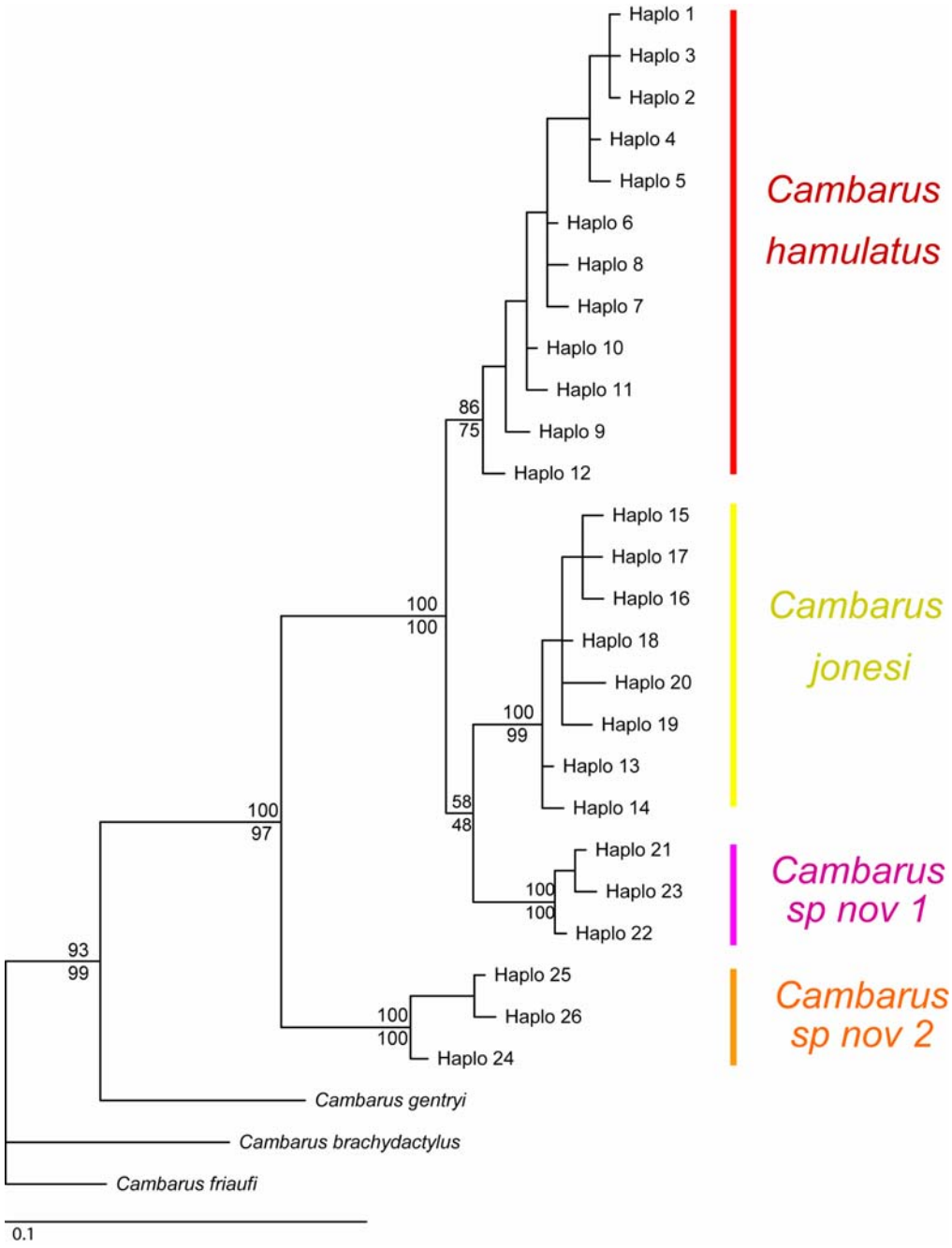


Figure 3. Phylogenetic relationships of the species with the subgenus *Aviticambarus* estimated from the combined dataset of three mtDNA genes (12S, 16S, CO1) and two nuclear genes (Histone H3 and GAPDH) totaling 2686 bases. The first values are bootstrap support of 500 replicates in PhyML (log likelihood = -8026.43) and values after the slash represent posterior probabilities from Bayesian analysis (log likelihood = -6811.02) since both analyses yielded similar topologies (Bayesian topology shown). Support values are not shown for intra-specific groupings.

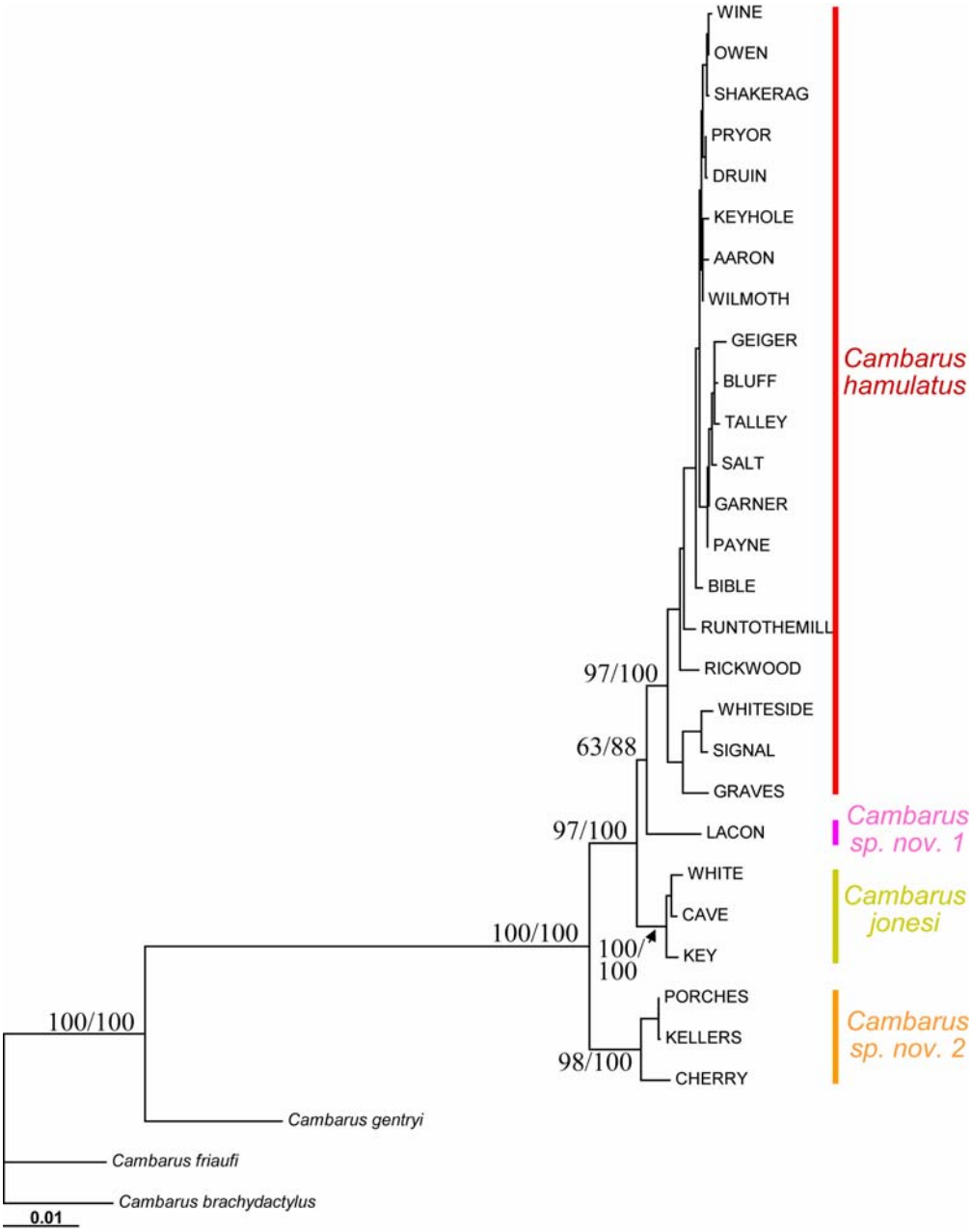


Figure 4. Haplotype network depicting the nesting levels used to infer historical processes which gave rise to current genetic structuring. Haplotype circles are colored to represent the four lineages: *Cambarus hamulatus* (haplotypes 1-12, red), *Cambarus jonesi* (haplotypes 13-20, yellow), *Cambarus sp. nov. 1* (haplotypes 21-23, pink), and *Cambarus sp. nov. 2* (haplotypes 24-26, orange). Ancestral haplotypes are represented by squares and empty circles represent unsampled or extinct haplotypes. The significance level connecting the network was nine steps, and *C. sp. nov. 1* and *C. sp. nov. 2* were outside the 95% confidence limit in addition to the three outgroup taxa.

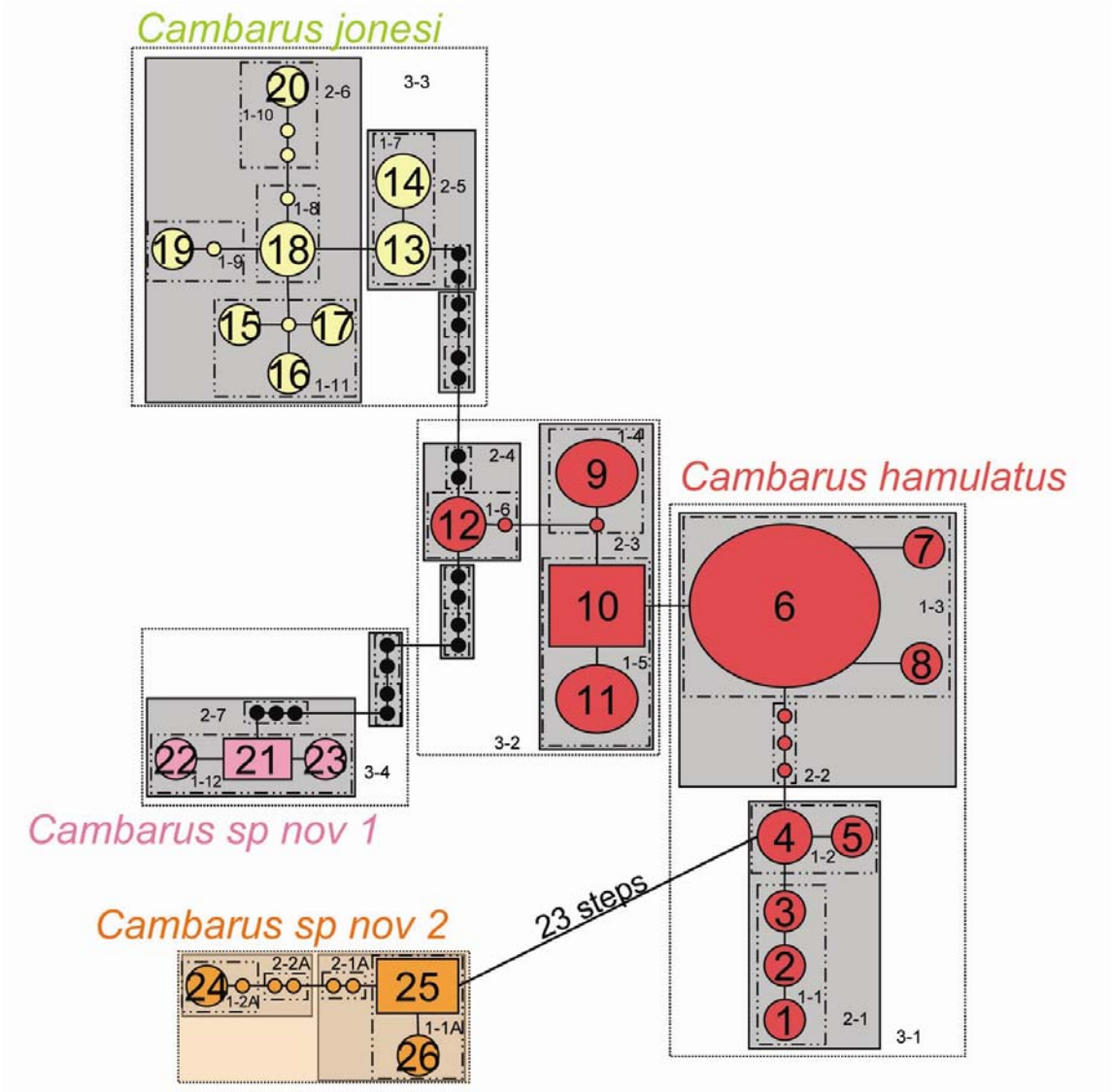
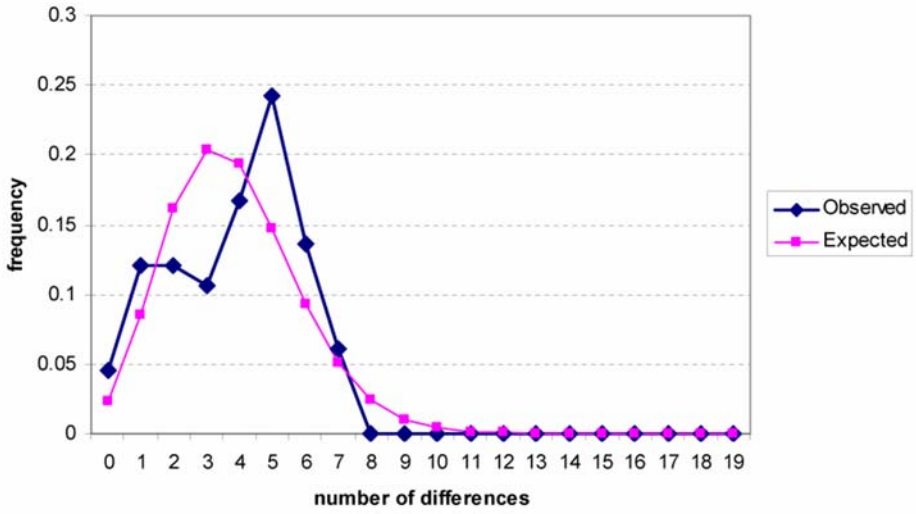


Figure 5. Mismatch distribution for the 16S haplotypes for *Cambarus hamulatus*. The observed frequency is represented by the diamond and thick solid line, and the expected frequency under the expansion model is depicted by thin solid line connecting square symbols.



Molecular Ecology (2005) **14**, 4259–4273 doi: 10.1111/j.1365-294X.2005.02755.x

Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes

JENNIFER E. BUHAY* and KEITH A. CRANDALL†
**Department of Integrative Biology, Brigham Young University, Provo, UT 84602-5255, USA, †Department of Integrative Biology, Department of Microbiology and Molecular Biology, and the Monte L. Bean Life Science Museum, Brigham Young University, Provo, UT 84602-5255, USA*

Abstract

Subterranean animals are currently viewed as highly imperiled, precariously avoiding extinction in an extreme environment of darkness. This assumption is based on a hypothesis that the reduction in visual systems and morphology common in cave faunas reflects a genetic inability to adapt and persist coupled with the perception of a habitat that is limited, disconnected, and fragile. Accordingly, 95% of cave fauna in the United States are presumed endangered due to surface environmental degradation and limited geographic distributions. Our study explores the subterranean phylogeography of stygobitic crayfishes in the southeastern United States, a global hotspot of groundwater biodiversity, using extensive geographic sampling and molecular data. Despite their endangered status, our results show that subterranean crayfish species have attained moderate to high levels of genetic diversity over their evolutionary histories with large population sizes and extensive gene flow among karst systems. We then compare the subterranean population histories to those of common surface stream-dwelling crayfishes. Our results show recent drastic declines in genetic variability in the surface crayfish and suggest that these species also warrant conservation attention.

Keywords: cave fauna, conservation genetics, crustaceans, endangered species, phylogeography, stygobite

Received 21 June 2005; revision received 19 July 2005; accepted 24 August 2005

Introduction

According to the Nature Conservancy, 95% of subterranean fauna in North America is considered vulnerable or imperiled using criteria similar to the IUCN-World Conservation Union Red List (Master 1991; Culver *et al*. 2000). The listings are based mostly on surface threats to groundwater systems (Danielopol *et al*. 2003), small geographic ranges (Culver *et al*. 2000), and habitat destruction, not in-depth species-specific biological studies. In fact, current scientific information on subterranean fauna is scarce, leaving the field of biospeleology and the unique biome in the dark. The convergent nature of cave life obscures species' relationships and geographic boundaries, while the inaccessibility of the underground microhabitat makes physical counts of census sizes almost impossible to confidently assess. Molecular genetic approaches are best employed in these situations to accurately estimate biodiversity and critically evaluate the conservation status of elusive organisms (DeSalle & Amato 2004).

Two hypotheses (as reviewed by Kane 1982) have been proposed concerning the genetic diversity, and hence the conservation status and extinction risk (Spielman *et al*. 2004), of subterranean fauna. Barr (1968) suggested that a genetic bottleneck initially occurs during the separation of the surface ancestor from its obligate cave-dwelling descendent. Barr suggested that this bottleneck is short in duration and that cave populations recover from the break in gene flow by range expansion and population growth into new uninhabited subterranean areas. In contrast, Poulson & White (1969) proposed that older fauna show

Correspondence: Jennifer E. Buhay, Fax: 801-422-0090; E-mail: crayfish@byu.edu

low genetic variability due to the long isolation and adaptation to the stable underground environment. They also suggested that the decrease in phenotypic variance in visual structures and morphological traits reflects a decreased genetic variability. Poulson & White (1969) also stressed the probable relationship between reduced genetic variability with the reduction of population size, reduced rate of population growth, longer maturation times, and longer lifespans. Previous studies (Avise & Selander 1972; Swofford *et al*. 1980; Koppelman & Figg 1995) on aquatic obligate cave species (stygobites) were consistent with the Poulson and White hypothesis, but each of the studies had sparse sampling across small geographic areas within the species' ranges and these studies were conducted using allozymes, which can underestimate genetic diversity. Our study tests these two alternative hypotheses for the first time using exceptional sampling and high-resolution genetic data from a group of subterranean crayfishes. We also compare our cave crayfish findings to those of two common surface stream-dwelling crayfish species for broader understanding of subsurface and surface freshwater habitats and conservation.

Materials and methods

Study organisms

One of the largest animals in caves are blind crayfish, which are found in all kinds of subterranean aquatic areas, including deep rivers and lakes, small seeps, rimstone pools, and mudholes. A group of stygobitic crayfishes in the genus *Orconectes* inhabits the karst groundwaters of the western escarpment of the Cumberland Plateau, ranging from eastern Kentucky south to northern Alabama (Hobbs & Barr 1972; Hobbs *et al*. 1977). As currently recognized, there are three obligate cave-dwelling *Orconectes* species along the plateau: *Orconectes incomptus*, *Orconectes australis* (with two subspecies, *australis* and *packardi*), and *Orconectes sheltae*, which was only known from one Mississippian Age cave in Alabama (Cooper 1975; Cooper & Cooper 1997) and is currently presumed extinct, with the last sighting by Hobbs & Bagley (1989). *O. incomptus* is found only in Ordovician Age limestone in an area just west of the escarpment. *O. australis* is found in Mississippian Age limestone along the escarpment, which was formed by the recession and erosion of the Cumberland Plateau in an eastward direction, allowing for cave development on the western side. The conservation categories for these species are: *Orconectes australis australis* (IUCN stable), *O. a. packardi* (IUCN vulnerable), *O. incomptus* (IUCN vulnerable), *O. sheltae* (unlisted).

To thoroughly investigate the genetic diversity and phylogeographic patterning of this unique assemblage, we collected mostly tissue samples (a claw or leg which are regenerated) from 421 individuals from 67 caves spanning the entire geographic range (Table 1). Nondestructive sampling involved returning the captured individual to the capture site immediately after removal of claw or leg. In a few cases, one or two voucher male specimens (preserved in 90% ethanol at the Monte L. Bean Museum at Brigham Young University) were taken from caves discovered after Hobbs *et al*.'s (1977) distribution list of cave crayfish localities to serve as voucher specimens for these caves.

For comparison to surface species, we chose two common surface stream-dwelling *Orconectes* species for which we have substantial molecular data and thoroughly sampled distributions as part of other research investigations. *Orconectes luteus* is a wide-ranging surface species throughout Missouri, while *Orconectes juvenilis* has a restricted range in the Upper Cumberland River and Kentucky River basins of Kentucky. Both *O. luteus* and *O. juvenilis* are assigned to the subgenus *Procericambarus* of the genus *Orconectes* and are IUCN stable species.

Table 1 List of cave *Orconectes* taxa, sampled caves, mtDNA 16S haplotype with number of individuals sequenced in parentheses, 3-step nested clade groupings, geographic information, and geologic age of cave sites used in this study

Species	Cave name	16S Haplotype (# of individuals)	3-step clade	State: county	Geologic age
incomptus	Cherry†	19(2)	3-3	TN: Jackson	Ordovician
incomptus	Flynn Creek	17(1)	3-3	TN: Jackson	Ordovician
incomptus	North Fork	18(2), 20(3)	3-3	TN: Jackson	Ordovician
a. packardi	Teamers	1(1), 2(2)	3-1	KY: Rockcastle	Mississippian
a. packardi	Duvalts	2(1)	3-1	KY: Rockcastle	Mississippian
a. packardi	Pine Hill	2(1)	3-1	KY: Rockcastle	Mississippian
a. packardi	Fletcher Spring	7(2)	3-2	KY: Rockcastle	Mississippian
a. packardi	Cedar Creek	7(14)	3-2	KY: Pulaski	Mississippian
a. packardi	Dykes Bridge	7(3)	3-2	KY: Pulaski	Mississippian
a. packardi	Dave's	6(8), 7(2)	3-1, 3-2	KY: Pulaski	Mississippian
a. packardi	Big Sink	7 (20)	3-2	KY: Pulaski	Mississippian
a. packardi	Hail	3(4), 4(3), 5(1)	3-1	KY: Pulaski	Mississippian

Table 1 *Continued*

Species	Cave name	16S Haplotype (# of individuals)	3-step clade	State: county	Geologic age
a. packardi	Wells	6(3), 7(1)	3-1, 3-2	KY: Pulaski	Mississippian
a. packardi	Jugornot	8(3), 12(13), 13(1), 14(2), 15(1), 16(2)	3-2	KY: Pulaski	Mississippian
a. packardi	Coral	3(1)	3-1	KY: Pulaski	Mississippian
a. packardi	Sloans Valley†	9(1), 10(2), 11(1)	3-2	KY: Pulaski	Mississippian
sp. nov.	Redmond Creek	24(9)	3-4	KY: Wayne	Mississippian
sp. nov.	Grayson Gunner	23(1)	3-4	KY: Wayne	Mississippian
sp. nov.	Stream	24(2), 25(2)	3-4	KY: Wayne	Mississippian
sp. nov.	Tonya's	23(7)	3-4	KY: Wayne	Mississippian
sp. nov.	Buffalo Saltpeter	23(3)	3-4	KY: Clinton	Mississippian
sp. nov.	Clinton	21(5), 22(1)	3-4	TN: Pickett	Mississippian
sp. nov.	Cornstarch	21(9)	3-4	TN: Fentress	Mississippian
sp. nov.	Redbud	21(1)	3-4	TN: Fentress	Mississippian
a. australis	Fallen Entrance	27(6)	3-6	TN: Fentress	Mississippian
a. australis	Skillmans Mark	27(3), 30(1)	3-6	TN: Fentress	Mississippian
a. australis	Mountain Eye	27(4)	3-6	TN: Fentress	Mississippian
a. australis	Mill Hollow	27(16), 28(1), 50(1), 51(3)	3-6, 3-8	TN: Overton	Mississippian
a. australis	Raven Bluff	37(1)	3-6	TN: Overton	Mississippian
a. australis	Bailey's Webb	27(5)	3-6	TN: Overton	Mississippian
a. australis	Capshaw	27(12), 29(1)	3-6	TN: Putnam	Mississippian
a. australis	Knieps Spring	27(4)	3-6	TN: Putnam	Mississippian
a. australis	Blindfish	26(1), 27(2), 31(3), 32(1), 33(1)	3-6	TN: Putnam	Mississippian
a. australis	Virgin Falls	40(4)	3-7	TN: White	Mississippian
a. australis	Merrybranch	34(1), 35(7), 36(1), 40(22), 41(1), 42(1), 43(1), 44(4), 45(1)	3-6, 3-7	TN: White	Mississippian
a. australis	Lost Creek Resurgence	40(1)	3-7	TN: White	Mississippian
a. australis	Rumbling Falls	40(6)	3-7	TN: VanBuren	Mississippian
a. australis	Winching Hollow Water	35(9), 40(3)	3-6, 3-7	TN: VanBuren	Mississippian
a. australis	Glencora Spring	27(1), 40(4)	3-6, 3-7	TN: VanBuren	Mississippian
a. australis	Waterfall Hollow	54(7)	3-8	TN: VanBuren	Mississippian
a. australis	Lost Cove	51(10), 53(1)	3-8	TN: VanBuren	Mississippian
a. australis	Camps Gulf	40(2), 54(1)	3-7, 3-8	TN: VanBuren	Mississippian
a. australis	Laurel Creek	40(1), 51(17)	3-7, 3-8	TN: VanBuren	Mississippian
a. australis	Lower Norton Spring	49(1), 51(3)	3-8	TN: VanBuren	Mississippian
a. australis	Rocky River	46(5), 47(2)	3-8	TN: Warren	Mississippian
a. australis	Jaco Spring	48(4)	3-8	TN: Warren	Mississippian
a. australis	Cumberland Caverns*	46(1), 51(4)	3-8	TN: Warren	Mississippian
a. australis	Blowing	38(5)	3-7	TN: Warren	Mississippian
a. australis	Woodlee	39(1)	3-7	TN: Grundy	Mississippian
a. australis	Dry	39(1)	3-5	TN: Grundy	Mississippian
a. australis	Red Trillium	61(2)	3-5	TN: Grundy	Mississippian
a. australis	Big Mouth	61(4)	3-5	TN: Grundy	Mississippian
a. australis	Crystal	61(5)	3-5	TN: Grundy	Mississippian
a. australis	Smith Hollow NR1	61(4), 63(1)	3-5	TN: Grundy	Mississippian
a. australis	Lusk	51(1), 61(7), 64(1)	3-5, 3-8	TN: Coffee	Mississippian
a. australis	Pearson	61(26), 62(1)	3-5	TN: Franklin	Mississippian
a. australis	Wet	61(2)	3-5	TN: Franklin	Mississippian
a. australis	Dripping Spring	59(1)	3-5	TN: Franklin	Mississippian
a. australis	Witherspoon	51(7)	3-8	TN: Franklin	Mississippian
a. australis	Floorless	51(1), 52(1)	3-8	TN: Franklin	Mississippian
a. australis	Larkin Spring	65(2)	3-5	AL: Jackson	Mississippian
a. australis	Limrock Blowing	65(28), 67(1), 69(1)	3-5	AL: Jackson	Mississippian
a. australis	Doug Green	56(1)	3-5	AL: Jackson	Mississippian
a. australis	Langston	55(1)	3-5	AL: Jackson	Mississippian
a. australis	Scott	65(3)	3-5	AL: Madison	Mississippian
a. australis	Hering	57(1), 65(12), 66(1)	3-5	AL: Madison	Mississippian
a. australis	Shelta†	58(4), 60(1), 65(1), 68(1)	3-5	AL: Madison	Mississippian

*Represents a known introduced population from a nearby cave; †represents type locality.

Data collection

Genomic DNA was extracted using standard methods and the 16S mtDNA gene was amplified during polymerase chain reaction (PCR) with primers 16sf-cray: GACCGTGCKAAGGTAGCATAATC and 16s-1492r: GGTTACCTTGTTACGACTT (Crandall & Fitzpatrick 1996). The 16S mtDNA is the most variable gene for freshwater crayfishes (Crandall 1997; Fetzner & Crandall 2003). Cycle-sequencing reactions were run with purified PCR products and the BigDye Ready-Reaction kit on a PerkinElmer Thermocycler. Reactions were cleaned using Millipore plates and then sequenced using an ABI377 automated DNA sequencer. Sequences were edited and aligned by eye using BIOEDIT (Hall 1999). GenBank (www.ncbi.nlm.nih.gov) Accession nos of the 16S mtDNA haplotypes used for this study are: *Orconectes a. packardi* AY853595–AY853610; *O. incomptus* AY853611–AY853614; *O.* sp. nov. AY853615–AY853619; *O. a. australis* AY853620–AY853663; *Cambarus gentryi* AY853664; and *Cambarus graysoni* AY853665. R. Ziemba collected samples of *O. juvenilis* (n = 100 individuals), which we sequenced for 16S (unpublished data, available upon request from R. Ziemba). The *O. luteus* (n = 393 individuals) aligned 16S data set (Fetzner & Crandall 2003; GenBank AF376483–AF376521) was provided by J. Fetzner. Both surface species were amplified in PCR and sequenced using primers 16s-1492r and 16s-17sub: ATASRGTCTRACCTGCCC (Fetzner & Crandall 2003).

Phylogenetic analyses

Phylogenetic analyses included 69 unique haplotypes (485 base pairs) from the 421 cave individuals and two outgroup sequences from the closest relatives *C. gentryi* and *C. graysoni* (Sinclair *et al*. 2004; Buhay *et al*., unpublished). The Bayesian analysis (Ronquist & Huelsenbeck 2003) was run for 10 million generations using four chains, sampling 1/1000 trees with parameters nst = 6 and rates = adgamma. We discarded the burn-in (first 1001 trees of 10 001 total determined by Tracer (http://evolve.zoo.ox.ac.uk/software.html), checked for convergence using Tracer, and constructed a 50% majority rule consensus tree. Five independent runs of the same data set with random start trees resulted in nearly identical results. Posterior probabilities (PP) greater than 95% are considered significant support for a clade (Huelsenbeck & Ronquist 2001). The maximum-likelihood analysis was run in PAUP* (Swofford 2001) by heuristic search (fast-stepwise addition with random seed) with 500 replicates using the TrN + I + G model of evolution selected by MODELTEST (Posada & Crandall 1998). Nodal support was assessed using 100 bootstrap (BS) replicates (Felsenstein 1985) with strong clade support of 70% (Hillis & Bull 1993).

Genetic diversity and effective population sizes

To address current and recent historical levels of variation, genetic diversity and effective population sizes within each surface and cave lineage were determined using several methods. We used different estimators of the parameter $\theta = 2N_{e(f)}\mu$ for maternally inherited mitochondrial DNA, to determine effective population size (N_e) with a mutation rate μ (2.2×10^{-8} substitutions per site per year; based on Cunningham *et al*. 1992 estimate for crabs) with generation times of 2 years for surface-dwelling species (Hobbs 1991) and 10 years for stygobitic species (Cooper 1975), and an equal sex ratio (Cooper 1975).

Current genetic diversity (θ_π; Nei 1987 equations 10.5 or 10.6, and the standard error, equation 10.7) was assessed using DNASP 4.0 (Rozas *et al*. 2003). Watterson's (1975) historical genetic diversity estimates (θ_W) were determined using LAMARC (http://evolution.genetics.washington.edu/lamarc.html; Kuhner *et al*. 2004). Current genetic diversity estimates (θ_π) are based on pairwise differences between sequences, while historical diversity estimates (θ_W) are based on the number of segregating sites among the sequences. These two methods used together provide insight into population dynamics over recent evolutionary history (Templeton 1993; Crandall *et al*. 1999; Pearse & Crandall 2004). Differences between current diversity and recent historical diversity are indicative of recent bottlenecks (if $\theta_\pi < \theta_W$) or recent population growth (if $\theta_\pi > \theta_W$) (Templeton 1993; Sinclair *et al*. 2002; Roman & Palumbi 2003; Yu *et al*. 2003).

Pairwise comparisons were used for genealogical estimates of diversity (B_1, θ_2, $\theta_{Ancestor}$) and divergence times using the program IM (Isolation–Migration Model: Nielsen & Wakeley 2001; Hey 2005; Won & Hey 2005; http://lifesci.rutgers.edu/~heylab/heylabsoftware.htm#IM). The HKY (Hasegawa–Kishino–Yano) model with an inheritance scalar of 0.25 for mitochondrial DNA was used with a random seed to initiate the run. A burn-in of 200 000 steps was discarded before recording genealogical steps, and each comparison was run until the effective sample sizes (ESS) were larger than 1000, and in most cases, over 1 million. Multiple independent runs with random start seeds were performed to ensure values were converging on similar estimates. Maximum-likelihood estimates of diversity were used to determine bottleneck (< 1) or growth trends (> 1) between descendent pairs and their ancestors (Descendents : Ancestor ratio) to test the two competing hypotheses about subterranean genetic diversity (Poulson & White 1969 and Barr 1968). Descendent : Ancestor ratios were computed by $(\theta_1 + \theta_2)/\theta_{Ancestor}$ for each pair.

Phylogeographic analyses

Nested clade analysis (NCA: Templeton *et al*. 1995; Templeton 1998) was used to test the null hypothesis of no genetic

differentiation between sampled sites and provide insight into historical processes. The program TCS (Clement *et al*. 2000) was used to construct the haplotype network and GEODIS (Posada *et al*. 2000) was used to test for significant associations between geographic cave locations and genetic distances over 5000 random permutations. Latitude and longitude coordinates of cave localities (at the entrance) were used for the geographic analysis. Haplotypes with the most connections and the highest frequencies are thought to be older, while haplotypes on the tips are more recently evolved. Clade distances (D_c) represent geographic ranges of the clades at each step level. Nested clade distances (D_n) represent the average distances of samples with a particular haplotype with respect to the geographic centre of the clade. Inferences about the historical processes that gave rise to the current genetic patterns were made using the 2004 inference key from A. R. Templeton (http://darwin.uvigo/es/software/geodis.html).

Results

Phylogenetic analysis of 16S mtDNA haplotypes

There are several operational methods available to delineate species boundaries using statistically testable frameworks, as reviewed by Sites & Crandall (1997) and Sites & Marshall (2003). The Genealogical Concordance Species concept (Avise & Ball 1990; Baum & Shaw 1995) is a lineage-based extension of the phylogenetic species concept, in which there is concordance among multiple characters (genetic, environmental, geographic, etc.). A genealogical species is a group of organisms whose members are more closely related to each other ('exclusivity') than to any other organisms outside the group (Baum & Shaw 1995).

We determined the phylogenetic relationships among the two extant species (*Orconectes incomptus* and *Orconectes australis*) using sequence data from the mitochondrial 16S gene (485 base pairs) and identified four distinct lineages: *O. a. packardi*, *O. incomptus*, *O. a. australis*, and *O.* sp. nov. (Fig. 1 and Table 1), each with significant posterior probability support. The cave-dwelling *Orconectes* members are most closely related to burrowing members of the genus *Cambarus* (Crandall & Fitzpatrick 1996; Fetzner 1996; Sinclair *et al*. 2004), rather than to the surface-dwelling members of *Orconectes*, as was previously thought based on similar (convergent) male morphology (Hobbs & Barr 1972), and accordingly, *Cambarus gentryi* and *Cambarus graysoni* were used as the closest outgroup taxa (Sinclair *et al*. 2004; Buhay & Crandall, unpublished).

The most basal member, *O. a. packardi*, was represented by 16 unique mtDNA 16S haplotypes from 13 Mississippian Age caves and 93 individuals, and is distributed from Rockcastle County, Kentucky, south to Pulaski County, Kentucky (Fig. 1: range shown as blue circles, haplotypes 1–16). *O. incomptus* was represented by four unique haplotypes from three Ordovician Age caves in Jackson County, Tennessee (Fig. 1: range shown as pink triangles, haplotypes 17–20). A new species, *O.* sp. nov., found along the Kentucky–Tennessee border (Wayne and Clinton counties, Kentucky, south to northern Fentress County, Tennessee), included five unique haplotypes from eight Mississippian Age caves and 40 individuals (Fig. 1: range shown as green pentagons, haplotypes 21–25). *O. a. australis* was represented by 321 individuals from southern Fentress County, Tennessee south to Madison County, Alabama and included 44 unique haplotypes from 43 Mississippian Age caves (Fig. 1: range shown as orange squares, haplotypes 26–69). Genetic data were acquired from type locality specimens: *O. a. packardi* (Sloans Valley Cave, Pulaski County, Kentucky), *O. incomptus* (Cherry Cave, Jackson County, Tennessee) and *O. a. australis* (Shelta Cave, Madison County, Alabama), and this information was used to clarify species boundaries and their geographic distributions.

Each of these lineages will be considered distinct species based on genetic and geographic concordance (Avise & Ball 1990; Baum & Shaw 1995). Rather than two species (*O. australis* and *O. incomptus*), there are five stygobitic cave *Orconectes* species on the Cumberland Plateau, including the unsampled, possibly extinct *Orconectes sheltae*.

Genetic variation, effective population sizes, and divergence times

Estimates of current (θ_π) and historical (θ_W) genetic diversity were moderate to high (Nei 1987) for the cave dwellers, with the exception of *O.* sp. nov. (Table 2). Similarly, current

Table 2 Current ($\theta_\pi \pm$ SE) and Historical-based (θ_W) estimates of genetic diversity and corresponding effective population sizes for obligate cave-dwelling *Orconectes* species and surface-dwelling *Orconectes* species

	Current		Historical	
	θ_π	N_e	θ_W	N_e
Cave species				
O. a. packardi	0.00455 ± 0.00043	41 364	0.00606	55 082
O. incomptus	0.00508 ± 0.00092	46 182	0.00477	43 375
O. sp. nov.	0.00238 ± 0.00027	21 636	0.00242	22 034
O. a. australis	0.00894 ± 0.00020	81 273	0.01593	144 777
Surface species				
O. juvenilis	0.00394 ± 0.00024	179 091	0.03179	1 445 182
O. luteus	0.02501 ± 0.00015	1 136 818	0.06076	2 761 955

$\mu = 2.2 * 10^{-8}$ substitutions per site per year. Surface-dweller generation time = 2 years, cave-dweller generation time = 10 years.

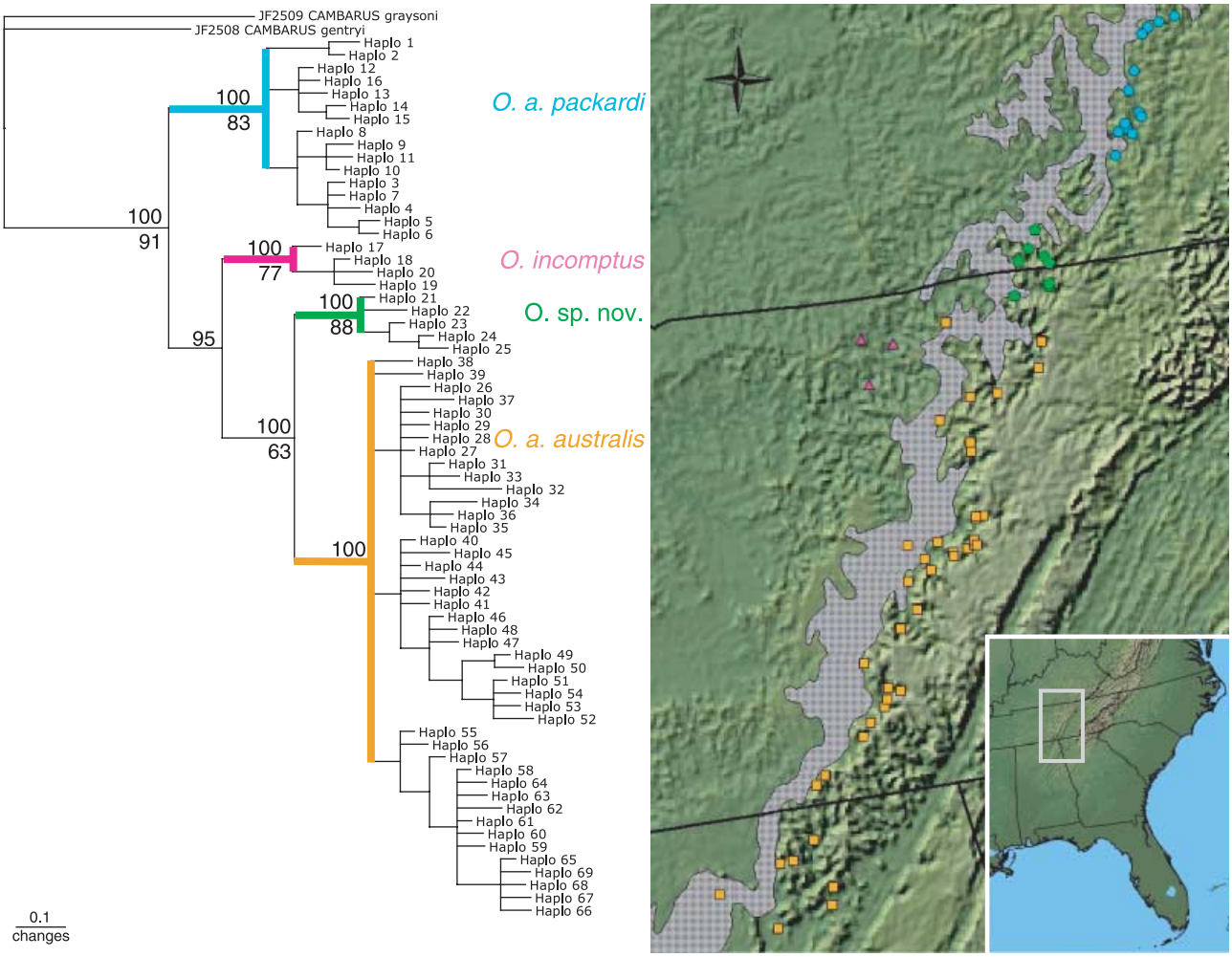


Fig. 1 Geographic distribution (on right) represented by sampled localities for *Orconectes australis packardi* (blue circles), *Orconectes* sp. nov. (green pentagons), and *Orconectes australis australis* (orange squares) along the western escarpment (dark grey shading) of the Cumberland Plateau in Mississippian Age caves at elevations between 180 and 450 m. *Orconectes incomptus* (pink triangles) is found in the area just west of the escarpment in Ordovician Age caves at 150–180 m in elevation. Phylogenetic relationships (on left) are based on 69 haplotypes of 16S mtDNA sequence data using similar results from maximum-likelihood and Bayesian methods. Colours marked on tree match cave species colours from distribution map. *Cambarus graysoni* and *Cambarus gentryi* were used as outgroup taxa. Numbers below branches indicate bootstrap support and numbers above branches indicate posterior probabilities.

effective population sizes (N_e) were also higher than expected, suggesting the occurrence of a vast groundwater network unknown to humans, but as accessible habitat to the stygobitic crayfish. Surprisingly, current (θ_π) and historical (θ_W) estimates for the stygobites were similar (Table 2, with exception of *O. a. australis* which exhibited decline), whereas both surface species estimates show serious recent declines ($\theta_\pi < \theta_W$).

We used a coalescent-based method (Nielsen & Wakeley 2001) to determine genetic diversity over the genealogical histories of each cave species to test the two competing hypotheses regarding genetic diversity of ancestors vs. descendents. Using pairwise species comparisons, we determined genealogical diversity (θ_1 and θ_2) for each crayfish species and $\theta_{Ancestor}$ for their common ancestor, along with their times since divergence (Table 3). These results show a growth trend (descendents/ancestor ratio > 1) after the initial split from the ancestors in cave species comparisons (Fig. 2).

The estimated divergence times for the cave crayfish species are much older than previous speculation (Hobbs *et al*. 1977). Given the broad credibility intervals (90% highest posterior probability densities; HPD) for the *O. a. packardi*–*O. incomptus* and *O. incomptus*–*O.* sp. nov. comparisons, it appears that more loci are needed to resolve divergence times for these species. It is also possible that

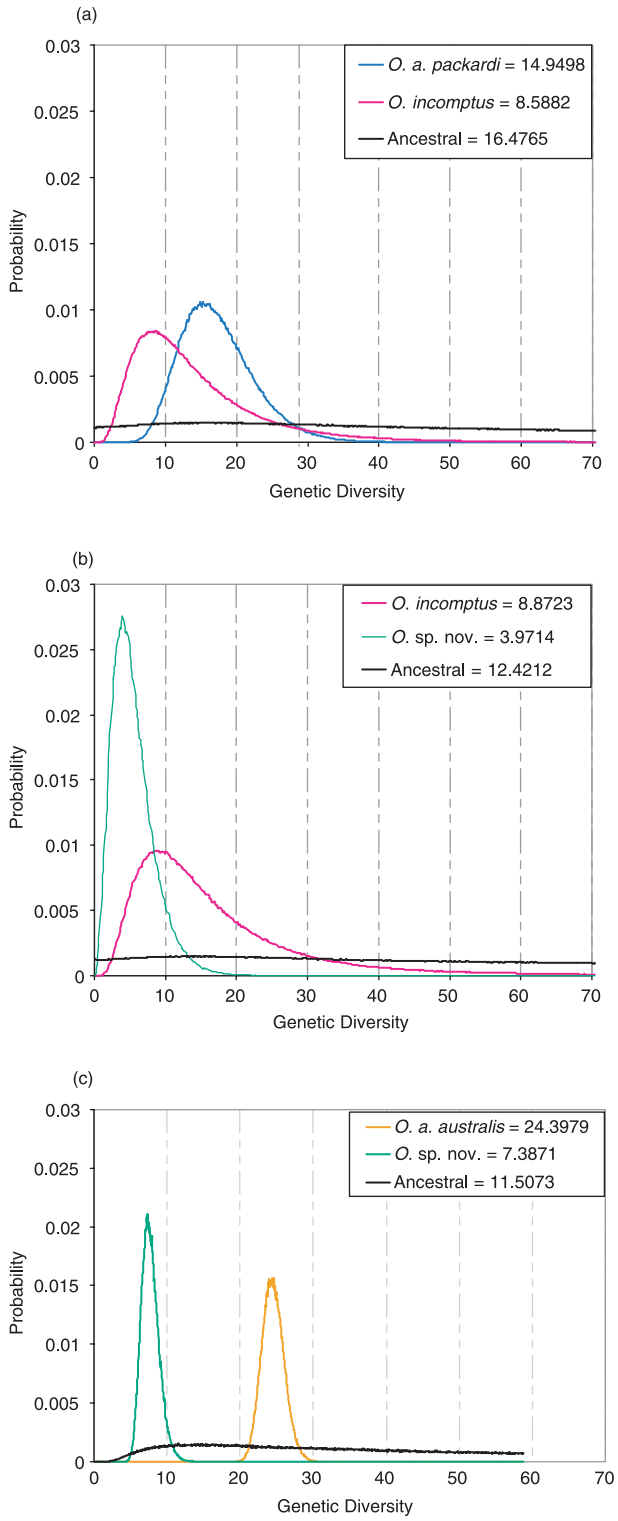


Fig. 2 The marginal posterior probability distributions for the IM model parameter of cave genetic diversity scaled by the neutral mutation rate. Curves are shown for the pairwise analyses of (a) *Orconectes australis packardi* (in blue) vs. *Orconectes incomptus* (in pink) (b) *O. incomptus* (in pink) vs. *O.* sp. nov. (in green), and (c) *O.* sp. nov. (in green) vs. *Orconectes australis australis* (in orange) with their corresponding ancestral (in black) diversities.

more individuals of *O. incomptus* are needed for the IM pairwise analyses, since only eight individuals from three caves of the 10 known sites were sampled for this study. *O. incomptus* is listed in Tennessee as a 'management concern species' and as a 'vulnerable species' by the International Union for Conservation of Nature and Natural Resources (IUCN) which required that sampling restrictions be placed on the collection permit. Interestingly, the split between *O.* sp. nov. and *O. a. australis* was estimated to be 110 million years ago (Ma) (90% HPD interval: 105–116 Ma), in the mid-Cretaceous, which was speculated to be the beginnings of cave invasion for the genus *Cambarus* (Hobbs & Barr 1960). The lower bounds of the 90% HPD intervals for the other two comparisons (*O. a. packardi*–*O. incomptus* at 125 Ma; *O. incomptus*–*O.* sp. nov. at 102 Ma) are similar to that of the *O.* sp. nov.–*O. a. australis* split. Such calculations necessarily make a number of simplifying assumptions and the resulting dates should be taken with caution; however, as outlined below, these divergence times nicely correspond to geological events that might cause such divergences.

Nested clade analysis of cave crayfish

To explain how the cave species attained high levels of genetic variation, we used NCA to uncover the major historical processes and patterns (Templeton 2001). A statistical parsimony network was constructed using a 95% confidence interval, which resulted in 69 unique haplotypes, thirty-four 1-step clades, fourteen 2-step clades, eight 3-step clades, and three 4-step clades in the total cladogram (Table 4, Fig. 3). The statistical parsimony analysis revealed two haplotypes as ancestral, *O. a. packardi* haplotype 7 and *O. a. australis* haplotype 27, and these are shown as rectangles on Fig. 4. *O. a. packardi* haplotype 8 is connected to *O. incomptus* haplotype 17 by 10 mutational steps (the significant 95% level was nine steps). *Cambarus gentryi* and *C. graysoni* were outside the 95% level, at 21 and 25 mutational steps, respectively, from haplotype 2 of *O. a. packardi*.

To geographically illustrate the historical speciation routes, we used the eight 3-step clades because they mostly resulted in significant inferences of 'contiguous range expansion' or 'isolation by distance' and they show 'big picture' historical biogeographic patterns (Table 5). On Fig. 4, *O. a. packardi* is shown as clades 3-1 (light blue) and clade 3-2 (dark blue) in the network and as circles on the corresponding map, and *O. incomptus* is clade 3-3 (pink) and is represented on the map as pink triangles. Clade 3-4 (green) is *O.* sp. nov. and is marked as green pentagons on the map, while four 3-step clades (3-5 through 3-8) comprise *O. a. australis* (marked as squares on the map of Fig. 4). The 3-step clades of *O. a. australis* geographically overlap extensively in central Tennessee, with several *australis* caves containing haplotypes from different 3-step clades (Table 1).

Table 3 Genealogical estimates of genetic diversity, descendents/ancestor ratio, and divergence time of four stygobitic *Orconectes* species and the ancestral species for each pairwise comparison estimated by IM

	θ	$\theta_{Ancestor}$	Descendents/Ancestor ratio	Time since divergence (in millions of years)
Cave species				
O. a. packardi	14.9498			
	8.82–25.51	16.4765	1.43	282.5
O. incomptus	8.5882	0.06–108.21		125.5–454.5
	2.10–27.29			
O. incomptus	8.8723			
	1.94–34.39	12.4212	1.03	356.1
O. sp. nov.	3.9714	0.08–146.94		102.7–454.4
	0.92–10.05			
O. sp. nov.	7.3871			
	5.56–9.74	11.5073	2.76	110.2
O. a. australis	24.3979	5.56–52.71		105.5–116.4
	21.87–27.11			

Upper values are the maximum-likelihood estimates and the lower values represent the confidence interval range for the 90% highest posterior density. Descendents/Ancestor Ratio = $(\theta_1 + \theta_2)/\theta_{Ancestor}$. A mutation rate of 2.2% per million years was used to determine time since divergence.

Table 4 Results of the nested clade analysis of *Orconectes* 16S mtDNA haplotypes based on 5000 permutations in GEODIS

0-step clades			1-step clades			2-step clades			3-step clades			4-step clades		
Haplotype	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n
1	0	5.6	1-1			2-1	3.7	21.1	3-1	19.2L	18.9	4-1	16.2S	148.7L
2	3.3	3.5												
I-T	3.3	–2.1												
3	3.4	3.8S	1-2	3.8L	3.9L	2-2	3.7S	17.7S						
4	0	4.1												
I-T	3.4	–0.3S												
5	0	4.0	1-3	3.5	3.5S									
6	3.5	3.5												
						I-T	–0.1	–3.4						
7	9.9	9.8	1-4	10.1S	10.3S	2-3	12.5	12.7	3-2	12.6S	13.6S			
8	0	18.8L												
9			1-5	0.0S	22.5L									
10														
11														
			I-T	10.1	–12.2S									
12			1-6			2-4	0.0S	12.1						
13														
14														
15			1-7											
16														
						I-T	12.5L	0.6	I-T	–6.6S	–5.3S			
17			1-8	0.0	10.7	2-5			3-3	9.0S	43.5L	4-2	33.9S	50.4S
19			1-9	0.0	8.3									
18			1-10	0.0S	8.1S									
20														
			I-T	0.0S	–1.0									
22			1-11	0.0	13.6	2-6			3-4	8.4S	34.3			
21	5.6S	10.1	1-12	9.4L	9.5L									
23	4.7S	8.8												
24	1.4	1.6	1-13	1.6S	4.8S									
25	0.0	1.6												
I-T	1.4	0.0	I-T	7.9L	4.1L									
26	0.0	23.1	1-19	21.5	21.6	2-10	20.7S	21.1S	3-6	23.9S	30.7S			
27	18.7	18.7												

Table 4 *Continued*

0-step clades			1-step clades			2-step clades			3-step clades			4-step clades		
Haplotype	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n
28	0.0	0.5												
29	0.0	23.4												
30	0.0	25.4												
I-T	18.4	2.4												
31			1-20	0.0S	27.4									
33														
37			1-24	0.0	24.2									
			I-T	20.1S	–2.2									
32			1-21			2-11	0.0	18.0						
35	5.2	5.4S	1-22	5.6S	5.8	2-12	6.0S	50.5L						
36	0.0	13.0												
I-T	5.2	–7.6S												
34			1-23	0.0	12.7L									
						I-T	15.1L	–27.7S	I-T	9.7	–11.7S			
38			1-32	0.0S	6.2L	2-9	4.2	26.9L	3-7	19.2S	49.1	4-3	47.8S	86.3
39			1-33	0.12	3.1S									
40	6.2	6.2	1-25	6.2	6.2	2-13	6.2S	15.0S						
41	0.0	5.0												
42	0.0	5.0												
44	0.0	5.0												
I-T	6.1	1.2												
43			1-26	0.0	4.9									
45			1-27	0.0	4.9									
			I-T	6.2	1.3									
46	3.5	5.1	1-29	4.6S	4.6S	2-14	6.1S	17.9S	3-8	33.6S	48.2			
47	0.0	3.7												
48	0.0	4.5												
I-T	3.5	0.9												
49	0.0	11.1S	1-28	18.6	17.0									
50	0.0	58.0												
I-T	0.0	–46.9												
51	37.8	35.19	1-31	34.2S	36.5S	2-15	38.6L	36.5L						
53	0.0	22.8												
54	0.8S	31.0												
I-T	37.1L	5.1												
52			1-30	0.0	66.2									
			I-T	34.2S	–29.7S	I-T	–32.5S	–18.5S						
61	8.1	8.1	1-14	8.1S	13.6S	2-7	20.5S	35.4	3-5	35.7S	46.5			
63	0.0	2.8												
64	0.0	15.2												
I-T	8.1	–0.9												
62			1-15	0.0	3.7									
58	0.0	31.2	1-16	26.0	48.5L									
59	0.0	22.2S												
60	0.0	31.2												
I-T	0.0	4.5L	I-T	9.9S	13.5									
57			1-34	0.0	19.1 L	2-8	14.1S	36.1						
55	0.0S	5.5	1-18	5.5	12.1									
56	0.0S	5.5												
I-T	0.0	0.0												
65	13.7	13.9	1-17	14.5	15.0									
66	0.0	14.5L												
67	0.0	12.4												
68	0.0	30.0L												
69	0.0	12.4												
I-T	13.8	–3.4S	I-T	–10.8	–0.65	I-T	–6.4	0.7	I-T	–15.7S	2.0	I-T	–5.1	–53.3S

Clade (D_c) and nested clade (D_n) distances are given. An 'S' indicates the distance is significantly small at the 5% level and an 'L' indicates the distance is significantly large. In clades with both tip and interior nested clades, the average distance I-T is given. Shaded regions indicate interior groupings.

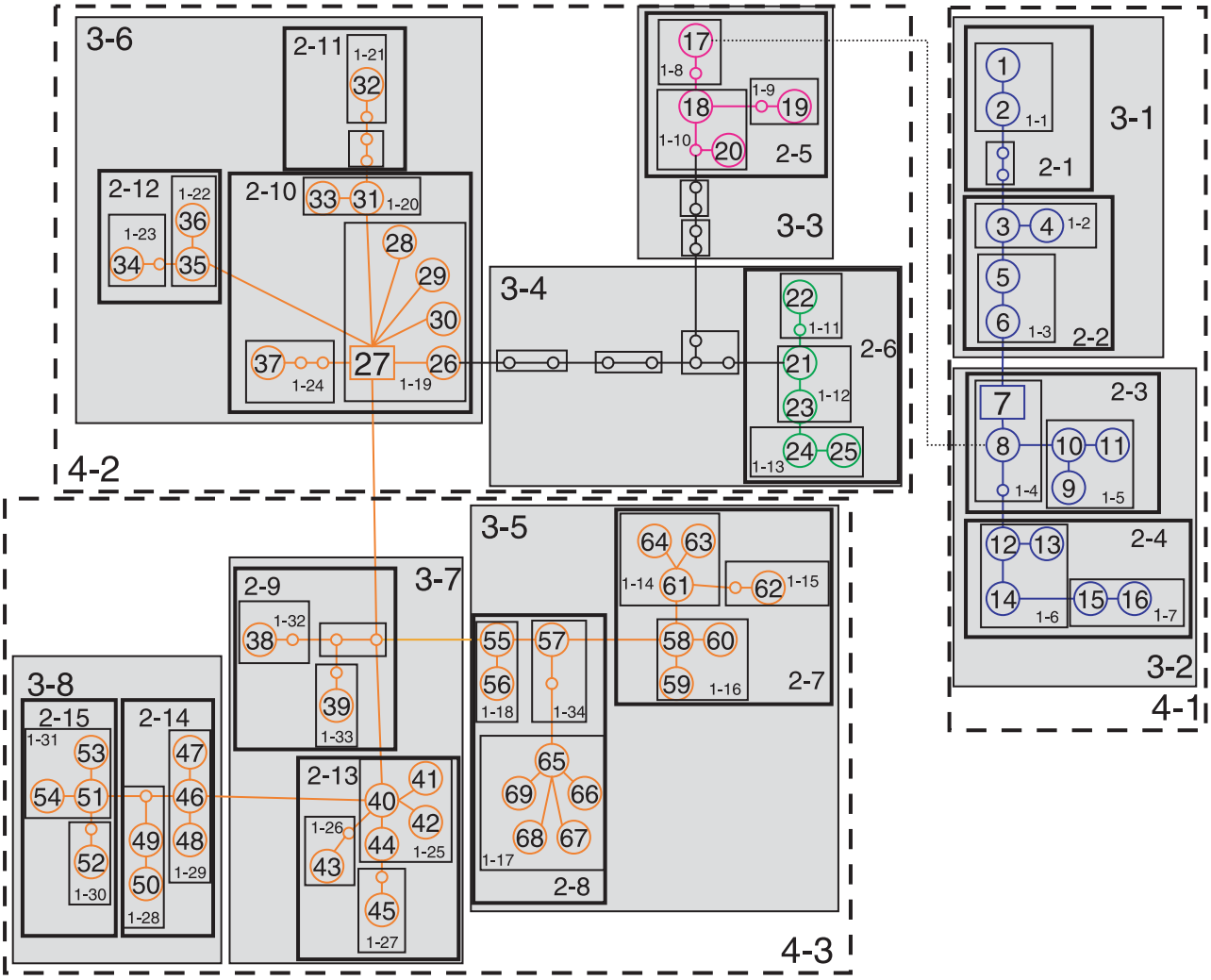


Fig. 3 Haplotype network showing the nesting levels used to infer historical processes. Haplotype circles are coloured to represent four distinct lineages: *Orconectes australis packardi* (blue), *Orconectes incomptus* (pink), *O.* sp. nov. (green), and *Orconectes australis australis* (orange). Empty circles in the network represent unsampled, possibly extinct haplotypes. The total cladogram includes clades 4-1, 4-2, and 4-3.

Discussion

It was hypothesized that the surface ancestor to the cave *Orconectes* originally expanded in a northeast direction from the Mississippi embayment, spawning obligate cave-dwelling species along the Cumberland Plateau en route to the northern Appalachian Mountains (Hobbs & Barr 1972). On the contrary, our phylogenetic and NCA results show that *Orconectes australis packardi*, which is distributed across the northern end of the Cumberland Plateau, is the most basal member of the cave assemblage. This suggests that the surface ancestor (a member of the burrowing genus *Cambarus*) ranged somewhere in eastern Kentucky and gave rise to the stygobitic species *O. a. packardi*. The other stygobitic species then diverged from a common ancestor with *O. a. packardi*. The southern limit of the cave *Orconectes* distribution is the area just north of the Fall Line in Alabama, the prehistoric Atlantic Ocean coastline.

Our estimates of divergence times, although based on one mtDNA region, place the oldest cave *Orconectes* species on the plateau present during the Cretaceous period, which was the suggested time period for cave invasion by surface members of the genus *Cambarus* (Hobbs & Barr 1960). This time frame also correlates with the age estimates of the oldest passages in plateau caves and the beginnings of the eastward recession of the Cumberland Plateau (Barr 1961). It appears that the long evolutionary histories of crayfishes in the stable underground environment have allowed them to persist and accumulate genetic diversity, despite environmental changes on the surface, long generation times, and isolation over the past millions of years. Poulson & White (1969) speculated that older cave

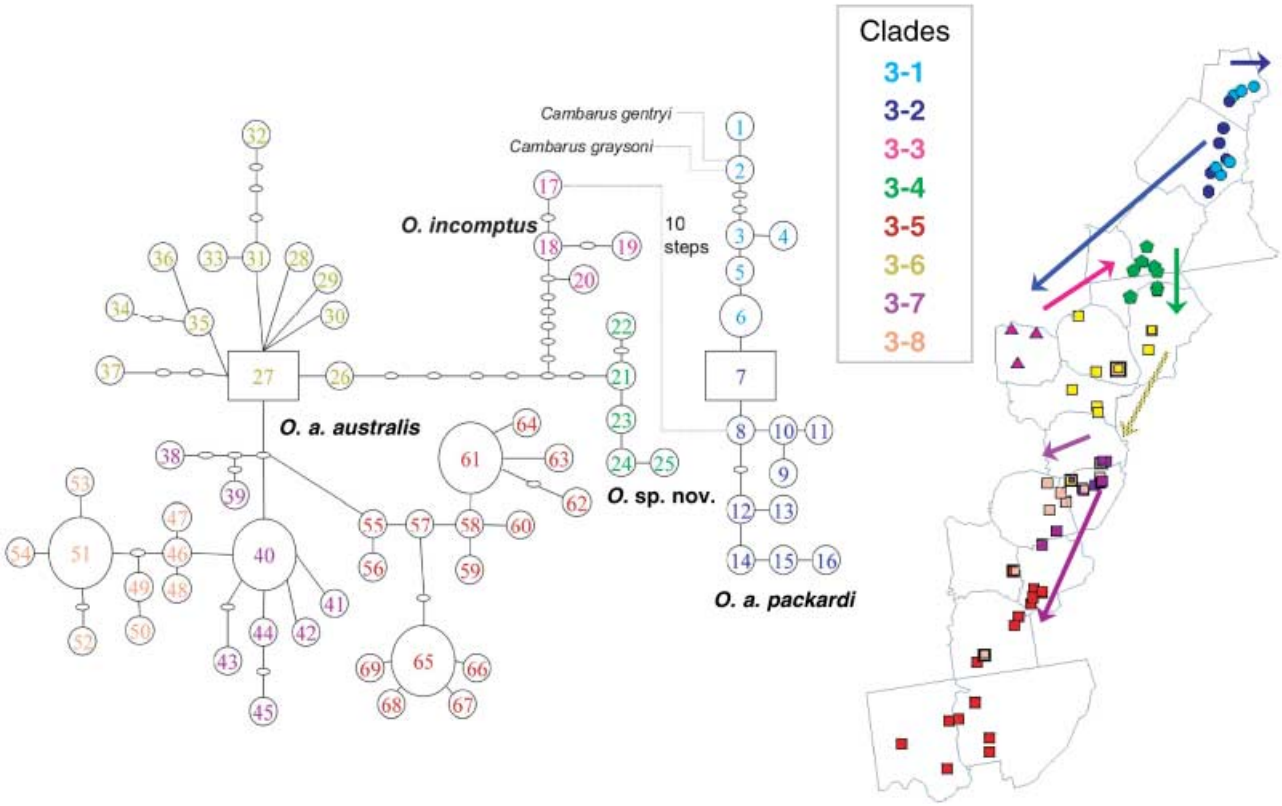


Fig. 4 Haplotype network on left is geographically illustrated using the eight 3-step nested clades, which are corresponding marked by the same colours on the map with grey county outlines to the right. *Orconectes australis packardi* (haplotypes 1–16 in network; circles on map) was outside the 95% confidence limit, while *Orconectes incomptus* (haplotypes 17–20 in network; triangles on map), *Orconectes* sp. nov. (haplotypes 21–25 in network; pentagons on map), and *Orconectes australis australis* (haplotypes 26–69; squares on map) were connected within the 95% confidence level. Coloured arrows on the dot map of sampled caves show routes of contiguous range expansion by the leading-edge expanding clade. Empty circles in the network represent unsampled, possibly extinct haplotypes. The outgroups *Cambarus gentryi* and *Cambarus graysoni* were outside the 95% limit and connected to haplotype 2 of *O. a. packardi*.

species would show low levels of diversity due to the long period of isolation underground, but it appears that levels of diversity for the cave crayfish species are not related to their estimated old divergence times.

One of the arguments made by Culver *et al*. (2000) for the endangered status of cave fauna was restricted geographic ranges, as most United States cave-adapted fauna (61%) are limited to caves in a single county. Although this is a common and practical approach for identifying possible conservation concerns for endemics and rare species as well as habitat types, species-specific information, particularly thorough geographic surveys (Van Jaarsveld *et al*. 1998) and demographic and genetic studies (Lande 1988) are critical pieces of information in assessing the requirements needed for species survival. In this study, *O. a. australis*, with the largest range of the stygobitic *Orconectes*, is now currently known from 11 counties and has the highest genetic diversity of the cave crayfish species; but *O. incomptus*, with the smallest geographic range, and currently only known from nine caves in Jackson County and one cave in Putnam County, Tennessee, has the second highest diversity of the assemblage. *O. a. packardi* is currently known from three Kentucky counties, and *O.* sp. nov. is distributed across four counties in Kentucky and Tennessee, with moderate and low levels of genetic diversity, respectively. In our case, geographic range is not reflective of genetic diversity or conservation status for these cave species. Rather, the decline in genetic diversity over recent history ($\theta_\pi < \theta_W$; Templeton 1993; Sinclair *et al*. 2002; Roman & Palumbi 2003) is a better indicator for conservation concern with *O. a. australis* (currently 0.00894 from historically 0.01593), along with the low levels of diversity for *O.* sp. nov. (currently and historically, 0.00238 and 0.00242). It is interesting that the IUCN stable cave crayfish species, *O. a. australis*, shows a recent loss of diversity, whereas, the two IUCN vulnerable cave species, *O. incomptus* and *O. a. packardi*, show little difference between historical and current diversity estimates.

We show in Fig. 4 a series of colonizations beginning in Kentucky with *O. a. packardi* and progressing down the

Table 5 Nested Contingency Results based on 5000 permutations in GEODIS

Clade	Chi-squared	Probability	Inference chain	Inferred pattern
1-2	0.6857	1.0	1-2-11-17 — No	Inconclusive Outcome
1-4	45.00	0.0002*	1-19-20-2-11-17-4 — No	RGF w/IBD
1-12	26.00	0.0*	1-19-20-2-11-12 — Yes	CRE
1-16	6.00	0.33	1-2-11-17-4 — No	RGF w/IBD
1-17	28.6037	0.17	1-2-11-17 — No	Inconclusive Outcome
1-18	2.000	1.0	1-19-20 — No	Inadequate Geographic Sampling
1-22	1.1953	0.46	1-2-11-17 — No	Inconclusive Outcome
1-28	2.00	1.0	1-19-20-2-11-17 — No	Inconclusive Outcome
1-31	58.9	0.04*	1-2-3-4 — No	RGF w/IBD
2-2	16.354	0.0002*	1-2-11-17–4 — No	RGF w/IBD
2-3	49.0	0.0*	1-19-20-2-11-12 — No	CRE
2-5	16.0	0.0066*	1-19-20 — No	Inadequate Geographic Sampling
2-6	45.385	0.030*	1-2-3-4 — No	RGF w/IBD
2-7	60.093	0.0178*	1-2-3-4 — No	RGF w/IBD
2-8	55.801	0.0136*	1-19-20-2-11-17-4 — No	RGF w/IBD
2-9	7.00	0.05*	1-19-20 — No	Inadequate Sampling
2-10	52.60	0.0174*	1-2-3-4 — No	RGF w/IBD
2-12	1.0588	1.0	1-2-11-17 — No	Inconclusive Outcome
2-14	14.00	0.0238*	1-19-20-2-11-17-4 — No	RGF w/IBD
2-15	27.4909	0.0736	1-2-11-12 — No	CRE
3-1	25.00	0.0*	1-19-20-2-11-12 — No	CRE
3-2	55.13	0.0*	1-2-3-4 — No	RGF w/IBD
3-5	106.27	0.0*	1-2-11-12 — No	CRE
3-6	89.98	0.012*	1-2-3-5-6-7 — Yes	RGF w/some LDD
3-7	57.00	0.0*	1-19-20-2-11-17-4 — No	RGF w/IBD
3-8	55.63	0.0*	1-2-11-12 — No	CRE
4-1	81.044	0.0*	1-19-20-2-11-12 — No	CRE
4-2	257.00	0.0*	1-19-20-2-11-12 — No	CRE
4-3	442.20	0.0*	1-2-11-12 — No	CRE
Total	851.25	0.0*	1-19-20-2-11-12 — No	CRE

*indicates significance at the $P < 0.05$ level. Inferences were made using with Templeton's (2004) revised key. Abbreviations for the inferences are: CRE, contiguous range expansion; RGF, restricted gene flow; IBD, isolation by distance; LDD, long-distance dispersal.

Cumberland Plateau in a leading-edge small-stepwise manner, following the flow of prehistoric waters. This colonization pattern is consistent for animal groups limited by mountain landscapes and by dispersal ability, particularly in response to glacial advance and retreat cycles (Hewitt 1996, 2000). Stygobitic crayfishes are severely limited in dispersal abilities by both subterranean and surface barriers, except during high water levels when they can migrate (or wash) out of caves into a limestone-based surface stream across short distances, and into a nearby underground system via a spring resurgence or cave entrance. These findings suggest that prehistoric groundwater levels were much higher, and allowed for subterranean fauna to disperse over the surface landscape in small distances. Phreatic caves form below the water table, and as karst dissolves and creates voids, the water table lowers to fill in the spaces, which increases groundwater habitat for stygobites (White 1988). Although the major surface rivers along the Cumberland Plateau historically and currently flow in a southern direction, ongoing cave development and subsequent groundwater lowering have probably lead to isolation by distance and the prevention of further stepwise range expansion of the species and clades.

Contiguous range expansion followed by periods of isolation appears to be the main mechanism for the increased variation within the cave crayfish species. A similar trend has been reported for invasive and introduced species (Tsutsui *et al*. 2000; Kolbe *et al*. 2004) in which genetic diversity and population size accumulates and recovers, rather than resulting in a series of bottlenecks leading to lower diversity and extirpation. One example (Sbordoni 1982) has also been documented for a troglobitic beetle species in Italy in which 50 individuals were introduced into an isolated cave with no beetles. After 30 years, the estimated census size was 15 000 individuals with a greater genetic diversity than the original 'founder' population.

Clearly, pre-adaptation and continued expansion into suitable habitat of the subterranean environment allowed cave crayfish to successfully and repeatedly colonize new areas, regardless of population size or genetic diversity of the founder populations.

Orconectes australis packardi, *O. incomptus*, and *O.* sp. nov. are currently distributed across small geographic ranges (four counties or less), possibly due to the hydrologic impacts of the prehistoric watercourses of the Cumberland River. Caves in the path of the Cumberland River during its formation would have been completely submerged by surface waters. The missing haplotypes in the parsimony network may be evidence of past drainage evolution events between the ancestors of *O. a. packardi* and *O. incomptus*, and *O. incomptus* and *O.* sp. nov. leading to local extirpations, range restrictions, and lower diversity in those species compared to *O. a. australis*.

Orconectes luteus and *O. juvenilis* are currently listed as IUCN stable species in conservation status based on the fact that they are widespread throughout their ranges (Taylor *et al*. 1996), but it appears that they are in need of some protection and study (based on the large discrepancy between θ_π and θ_W for both common surface dwellers). The stable underground environment may provide enough suitable 'habitat pockets and hideouts' to buffer the subterranean biota from the direct impacts of ongoing surface pressures, but it appears that the surface species are not so fortunate. It is surprising that species considered to be common stream inhabitants show a reduction in population sizes whereas most of the cave species show consistent population sizes over evolutionary time.

We also hope that these findings shed light on the conservation status of other subterranean taxa and propel biospeleogists to test their assumptions concerning biodiversity. We suggest that management strategies be redirected toward molecular genetic assessments of effective population sizes and diversity (Thorpe *et al*. 1995) for cave species and other elusive fauna considered to be on the brink of extinction because of a lack of scientific information (Holmes 2001). Current cave conservation activities focus on general efforts to protect subterranean habitat by purchasing karstlands, avoiding pollution catastrophes, and gating highly visible entrances. Although these are important defences for the protection of the biome, the ultimate goal of cave conservation is the sustainability of each unique obligate cave-dwelling species. Stochastic factors are well-known causes of biodiversity losses, yet, current research shows that the genetic factors, specifically loss of heterozygosity and inbreeding, can play major roles in driving endangered and threatened species to extinction (Brook *et al*. 2002; Spielman *et al*. 2004). We hope this research will turn the efforts of conservation agencies toward protecting gene flow routes and areas of connectivity to prevent future imperilment of the amazing fauna under our feet and the common inhabitants in our backyards.

Acknowledgements

Many dedicated cavers assisted with the fieldwork, especially N. Mann, G. Moni, M. Niemiller, W. Simpson, K. Toepke, W. Walden, and D. Withers, as well as the local caving groups of the National Speleological Society: Dayton Underground, Central Ohio, Greater Cincinnati, Spencer Mountain, Nashville, Golden Pond, Huntsville, and the Tennessee Cave Survey. We are also grateful to the Tennessee Department of Conservation for landowner reconnaissance and transportation and to Alabama, Tennessee, and Kentucky wildlife agencies for collecting permits. Special thanks to the Elmore, Evans, and Gulley households for part-time residence. T. Barr and J. Cooper provided valuable insight and continual support, and M. Achtman reviewed an earlier draft. J. Fetzner Jr aided with DNA extractions and provided *O. luteus* data. R. Ziemba provided *O. juvenilis* data. This project was funded by the Cave Research Foundation, National Speleological Society, University of Alabama Graduate School and Biology Department, Brigham Young University Graduate Studies and Integrative Biology Department, and United States Fish and Wildlife to J.E.B., and the National Science Foundation (DDIG DEB 0508580) to K.A.C. and J.E.B.

References

Avise JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology*, **7**, 45–67.

Avise JC, Selander RK (1972) Genetics of cave-dwelling fishes of the genus *Astyanax*. *Evolution*, **26**, 1–19.

Barr TC Jr (1961) The caves of Tennessee. *Tennessee Division of Geology Bulletin*, **64**, 1–567.

Barr TC Jr (1968) Cave ecology and the evolution of troglobites. In: *Evolutionary Biology* (eds Dobzhansky T, Hecht M, Steere WC), vol. 2, pp. 35–102. Appleton-Century-Crofts, New York.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to Plant Biosystematics* (eds Hoch PC, Stephenson AG), pp. 289–303. Missouri Botanical Garden, St. Louis, Missouri.

Brook BW, Tonykn DW, O'Grady J, Frankham R (2002) Contribution of inbreeding to extinction risk in threatened species. *Conservation Ecology*, **6**, 16.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cooper JE (1975) *Ecological and behavioral studies in Shelta Cave, Alabama with emphasis on decapod crustaceans*. PhD Dissertation, University of Kentucky.

Cooper JE, Cooper MR (1997) New troglobitic crayfish of the genus *Orconectes*, subgenus *Orconectes* (Decapoda: Cambaridae), endemic to Shelta Cave, Huntsville, Alabama. *Journal of Cave and Karst Studies*, **59**, 119–127.

Crandall KA (1997) Genetic variation within and among crayfish species. *Freshwater Crayfish*, **11**, 135–145.

Crandall KA, Fitzpatrick JF Jr (1996) Crayfish molecular systematics: using a combination of procedures to estimate phylogeny. *Systematic Biology*, **45**, 1–26.

Crandall KA, Posada D, Vasco D (1999) Effective population sizes: missing measures and missing concepts. *Animal Conservation*, **2**, 317–319.

Culver DC, Master LL, Christman MC, Hobbs HH III (2000) Obligate cave fauna of the 48 contiguous United States. *Conservation Biology*, **14**, 386–401.

Cunningham CW, Blackstone NW, Buss LW (1992) Evolution of king crabs from hermit crabs ancestors. *Nature*, **355**, 539–542.

Danielopol DL, Griebler C, Gunatilaka A, Notenboom J (2003) Present state and future prospects for groundwater ecosystems. *Environmental Conservation*, **30**, 104–130.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews in Genetics*, **5**, 702–712.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Fetzner JW Jr (1996) Biochemical systematics and evolution of the crayfish genus *Orconectes* (Decapoda: Cambaridae). *Journal of Crustacean Biology*, **16**, 111–141.

Fetzner JW Jr, Crandall KA (2003) Linear habitats and the nested clade analysis: an empirical evaluation of geographic vs. river distances using an Ozark crayfish (Decapoda: Cambaridae). *Evolution*, **57**, 2101–2118.

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLOS Biology*, **3**, e193.

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**, 182–192.

Hobbs HH III (1991) Decapoda. In: *Ecology and Classification of North American Freshwater Invertebrates* (eds Thorp JH, Covich AP), pp. 823–858. Academic Press, San Diego, California.

Hobbs HH III, Bagley FM (1989) *Shelta Cave Management Plan, Biology Subcommittee of the Shelta Cave Management Committee*. 78 p.

Hobbs HH Jr, Barr TC Jr (1960) The origins and affinities of the troglobitic crayfishes of North America (Decapoda, Astacidae). I. The genus *Cambarus*. *American Midland Naturalist*, **64**, 12–33.

Hobbs HH Jr, Barr TC Jr (1972) Origins and affinities of the troglobitic crayfishes of North America (Decapoda: Astacidae). II. Genus *Orconectes*. *Smithsonian Contributions to Zoology*, **105**, 1–84.

Hobbs HH Jr, Hobbs HH III, Daniel MA (1977) A review of the troglobitic decapod crustaceans of the Americas. *Smithsonian Contributions to Zoology*, **244**, 1–176.

Holmes EE (2001) Estimating risks in declining populations with poor data. *Proceedings of the National Academy of Sciences, USA*, **98**, 5072–5077.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Kane TC (1982) Genetic patterns and population structure in cave animals. In: *Environmental Adaptation and Evolution* (eds Mossakowski D, Roth G), pp. 131–149. Gustav Fisher & Stuttgart, New York.

Kolbe JJ, Glor RE, Schettino LR, Lara AD, Larson A (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Koppelman JB, Figg DE (1995) Genetic estimates of variability and relatedness for conservation of an Ozark cave crayfish species complex. *Conservation Biology*, **9**, 1288–1294.

Kuhner MK, Yamato J, Beerli P (2004) LAMARC *V.1.2.1*. University of Washington. http://evolution.gs.washington.edu/lamarc.html.

Lande R (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455–1460.

Master LL (1991) Assessing threats and setting priorities for conservation. *Conservation Biology*, **5**, 559–563.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Pearse DE, Crandall KA (2004) Beyond F_{ST}: analysis of population genetic data for conservation. *Conservation Genetics*, **5**, 585–602.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Poulson TL, White WB (1969) The cave environment. *Science*, **165**, 971–981.

Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science*, **301**, 508–510.

Ronquist F, Huelsenbeck JP (2003) MRBAYES3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rozas J, Sánchez-DelBarrio JC, Messegyer X, Rozas R (2003) DNASP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Sbordoni V (1982) Advances in speciation of cave animals. In: *Mechanisms of Speciation* (ed. Barigoozi C), pp. 219–240. Alan R. Liss, New York.

Sinclair EA, Costello B, Courtenay JM, Crandall KA (2002) Detecting a genetic bottleneck in Gilbert's Potoroo (*Potorous gilbertii*) (Marsupialia: Potoroidae), inferred from microsatellite and mitochondrial DNA sequence data. *Conservation Genetics*, **3**, 191–196.

Sinclair EA, Fetzner JF Jr, Buhay JE, Crandall KA (2004) Proposal to complete a phylogenetic, taxonomy, and systematic revision for freshwater crayfish (Astacidae). *Freshwater Crayfish*, **14**, 21–29.

Sites JW Jr, Crandall KA (1997) Testing species boundaries in biodiversity studies. *Conservation Biology*, **11**, 1289–1297.

Sites JW Jr, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution*, **18**, 462–470.

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences, USA*, **101**, 15261–15264.

Swofford DL (2001) PAUP*: *Phylogenetic Analysis Using Parsimony* (and Other Methods)*, Version 4.0b7 Beta. Sinauer Associates, Sunderland, Massachusetts.

Swofford DL, Branson BA, Sievert GA (1980) Genetic differentiation of cavefish populations. *Isozyme Bulletin*, **13**, 109–110.

Taylor CA, Warren ML Jr, Fitzpatrick JF Jr. *et al.* (1996) Conservation status of crayfishes of the United States and Canada. *Fisheries*, **21**, 25–38.

Templeton AR (1993) The 'Eve' hypothesis: a genetic critique and reanalysis. *American Anthropologist*, **95**, 51–72.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, **10**, 779–791.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **4**, 789–809.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographic distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.

Thorpe JP, Smartt J, Allcock AL *et al*. (1995) Genetic diversity as a component of biodiversity. In: *Global Biodiversity Assessment* (ed. Heywood VH), pp. 57–87. Cambridge University Press, New York.

Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences, USA*, **97**, 5948–5953.

Van Jaarsveld AS, Freitag S, Chown SL *et al*. (1998) Biodiversity assessment and conservation strategies. *Science*, **279**, 2106–2108.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theories in Population Biology*, **7**, 256–276.

White WB (1988) *Geomorphology and Hydrology of Karst Terrains*. Oxford Press, New York.

Won Y-J, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

Yu N, Jensen-Seaman MI, Chemnick L *et al*. (2003) Low nucleotide diversity in chimpanzees and bonobos. *Genetics*, **164**, 1511–1518.

This study represents part of Jen Buhay's doctoral research on the evolutionary history of cave crayfishes with Keith Crandall. Buhay's research interests focus on population genetics and phylogeography of surface and subterranean freshwater fauna. Keith Crandall is a Professor in the Integrative Biology and Molecular Biology Departments at Brigham Young University. Crandall works on a wide variety of organisms in an evolutionary context.

Animal Conservation. Print ISSN 1367-9430

Surface to subsurface freshwater connections: phylogeographic and habitat analyses of *Cambarus tenebrosus*, a facultative cave-dwelling crayfish

J. B. Finlay[1], J. E. Buhay[1] & K. A. Crandall[1,2]

1 Department of Integrative Biology, Brigham Young University, Provo, UT, USA
2 Monte L. Bean Life Science Museum, Brigham Young University, Provo, UT, USA

Keywords
crustacean; invasive species; cave; subterranean; population genetic.

Correspondence
Jennifer E. Buhay, Department of Integrative Biology, Brigham Young University, 401 Widtsoe Building, Provo, UT 84602-5255, USA.
Email: crayfish@byu.edu

Received 18 November 2005; accepted 26 April 2006

doi:10.1111/j.1469-1795.2006.00046.x

Abstract

This study examined the phylogeography and population demographics of *Cambarus tenebrosus*, which has an unusually large distribution for a freshwater crayfish species, encompassing the Interior Lowlands and Cumberland Plateau of the eastern United States. This facultative cave-dweller provides a unique perspective on the biologic connections between surface and subsurface freshwater ecosystems, which are considered to be highly imperiled due to pollution and habitat degradation. The 16S mitochondrial gene was sequenced for 233 individuals from 84 cave and 20 surface locations throughout the range, with most sampling concentrated around the Cumberland Plateau of the southern Appalachians, to assess conservation status of this species and examine the extent of gene flow between the two habitat types. Cave and surface populations formed a single monophyletic group relative to *Cambarus striatus*, and clades showed strong geographical associations, but lacked habitat structuring. Occupation of subterranean environments does not appear to be a recent event in the evolutionary history of the species. The large amount of genetic diversity within the species, coupled with its ability to inhabit surface and subsurface environments, suggests that this species may pose a threat as a possible invasive species in other karst-dominated landscapes.

Introduction

The number of faunal extinctions occurring in North American freshwater environments has been steadily increasing (Master, 1990; Williams *et al*., 1993; Taylor *et al*., 1996; Schuster, 1997), and it has been estimated that the number of freshwater species in North America is decreasing at a rate of 4% per decade, which rivals extinction rates in tropical rainforests (Ricciardi & Rasmussen, 1999). Elevated extinction rates of freshwater fauna are typically associated with habitat destruction, organic pollution, stream regulation by dams and habitat fragmentation (Neves *et al*., 1997; Ricciardi, Neves & Rasmussen, 1998), yet current research also suggests that genetic factors play important roles in driving threatened and endangered species to extinction (Spielman, Brook & Frankham, 2004). Thus, it is important that the protection of freshwater environments be approached not only by reducing the impact of humans on the aquatic environment, but also by investigating the population structures and connectivity of its inhabitants using molecular assessments of conservation status.

The Nature Conservancy considers 95% of subterranean species in North America to be endangered or imperiled (Master, 1991; Culver *et al*., 2000). There is little doubt that subsurface groundwater fauna are threatened by surface pollution and habitat deterioration (Danielopol *et al*., 2003), but studies of aquatic cave organisms are sparse and often inconclusive, further adding to the enigmatic nature of the subterranean environment. Furthermore, information about the biological connections between surface and subsurface environments is lacking, and this study is the first species-specific genealogical investigation of any North American stygophilic (aquatic facultative cave-dwelling) species.

Cambarus tenebrosus (Hay, 1902) is unusual among freshwater crayfish species because it occupies both epigean (surface) and hypogean (subsurface) karst habitats. *Cambarus tenebrosus* also has a large range for a crayfish, extending from south-central Indiana southward to northern Alabama (Fig. 1). Because it is found in subterranean habitats typically occupied by obligate cave dwellers (stygobites), it was originally thought that *C. tenebrosus* was a transient member of the cave environment, perhaps being washed into the cave by accident. Hay (1902) refuted this hypothesis based partially on morphological characteristics indicative of stygobitic crayfishes, including the presence of reduced eyes and elongated limbs, which *C. tenebrosus*

possesses. These morphological characteristics, collectively referred to as troglomorphy, suggest that *C. tenebrosus* has partially adapted to subterranean life and, therefore, is not a passing member of the underground environment. A previous morphological study of *C. tenebrosus* showed no difference between individuals collected from surface and subsurface sites, but reflected overall intraspecific phenotypic plasticity (Taylor, 1997). This morphological plasticity might be caused by convergence due to similar environmental pressures in conjunction with active gene flow between the surface and cave habitats (Wiens, Chippindale & Hillis, 2003).

The objectives of this project were to (1) establish whether *C. tenebrosus* shows intraspecific geographic structuring of genetic variation, (2) test if there is a significant genetic association with the two habitats the crayfish occupies (cave vs. surface), and (3) provide molecular-based estimates of genetic diversity and effective population size for the species.

Materials and methods

Population sampling

Samples were collected at 104 sites (84 cave and 20 surface) throughout the range of *C. tenebrosus*, concentrating on areas of the Cumberland Plateau of the southern Appalachians and the Interior Lowlands which range from south-central Indiana to northern Alabama (Fig. 1; Table 1). A sample was considered subterranean or 'cave' if it was collected from an area not lit by natural light. Samples were included from the type locality at Mammoth Cave in Kentucky. In most cases, a non-destructive method of sampling was used, which involved collecting a leg from each individual and then returning the individual to the place of capture. Crayfish have the ability to regenerate lost limbs and, therefore, removing a limb during capture is not detrimental to the animal's survival (crayfish often lose their limbs in territorial battles). Tissue samples were stored in 95% ethanol, and each sample was given a unique identification number. Latitude and longitude coordinates were taken by a global positioning system (GPS) device at each sample site, including entrances to sampled caves. In a few cases, voucher specimens were taken and deposited at the Monte L. Bean Museum (BYU) and the North Carolina State Museum of Natural History. Additionally, *Cambarus striatus* (Hay, 1902), a closely related species (J. E. Buhay & K. A. Crandall, unpubl. data), was used as the outgroup to root the phylogenetic tree and haplotype network. The network analysis clearly shows *C. striatus* to be outside the 95% confidence interval for *C. tenebrosus* and phylogenetic analysis of the genus shows this species to be the sister taxon to *C. tenebrosus* (J. E. Buhay & K. A. Crandall, unpubl. data), making it an appropriate outgroup for this analysis.

DNA sequencing

DNA was extracted from the samples using a cell lysis protocol (Crandall *et al.*, 1999). The protocol called for 5–15 mg of vacuum-dried tissue to be placed in a tube with 800 μL of cell lysis solution (1.21 g Tris, 37.1 g EDTA, 20 g SDS per liter, pH 8.0). Nine microliters of proteinase K (20 mg mL^{-1}) was added to this solution and the samples were incubated overnight at 55 °C while mixing on a shaker for tissue digestion. After 180 μL of 5 M NaCl was added, the mixture was vortexed and centrifuged to pellet out the salt. The supernatant was transferred to a clean cryotube. Immediately after, 420 μL of ice cold isopropanol was added and this mixture was centrifuged at 20 000 *g* for 10 min to pellet the DNA. After discarding the supernatant, the DNA pellet was washed with 500 μL of 70% ethanol using a cell

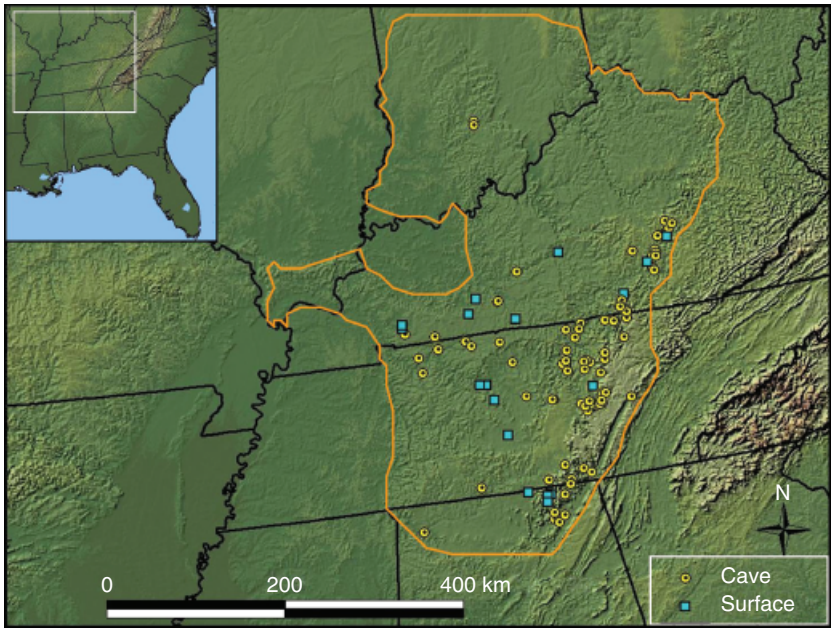


Figure 1 Outlined range (in orange, adapted from Taylor, 1997) of *Cambarus tenebrosus* extending from central Indiana to northern Alabama. Blue dots represent surface collection sites whereas yellow dots represent cave collection sites.

Table 1 Locations and sample sizes for all *Cambarus tenebrosus* used in this study

Site	Location	Habitat	State	Country	Sample size	Haplotype (no. of individuals)
1	Arthur Singleton	Cave	KY	Rockcastle	2	1(2)
2	Bakers	Cave	TN	Robertson	1	1(1)
3	Bartlett	Cave	TN	Putnam	4	1(3), 35(1)
4	Beaver Creek	Surface	TN	Wayne	1	47(1)
5	Bellamy	Cave	TN	Montgomery	3	1(3)
6	Bible Springs	Cave	TN	Marion	2	31(1), 32(1)
7	Big Bush Creek	Surface	KY	Green	2	1(1), 13(1)
8	Big Sink	Cave	KY	Pulaski	2	1(2)
9	Big Sulphur	Cave	KY	Trigg	2	3(2)
10	Blackpatch Hollow	Cave	TN	Robertson	3	1(2), 4(1)
11	Blind Fish	Cave	TN	Putnam	1	37(1)
12	Bluehole Resurgence	Cave	KY	Rockcastle	4	1(4)
13	Bluff River	Cave	AL	Jackson	2	49(1), 51(1)
14	Boone Hollow	Cave	TN	Clay	2	1(2)
15	Browns Creek	Surface	TN	Davidson	2	21(1), 22(1)
16	Bunkum	Cave	TN	Pickett	2	9(1), 10(1)
17	Camps Gulf	Cave	TN	Van Buren	2	37(1), 40(1)
18	Capshaw	Cave	TN	Putnam	4	37(4)
19	Car Parts	Cave	KY	Rockcastle	2	1(1), 5(1)
20	Cedar Creek	Cave	KY	Pulaski	4	1(4)
21	Cherry	Cave	TN	Jackson	1	11(1)
22	Climax	Cave	KY	Rockcastle	1	14(1)
23	Clinton	Cave	TN	Pickett	4	1(4)
24	Cornstarch	Cave	TN	Fentress	1	5(1)
25	Cummings Cove	Surface	TN	Van Buren	3	37(2), 58(1)
26	Dave's	Cave	KY	Pulaski	1	5(1)
27	Dillions	Cave	IN	Orange	2	56(1), 57(1)
28	Doug Green	Cave	AL	Jackson	1	27(1)
29	Dripping Spring	Cave	TN	Franklin	2	62(2)
30	Dumpling	Cave	KY	Pulaski	2	1(1), 5(1)
31	Dunbar	Cave	TN	Montgomery	1	5(1)
32	Duvalts	Cave	KY	Rockcastle	1	6(1)
33	Edmonson Branch	Surface	TN	Davidson	3	1(2), 46(1)
34	England Cove	Surface	TN	White	5	17(2), 20(1), 37(1), 44(1)
35	Estill Fork	Surface	AL	Jackson	3	33(1), 34(2)
36	Fancher	Cave	TN	Overton	3	1(1), 38(2)
37	Fletcher Spring	Cave	KY	Rockcastle	4	1(2), 7(2)
38	Flynn Creek	Cave	TN	Jackson	1	1(1)
39	Gallatin Steam Plant	Cave	TN	Wilson	1	1(1)
40	Garner Spring	Cave	TN	Franklin	2	33(2)
41	Garretts Mill	Cave	TN	Overton	3	40(3)
42	Grayson Gunner	Cave	KY	Wayne	3	1(3)
43	Hail	Cave	KY	Pulaski	3	1(2), 12(1)
44	Herring	Cave	TN	Rutherford	4	1(3), 8(1)
45	Hester Creek	Surface	AL	Madison	1	59(1)
46	Jared Hollow	Cave	TN	Putnam	3	1(2), 24(1)
47	John Griffin	Cave	KY	Jackson	1	14(1)
48	Kuykendall	Cave	TN	Putnam	12	37(7), 43(5)
49	Larkin Fork	Surface	AL	Jackson	1	33(1)
50	Larkin Spring	Cave	AL	Jackson	3	49(3)
51	Laurel Creek	Cave	TN	Van Buren	1	45(1)
52	Lick Fork	Surface	AL	Jackson	1	61(1)
53	Limrock Blowing	Cave	AL	Jackson	2	49(1), 50(1)
54	Lost Cove	Cave	TN	Franklin	1	33(1)
55	Lost Cove	Cave	TN	Van Buren	1	40(1)
56	Lost Creek	Cave	TN	White	3	1(2), 9(1)
57	Lost River	Cave	KY	Warren	2	14(2)
58	Mammoth	Cave	KY	Edmonson	3	1(2), 3(1)

Table 1 Continued

Site	Location	Habitat	State	Country	Sample size	Haplotype (no. of individuals)
59	Manning Spring	Cave	TN	Cumberland	1	60(1)
60	Markham	Cave	TN	Clay	1	1(1)
61	Martin Creek	Surface	TN	Putnam	3	9(2), 36(1)
62	McBrides	Cave	AL	Jackson	5	33(3), 49(2)
63	McKinney Pit	Cave	AL	Colbert	1	28(1)
64	Merrybranch	Cave	TN	White	2	1(1), 9(1)
65	Miller	Cave	TN	Warren	3	29(1), 30(1), 52(1)
66	Moore's Spring	Cave	TN	Giles	1	27(1)
67	Mud River	Surface	KY	Logan	2	1(2)
68	Muddy Creek	Surface	KY	Logan	1	1(1)
69	Natural Bridge	Cave	TN	Pickett	1	1(1)
70	North Fork Creek	Surface	TN	Bedford	2	23(2)
71	Norton Spring	Cave	TN	Warren	2	33 (1), 53(1)
72	Pearson Spring	Cave	TN	Franklin	3	26(1), 33(2)
73	Pennywinkle Spring	Cave	TN	Van Buren	1	40(1)
74	Pitman Creek	Surface	KY	Pulaski	1	48(1)
75	Pless	Cave	IN	Lawrence	1	56(1)
76	Pond Cave	Cave	TN	Cannon	1	1(1)
77	Price Valley	Cave	KY	Pulaski	2	1(2)
78	Redmond Creek	Cave	KY	Wayne	3	1(2), 5(1)
79	Richland Creek	Surface	TN	Davidson	2	18(1), 19(1)
80	Roundstone Creek	Surface	KY	Rockcastle	3	1(1), 7(1), 15(1)
81	Rumbling Falls	Cave	TN	Van Buren	2	37(2)
82	Sauta	Cave	AL	Jackson	2	49(2)
83	Sheldon	Cave	AL	Jackson	2	49(2)
84	Short Creek	Cave	KY	Pulaski	2	7(2)
85	Sinking Fork	Surface	KY	Trigg	1	2(1)
86	Skillmans Mark	Cave	TN	Fentress	1	5(1)
87	Spring at Fahey	Cave	TN	Putnam	1	25(1)
88	Spring off Little Creek	Cave	TN	Putnam	3	37(1), 41(1), 42(1)
89	State Trooper	Cave	KY	Warren	3	14(3)
90	Steele Branch	Surface	KY	Trigg	8	1(8)
91	Stout	Cave	TN	Putnam	2	35(1), 39(1)
92	Stream	Cave	KY	Wayne	4	1(2), 5(1), 16(1)
93	Sump Jump	Cave	TN	Robertson	6	1(6)
94	Thorp	Cave	TN	Clay	2	1(2)
95	Tonyas	Cave	KY	Wayne	2	1(2)
96	Trammel Creek	Surface	KY	Allen	1	1(1)
97	Trick or Treat	Cave	TN	Putnam	2	1(2)
98	Turkeyscratch	Cave	TN	Warren	2	33(1), 40(1)
99	Turner	Cave	TN	Houston	1	1(1)
100	Twin Levels	Cave	KY	Christian	1	5(1)
101	Upper Sheep	Cave	TN	White	1	1(1)
102	Waterfall Hollow	Cave	TN	Van Buren	1	37(1)
103	West Cemetery	Cave	TN	Putnam	1	37(1)
104	Winching Hollow Water	Cave	TN	Van Buren	3	37(1), 43(1), 44(1)

rotator for 1 h. The supernatant was removed and the DNA pellet was vacuum dried for 15 min at 55 °C, and then the pellet was re-suspended in 200 μL of double-distilled water.

The 16S mitochondrial gene was sequenced for all samples because it is highly variable and appropriate for population genetic or intraspecific studies (Fetzner & Crandall, 2003; Buhay & Crandall, 2005). The following reactants were used in each of the 50 μL reactions: 5 μL 10 × buffer, 8 μL dNTPs, 8 μL 25 mM magnesium chloride, 5 μL of each 10 mM primer, 0.3 μL Taq Polymerase and 1.5 μL DNA with water added to total 50 μL. The primers used were 16SF (5′ GAC CGT GCK AAG GTA GCA TAA TC 3′) and 1472 (5′ AGA TAG AAA CCA ACC TGG 3′) (Crandall & Fitzpatrick, 1996). Polymerase chain reaction (PCR) was performed on a Peltier Thermal Cycler machine (AB9800, Foster City, CA, USA) or a GeneAmp PCR System 9700 (AB9700, Foster City, CA, USA) using the following program: 96 °C for 3 min, followed by 45 cycles of 94 °C for 1 min, annealing between 45 and 47 °C for 1 min, and 72 °C for 1 min, followed by a final elongation at 72 °C

for 5 min. PCR products were examined on a 1.5% agarose gel using an ethidium bromide stain. The PCR products were purified using a Montage PCR_{96} plate (Millipore, Billerica, MA, USA). The PCR products were cycle sequenced using the ABI Big-dye Ready-Reaction kit with 1/4 or 1/8 of the normal reaction size, and sequences were generated on an Applied Biosystems (Foster City, CA, USA) 3730 XL Automated Sequencer at the BYU DNA Sequencing Center. Resulting sequences were edited using Sequencher 4.2 OS X (Gene Codes Corporation, Ann Arbor, MI, USA) and aligned by eye using MacClade 4.05 OS X (Madison & Madison, 2000).

Delimiting species

Although methods of diagnosing species remain a controversial issue in systematic biology (Sites & Marshall, 2003, 2004), they are highly relevant to conservation studies (Sites & Crandall, 1997) because the method of delimitation can have a significant impact on the number of species diagnosed (Agapow *et al*., 2004). We prefer a statistically testable method developed for use with molecular data for our study. Templeton's test of cohesion (Templeton, 1989) uses both historical and current processes to statistically delimit species boundaries through a suite of nested null hypotheses. The hypotheses are then used to determine correlations between genotype and geographic location, habitat or other ecological variables [nested clade analysis (NCA); Templeton, Routman & Phillips, 1995]. Under this definition, two organisms would be considered a single species if they are genetically and/or ecologically exchangeable (Templeton, 2001; Rader *et al*., 2005).

Phylogenetic analysis

The model of evolution that best fits the sequence data was determined using the program ModelTest 3.06 (Posada & Crandall, 1998), with the unique 16S haplotypes determined by TCS 1.18 (Clement, Posada & Crandall, 2000). A Bayesian phylogeny was obtained using MrBayes v3.0b4 (Huelsenbeck & Ronquist, 2001; Huelsenbeck *et al*., 2001) with over 20 Markov chains run simultaneously using only unique haplotypes, with each chain initiating at a random tree and parameters nst = 6 and rates = adgamma provided by ModelTest. This analysis was run for 20 million generations on 20 processors on a 64-node RackSaver computing cluster, taking samples from the chain every 1000th tree, totaling 20 001 trees. Using the sampled trees minus the burn-in determined by Tracer (http://evolve.zoo.ox.ac.uk/software.html), a majority-rule consensus tree was constructed. A posterior probability of 95% or greater is considered to be strong Bayesian support for a node (Huelsenbeck & Ronquist, 2001).

Genetic–geographic associations

NCA allows the partitioning of current population parameters (e.g. recent gene flow) from historical events (e.g. range expansion). NCA is a statistical approach that distinguishes among alternative hypotheses to explain contemporary and historical genetic patterns using haplotype diversity information coupled with geographic location information (Templeton *et al*., 1995; Templeton, 1998; Avise, 2000). Inferences about genetic patterns can be made by testing a null hypothesis of no association between the collecting locale and the genetic variability (Templeton *et al*., 1995).

To perform an NCA, a haplotype network was constructed using TCS 1.18 set at a 95% confidence level. The original haplotype network contained several loops, which would be ambiguous in the NCA. These loops were broken using the protocol of Crandall & Templeton (1993) and Templeton & Sing (1993), where the number of sequences in a haplotype and geographic location were most heavily considered. The network was then converted into a series of nesting groups (Templeton, Boerwinkle & Sing, 1987), with the haplotypes exhibiting the highest sequence frequency and most connections being ancestral to the others. According to coalescent theory, haplotypes found at the tips are more recently evolved than those in the interior of the network (Crandall & Templeton, 1993; Templeton, 2004).

To test the null hypothesis of no geographical association, two measurements were calculated by the program GeoDis 2.2 (Posada, Crandall & Templeton, 2000). The first is 'clade distance' (D_c), which measures the geographical range of a clade at each nested level. Distances were determined by GeoDis using the longitude and latitude coordinates taken at each sample site. Fetzner & Crandall (2003) suggested, for aquatic species, a 'river' distance (measuring the distance between two points following only linear water bodies) rather than great circle distance (which uses latitude–longitude coordinates). This approach was not taken for this project because aquatic distances are not known for subterranean basins due to unknown and inaccessible connections. Although the approach used in this study could have some effect on the lower (newer) nesting levels, the higher (older) nesting levels would presumably remain unaffected (Fetzner & Crandall, 2003). The second measurement calculated by GeoDis is 'nested clade distance' (D_n), which estimates the evolutionary distance between two haplotypes or clades from the center (oldest) nested clade. The output of GeoDis was used to answer a series of dichotomous questions in the NCA inference key (Templeton, 2004). These inferences help explain what type of event [such as contiguous range expansion (CRE) or restricted gene flow (RGF)] led to the current haplotype diversity of a species. The most recent version of the GeoDis inference key can be found at http://darwin.uvigo.es/software/geodis.html.

Genetic–habitat associations

GeoDis was used to test for significant associations between genetic and habitat (cave or surface) patterns for clades that include both habitat types. This was done by reducing the number of 'locations' in the GeoDis input file to two (cave and surface). These two new 'locations' were assigned different coordinates and the test of habitat association (χ^2)

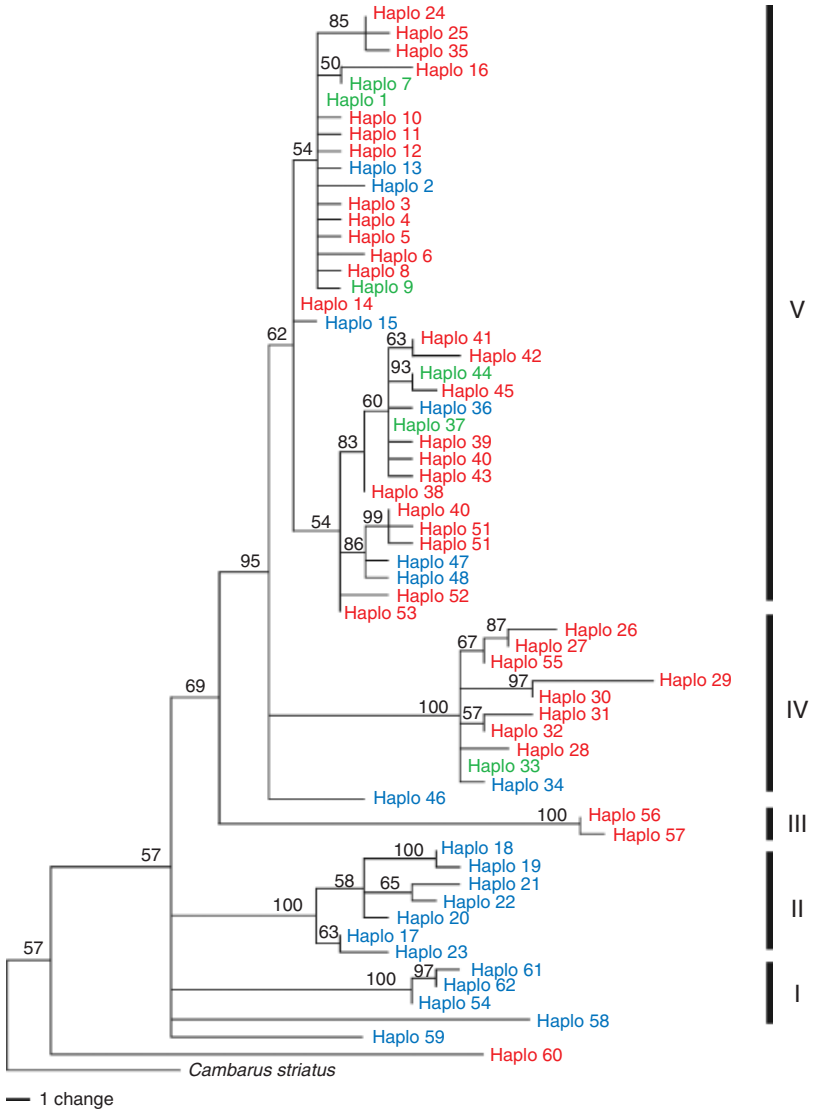


Figure 2 Phylogenetic relationships of 62 *Cambarus tenebrosus* haplotypes of 16S mtDNA sequences. The Bayesian analysis was run using the GTR + I + G (general time reversible plus proportional invariant plus gamma) model of evolution determined by ModelTest. The numbers above the branches indicate posterior probabilities. Haplotypes were colored according to habitat (red = cave, blue = surface and green = both cave and surface). The five main clades are labeled with roman numerals as I, eastern Cumberland Plateau; II, central Tennessee; III, Indiana; IV, western Cumberland Plateau; V, entire sampled range except Indiana. *Cambarus striatus* was used as an outgroup.

was performed over 5000 permutations. This effectively results in a permutation χ^2 test as described by Roff & Bentzen (1989).

To test the hypothesis that *C. tenebrosus* is a recent invader of the cave habitat versus a long-standing resident, we used Fisher's exact test to identify significant associations between tip haplotypes (more recent events) and interior haplotypes (older events) for cave and surface habitats in clades with both habitats represented. If the species was a recent invader into subsurface waters, a significant association would be expected between the cave habitats and the tip locations of the tree. Likewise, if the species was historically located in the cave, but recently invaded surface waters, a significant association would be observed between the cave and interior clades (or surface and tip clades). If no significant association was found, this would provide evidence for long-term residence in both cave and surface waters.

Demographic parameters

Current genetic diversity (θ_π; Tajima, 1983) and historical-based genetic diversity (θ_w; Watterson, 1975) were obtained using the computer program DnaSP 4.0 (Rozas *et al.*, 2003). Current genetic diversity was computed by pairwise differences between sequences whereas historical-based Watterson's θ was determined by the number of segregating sites. These two methods together provide a diversity comparison between current and recent historical diversity of a species for a conservation perspective (Templeton, 1993; Yu *et al.*, 2003; Buhay & Crandall, 2005). Recent losses of diversity (e.g. through selective sweeps or population bottlenecks) would

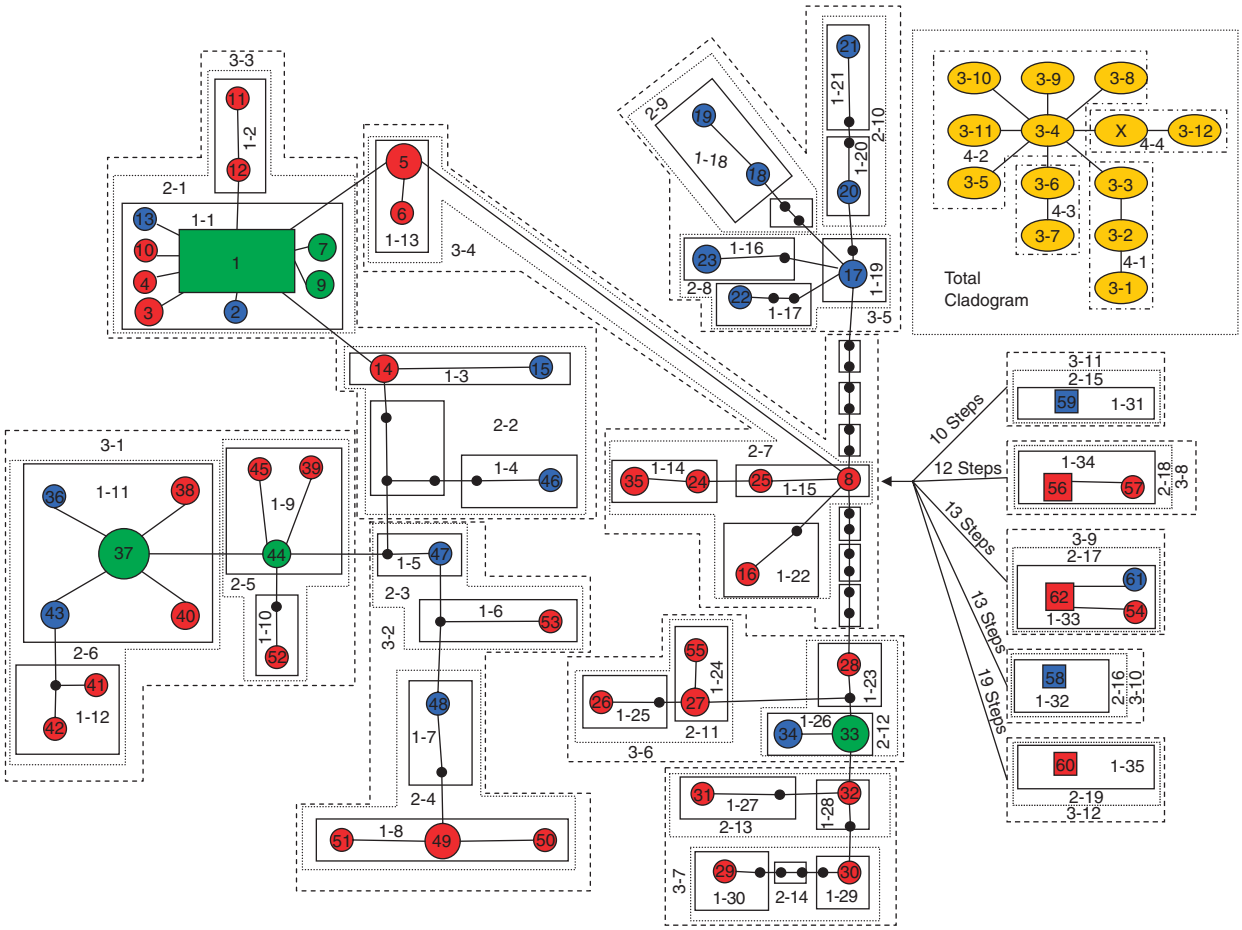


Figure 3 Haplotype network showing the nesting levels used to infer historical processes. Numbers indicate haplotypes (62 total) with black dots representing unsampled or possible extinct haplotypes. The rectangular shape designates the ancestral haplotype for that network. Haplotypes represented by larger numbers of individuals (frequency) are depicted as larger shapes, but the size is not proportional to frequency. Colors correspond to habitat (red = cave, blue = surface and green = both cave and surface). The total cladogram is shown in orange.

typically show $\theta_\pi < \theta_w$, whereas recent increases in genetic diversity (e.g. through population growth) would show $\theta_\pi > \theta_w$.

Results

Phylogenetic analyses

A total of 233 partial 16S (485 base pairs) mitochondrial DNA sequences from 104 collection sites was gathered for *C. tenebrosus*, which included 62 unique haplotypes (Table 1). These haplotypes are accessioned into GenBank as DQ087332–DQ087393. Bayesian analysis (Fig. 2) revealed that *C. tenebrosus* from both cave and surface habitats formed a monophyletic group relative to *C. striatus* (GenBank DQ087394). The cave and surface populations did not form separate monophyletic groups, indicating that there is ongoing gene flow between these two habitats. Additionally, the same haplotype was found in both surface and subsurface habitat types in six instances.

NCA

Haplotype connections $\leq$ nine substitutions for the 485 bp of the 16S mitochondrial gene were determined to be part of the 95% confidence set of network connections. All haplotypes were included in a single network created by TCS with the exception of haplotypes 54 and 56–62 (Fig. 3). Although these haplotypes were determined to be outside the 95% confidence level (by 13 or fewer mutational steps for every haplotype, except 60 which was 19 steps), they were still included in the analysis. Haplotypes 56 and 57 (Indiana cave sites) may have connected to the network had more sampling taken place in north-western Kentucky and southern Indiana (Fig. 1). *Cambarus striatus* was also outside the 95% confidence level, being 10 mutational steps from haplotype 8. The network mostly centered around a single ancestral haplotype (haplotype 1 in Fig. 3) that contained 88 sequences from 42 locations (both cave and surface) found throughout the range of *C. tenebrosus*, excluding Indiana. Nesting of the haplotype network resulted in 35 one-step clades, 19 two-step clades, 12 three-

Table 2 Results of the nested clade analysis of *Cambarus tenebrosus* 16S mtDNA haplotypes based on 5000 permutations

0-step clades			1-step clades			2-step clades			3-step clades			4-step clades		
Haplotype	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n
1	93	93.62	1-1	97.36L	97.24	2-1	95.67	96.23	3-3	98.37	100.14	4-1	95.81S	96.38S
2	0.00	177.66												
3	53.67	151.04L												
4	0.00	105.75												
7	11.84S	135.42												
9	36.95S	67.72												
10	0.00	47.63												
13	0.00	77.22												
I-T	69.18L	−16.79												
11	0.00	29.19S	1-2	43.73	45.88									
12	0.00	87.35L												
I-T	0.00	58.16L	I-T	53.63	51.36									
14	105.03L	105.04	1-3	103.93	105.09S	2-2	105.03	116.12						
15	0.00	90.61												
I-T	105.03L	14.44												
46			1-4	0.00	175.99									
			I-T	103.93	−70.90S									
40	23.27	23.36	1-9	24.23	23.90	2-5	23.56	27.81	3-1	24.62S	52.25S			
45	0.00	16.66												
39	0.00	46.28												
I-T	23.27	−8.11												
52			1-10	0.00	17.99									
			I-T	24.23	5.91									
41			1-12	0.00	23.04	2-6	22.41S	22.85S						
42														
36	0.00	37.36L	1-11	22.37	22.36									
37	21.35	21.40												
38	0.00	28.24												
43	23.51	24.68												
44	8.98	16.16												
I-T	18.22	−3.12	I-T	22.37	−0.68	I-T	1.15	4.96						
47			1-5	0.00	47.76	2-3	63.74	131.49	3-2	105.62	159.68L			
58			1-6	0.00	95.70									
48			1-7	0.00	250.57	2-4	78.58	99.56						
49	12.36	13.02	1-8	12.99S	46.72S									
50	0.00	6.90												
51	0.00	18.84												
I-T	12-36	0.15	I-T	−12.99L	203.84L	I-T	−14.83	31.93	I-T	74.63L	55.11L			
5	103.49	100.27	1-13	79.01	86.06	2-7			3-4	93.77S	102.43S	4-2	128.07L	125.60L
6	0.00	108.48												
I-T	103.49	−8.21												
35	0.64S	1.81	1-14	2.44S	61.05									
24	0.00	3.86L												
I-T	−0.64	2.05												
25	0.00	11.04S	1-15	17.68	74.72									
8	0.00	44.22												
16			1-22	0.00	37.37									
			I-T	64.91	28.66									
23			1-16	0.00	29.83	2-8	46.67	49.35	3-5	45.42S	90.31S			
22			1-17	0.00	48.65									
17			1-19	0.00	86.28									
			I-T	0.00	50.17									
18			1-18			2-9	0.00	35.86						
19														
20			1-20	0.00	85.24L	2-10	48.68	48.42						

Table 2 Continued

0-step clades			1-step clades			2-step clades			3-step clades			4-step clades		
Haplotype	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n
21			1-21	0.00	34.06									
			I-T	0.00	51.18	I-T	22.33	7.21						
56	1.99	2.24L	1-34			2-18			3-8	2.24S	269.13			
57	0.00	2.24												
I-T	1.99	0.00L												
54	0.00	35.10	1-33			2-17			3-9	25.67S	158.72			
61	0.00	25.55												
62	0.00	16.37												
I-T	0.00	−13.96												
58			1-32			2-16			3-10	0.00	78.77			
59			1-31			2-15			3-11	0.00	166.96			
									I-T	64.98	−36.50			
26			1-25	0.00	37.69	2-11	58.30	57.31	3-6	54.68	53.87	4-3	54.40S	131.65L
27	37.65S	52.99S	1-24	60.53	60.59									
55	0.00S	75.61												
I-T	37.65	−22.61S	I-T	60.53	22.90									
28			1-23	0.00	147.89	2-12	50.12	53.47						
33	26.76	26.32	1-26	26.77	34.51									
34	0.00	30.51												
I-T	26.76	−4.19				I-T	−8.18	−3.84						
31			1-27			2-13	0.00	29.02	3-7	34.83	57.79			
32			1-28											
30			1-29			2-14	0.00	43.53						
29			1-30											
						I-T	0.00	−14.51	I-T	19.85	−3.92			
60			1-35			2-19			3-12			4-4	0.00	84.02
												I-T	37.47L	25.25L

Clade (D_c) and nested clade (D_n) distances are given. S indicates that the distance is significantly small at the 5% level and L indicates that the distance is significantly large. In clades with both tip and interior nested clades, the average distance I-T is given. Shaded regions indicate interior grouping.

step clades, four four-step clades and the total cladogram (Fig. 3). The NCA returned 23 significantly large and 27 significantly small associations between genetic variance and geographic location (Table 2).

The NCA revealed significant genetic associations of clades and sampling locations at all clade levels except level two (Table 3). The null hypothesis of no geographic association was rejected at two 1-step clades (1-1 and 1-11), three 3-step clades (3-1, 3-2, and 3-3), all 4-step clades (4-1, 4-2 and 4-3) and the total cladogram. RGF with isolation by distance (IBD) was inferred for four of the nine significant clades (at 3- and 4-step levels; Table 3). An inference of RGF with long-distance dispersal (LDD) was determined for the total cladogram.

Habitat association

The habitat association χ^2 test revealed no significant association between current genetic patterning and habitat type (cave and surface), except for clades 1-26, 3-2 and 4-2 (Table 4). For Fisher's exact test, we counted 23 cave haplotypes occurring on the tips with 12 interior and 13 surface tip haplotypes with five interior. This resulted in no significant association between habitat (cave and surface) and relative age (recent and historical) of the tested haplotypes ($P = 0.76$).

Demographic parameters

Estimates of the current (θ_π) and recent historical (θ_w) genetic diversity for *C. tenebrosus* are extremely high (Nei, 1987) and independent of habitat type (Table 5). These diversity estimates are proportional to the effective population sizes ($\theta = 2N_e\mu$), suggesting that the number of breeding individuals is large in both cave and surface populations. The estimate of effective population size should not be considered a census of the total population of the species, as it only estimates the number of breeding individuals contributing to the gene pool. Interestingly, the recent historical diversity estimates (θ_w) are almost double those of the current diversity estimates (θ_π) (Table 5), showing a sharp decline, nearly 50% loss, in the recent history of the species (Sinclair *et al.*, 2002; Yu *et al.*, 2003; Buhay & Crandall, 2005).

Table 3 Nested contingency results based on 5000 permutations in GeoDis

Clade	χ^2	Probability	Inference chain	Inferred pattern
1-1	464.80	0.01*	1-2-3-5-6-7-8-No	IS
1-2	2.00	1.00	1-19-20-2-11-17-4-No	RGF with some IBD
1-3	8.00	0.37	1-2-3-4-No	RGF with some IBD
1-8	11.82	0.45	Nothing significant	NA
1-9	17.00	0.43	Nothing significant	NA
1-11	86.92	0.007*	1-2-11-17-No	IO
1-13	10.00	0.30	Nothing significant	NA
1-14	3.00	1.00	1-2-3-4-No	RGF with some IBD
1-15	2.00	1.00	1-19-20-2-11-17-4-No	RGF with some IBD
1-24	3.00	1.00	1-19-20-2-11-12-No	CRE
1-26	8.56	0.60	Nothing significant	NA
1-33	8.00	0.16	Nothing significant	NA
1-34	0.75	1.00	1-2-11-17-4-No	RGF with some IBD
2-1	70.65	0.14	1-2-3-4-No	RGF with some IBD
2-2	8.00	0.34	1-19-20-2-3-5-6-7-Yes	RGF/D with LDD
2-3	2.00	1.00	Nothing significant	NA
2-4	13.00	0.07	1-19-20-2-3-5-6-7-8-No	IS
2-5	10.00	0.30	Nothing significant	NA
2-6	22.63	0.10	Nothing significant	NA
2-7	34.13	0.88	1-2-3-4-No	RGF with some IBD
2-8	10.00	0.06	Nothing significant	NA
2-10	2.00	1.00	1-2-11-17-4-No	RGF with some IBD
2-11	4.00	1.00	Nothing significant	NA
2-12	14.00	0.34	Nothing significant	NA
3-1	42.11	0.00*	1-2-3-4-No	RGF with some IBD
3-2	16.00	0.03*	Nothing significant	NA
3-3	114.00	0.0002*	Nothing significant	NA
3-5	10.65	0.17	Nothing significant	NA
3-6	14.99	0.11	Nothing significant	NA
3-7	4.00	0.35	Nothing significant	NA
4-1	346.72	0.00*	1-2-3-4-No	RGF with some IBD
4-2	170.00	0.00*	1-2-11-12-No	CRE
4-3	23.00	0.02*	Nothing significant	NA
Total	612.58	0.00*	1-2-3-5-6-7-Yes	RGF/D with LDD

*indicates significance with a probability of 0.05 or less. Inferences were made using Templeton's (2004) revised key. RGF/D, restricted gene flow/dispersal; IBD, isolation by distance; CRE, contiguous range expansion; IS, inadequate sampling; IO, inconclusive outcome; LDD, long-distance dispersal.

Discussion

The support values for most nodes in Bayesian topology are markedly low. The polytomies in the tree are not a result of low overall genetic diversity, but rather are caused by small mutations in the 16S gene that cannot be resolved at the intraspecific level using a phylogenetic approach. However, some deep structure exists in the tree, showing four well-supported clades, which mainly cluster according to geography (Fig. 2). Clade I is localized near the eastern border between Alabama and Tennessee along the Cumberland Plateau. All the haplotypes in clade II are from the surface sites in central Tennessee. Clade III is localized to south-central Indiana and could represent a distinct evolutionarily significant unit (ESU), but more sampling is required in this area to support this conclusion. The haplotypes in clade IV are concentrated along the border separating Alabama and Tennessee to the west of those haplotypes found in clade I. Clade V is a mixture of both surface and cave populations and spans the entire sampled distribution of *C. tenebrosus*, except for Indiana.

RGF and contiguous range expansion were inferred for most of the significant phylogeographic patterns within *C. tenebrosus*, particularly in clade 4-2 (Fig. 3), which includes six (of 12 total) of the 3-step clades. This may explain why *C. tenebrosus* is found across such a large distribution for a freshwater crayfish species. The network was less informative at some nesting levels because of possible short isolation periods, insufficient geographic sampling or panmixia.

Samples of *C. tenebrosus* from Indiana (haplotypes 56 and 57) were separated by 12 steps and haplotype 60 from central Tennessee was 19 steps from the 95% network. With such extensive geographic overlap of the clades, particularly in Tennessee and Alabama, it becomes difficult to define boundaries for ESU designation within *C. tenebrosus*. Additional sampling in northern Kentucky and

Table 4 χ^2 test of habitat association executed in GeoDis. This test includes only clades with both cave and surface locales

Clade	χ^2	Probability
1-1	10.682	0.162
1-3	8.000	0.126
1-11	8.957	0.086
1-26	8.556	0.036*
1-33	4.000	0.497
2-1	0.498	1.000
2-2	3.938	0.226
2-4	14.000	0.074
2-6	0.354	1.000
2-12	0.268	1.000
3-1	1.607	0.332
3-2	9.905	0.026*
3-3	0.035	1.000
3-6	0.950	0.569
4-1	1.724	0.452
4-2	30.716	0.000*
4-3	0.726	0.617
Total	6.681	0.074

*indicates significance with a probability of 0.05 or less.

Table 5 Current (θ_π) and historical-based (θ_w) estimates of genetic diversity and corresponding effective population size estimates for *Cambarus tenebrosus* (collectively and segregated based on habitat)

	Current		Historical	
Cambarus tenebrosus	θ_π	N_e	θ_w	N_e
All samples (n=233)	0.02359	428 910	0.04394	798 910
Cave (n=187)	0.02142	389 450	0.04007	728 550
Surface (n=46)	0.02677	486 730	0.03501	636 550

Effective population sizes were determined using a substitution rate of 2.2% per million years with a generation time of 5 years (Buhay & Crandall, 2005).

southern Indiana may support the recognition of ESUs or even distinct species which do not overlap geographically with other clades. Our outgroup species, *C. striatus*, fell just outside the 95% confidence limit in the haplotype network at 10 mutational steps. Further sampling of *C. striatus* and other closely related species may provide additional insight into phylogenetic relationships with *C. tenebrosus*.

Cambarus tenebrosus appears to have occupied both cave and surface habitats throughout its evolutionary history. This is supported by the presence of haplotypes from both cave and surface habitats situated in the interior of the network. Therefore, rather than an incipient cave species, it appears that *C. tenebrosus* is a long-term inhabitant of caves and associated streams, despite the morphological changes typically associated with the obligate cave-dwelling species.

Despite having a relatively abrupt decrease in genetic diversity in recent history, *C. tenebrosus* still maintains an extremely high level of diversity. This high level of diversity is not surprising considering its unusually large range, its ability to survive in above-ground and below-ground aquatic habitats, and a certain degree of population subdivision among the major clades. *Cambarus tenebrosus* is an opportunistic crayfish, occupying almost any freshwater karst area, including subterranean areas with and without obligate cave-dwelling crayfish species. For the subterranean populations, open habitat increases as the limestone erodes, which creates new subterranean spaces and corridors (i.e. connections between two previously separated karst areas). These newly formed groundwater connections provide new habitat over time as well as access to other neighboring gene pools.

Cambarus tenebrosus is a robust species of freshwater crayfish in that it has attained an extremely high level of genetic diversity because it can thrive in two very different yet connected habitats. Important factors in shaping the genetic patterns of aquatic species are climatic fluctuations and glacial events (Graham & Grimm, 1990; Vrba, 1992; Roy *et al*., 1996). It appears that the cave populations of *C. tenebrosus* have slightly higher historical genetic diversity than surface populations. A higher genetic diversity in the caves may suggest that the underground environment possibly acted as refugia during glacial/interglacial periods when surface waters were in flux between drought during glacial periods and flooding periods during interglacials of south-flowing meltwaters.

Personal observations regarding the troglomorphisms of *C. tenebrosus* indicate that populations in the northern portion of the species' distribution were notably less sensitive to artificial light (e.g. flashlights) in the caves, whereas *C. tenebrosus* in northern Alabama caves were often startled by light and retreated. Moreover, *C. tenebrosus* in the more northern areas were mostly gray or light brown in body color, whereas the southern populations possessed more coloration such as light orange, green and pink. This might indicate that the crayfish expanded into the southern regions more recently, and have not had time to accumulate fixed troglomorphisms (such as loss of body pigmentation) in the southern populations.

Invasive species are acknowledged as a major economical threat as well as a threat to indigenous species (Vitousek *et al*., 1996; Pimentel *et al*., 2000; Mooney & Cleland, 2001) throughout the world. Invasive species are typically genetically diverse (Lee, 2002), thus providing a rich pool to draw from to adapt to new surroundings and to out-compete species that occupy a similar niche. *Cambarus tenebrosus* would certainly fall into this description of a potential invasive species because of its high levels of genetic variability and its capacity to thrive in cave and surface environments, particularly karst-dominated areas. Identifying possible invasive species is necessary to protect the overall biological diversity of freshwater systems (Lodge *et al*., 1998). By identifying potential invasive species, precautions can be taken to help avoid their introduction into new areas. Crayfishes are particularly troublesome because they are often used as fish bait and, therefore, are easily transferred artificially from one location to another. If this form of unnatural range expansion were to happen with *C. tenebrosus*, it would be especially problematic in both surface and cave environments.

Acknowledgements

We gratefully acknowledge support for this research from the National Science Foundation (DDIG DEB-0508580) (J.E.B., K.A.C.), Brigham Young University Graduate School and Integrative Biology Department (J.E.B.), United States Fish and Wildlife Service (J.E.B.), Brigham Young University Undergraduate Honors Program (J.B.F., K.A.C.) and Brigham Young University Office of Research and Creative Activities (J.B.F.). Comments by Horton Hobbs III, Jack Sites and an anonymous reviewer were very helpful. We also thank the many members of the National Speleological Society who assisted with the fieldwork, especially Gerald Moni, Nathaniel Mann and Kevin Toepke.

References

Agapow, P.M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C. & Purvis, A. (2004). The impact of species concept on biodiversity. *Q. Rev. Biol.* **79**, 161–179.

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

Buhay, J.E. & Crandall, K.A. (2005). Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. *Mol. Ecol.* **14**, 4259–4273.

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

Crandall, K.A., Fetzner, J.W. Jr., Lawler, S.H., Kinnersley, M. & Austin, C.M. (1999). Phylogenetic relationships among the Australian and New Zealand genera of freshwater crayfishes. *Austr. J. Zool.* **47**, 199–214.

Crandall, K.A. & Fitzpatrick, J.F. Jr. (1996). Crayfish molecular systematics: using a combination of procedures to estimate phylogeny. *Syst. Biol.* **45**, 1–26.

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

Culver, D.C., Master, L.L., Christman, M.C. & Hobbs, H.H. III (2000). Obligate cave fauna of the 48 contiguous United States. *Conserv. Biol.* **14**, 386–401.

Danielopol, D.L., Griebler, C., Gunatilaka, A. & Notenboom, J. (2003). Present and future prospects for groundwater ecosystems. *Environ. Conserv.* **30**, 104–130.

Fetzner, J.W. Jr. & Crandall, K.A. (2003). Linear habitats and the nested clade analysis: an empirical evaluation of geographic versus river distances using an Ozark crayfish (Decapoda: Cambaridae). *Evolution* **57**, 2101–2118.

Graham, R.W. & Grimm, E.C. (1990). Effects of global climate change on the patterns of terrestrial biological communities. *Trends Ecol. Evol.* **5**, 289–292.

Hay, W.P. (1902). Observations on the crustacean fauna of the region about Mammoth Cave, Kentucky. *Proc. US Natl. Mus.* **25**, 223–236.

Huelsenbeck, J.P. & Ronquist, F. (2001). MrBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R. & Bollback, J.P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310–2314.

Lee, C.E. (2002). Evolutionary genetics of invasive species. *Trends Ecol. Evol.* **17**, 386–391.

Lodge, D.M., Stein, R.A., Brown, K.M., Covich, A.P., Bronmark, C., Garvey, J.E. & Klosiewski, S.P. (1998). Predicting impact of freshwater exotic species on native biodiversity: challenges in spatial scaling. *Aust. J. Ecol.* **23**, 53–67.

Madison, D.R. & Madison, W.P. (2000). *MacClade 4. Analysis of phylogeny and character evolution*. Sunderland, MA: Sinauer Associates.

Master, L. (1990). The imperiled status of North American aquatic animals. *Biodiver. Network News* **3**, 1–2, 7–8.

Master, L. (1991). Assessing threats and setting priorities for conservation. *Conserv. Biol.* **5**, 559–563.

Mooney, H.A. & Cleland, E.E. (2001). The evolutionary impact of invasive species. *Proc. Natl. Acad. Sci. USA* **98**, 5446–5451.

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

Neves, R.J., Bogan, A.E., Williams, J.D., Ahlstedt, S.A. & Hartfield, P.W. (1997). Status of aquatic mollusks in the southeastern United States: a downward spiral of diversity. In *Aquatic fauna in peril: the southeastern perspective*: 43–86. Benz, G.W. & Collins, D.E. (Eds). Special Publication 1, Southeast Aquatic Research Institute. Decatur, GA: Lenz Design and Communications.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2000). Environmental and economic costs of nonindigenous species in the United States. *Bioscience* **50**, 53–65.

Posada, D. & Crandall, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

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

Rader, R.B., Belk, M.C., Shiozawa, D.K. & Crandall, K.A. (2005). Empirical tests of ecological exchangeability. *Anim. Conserv.* **8**, 239–247.

Ricciardi, A., Neves, R.J. & Rasmussen, J.B. (1998). Impending extinctions of North American freshwater mussels (Unionoida) following the zebra mussel (*Dreissena polymorpha*) invasion. *J. Anim. Ecol.* **67**, 613–619.

Ricciardi, A. & Rasmussen, J.B. (1999). Extinction rates of North American freshwater fauna. *Conserv. Biol.* **13**, 1220–1222.

Roff, D.A. & Bentzen, P. (1989). The statistical analysis of mitochondrial DNA polymorphisms: chi-square and the problem of small samples. *Mol. Biol. Evol.* **6**, 539–545.

Roy, K., Valentine, J.W., Jablonski, D. & Kidwell, S.M. (1996). Scales of climatic variability and time averaging in

Pleistocene biotas: implications for ecology and evolution. *Trends Ecol. Evol.* **11**, 458–563.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Schuster, G.A. (1997). Resource management of freshwater crustaceans in the southeastern United States. In *Aquatic fauna in peril: the southeastern perspective*: 269–282. Benz, G.W. & Collins, D.E. (Eds). Special Publication 1, Southeast Aquatic Research Institute. Decatur, GA: Lenz Design and Communications.

Sinclair, E.A., Costello, B., Courtenay, J.M. & Crandall, K.A. (2002). Detecting a genetic bottleneck in Gilbert's Potoroo (*Potorous gilbertii*) (Marsupialia: Potoroidae), inferred from microsatellite and mitochondrial DNA sequence data. *Conserv. Genet.* **3**, 191–196.

Sites, J.W. Jr. & Crandall, K.A. (1997). Testing species boundaries in biodiversity studies. *Conserv. Biol.* **11**, 1289–1297.

Sites, J.W. Jr. & Marshall, J.C. (2003). Delimiting species: a renaissance issue in systematic biology. *Trends Ecol. Evol.* **18**, 462–470.

Sites, J.W. Jr. & Marshall, J.C. (2004). Operational criteria for delimiting species. *Annu. Rev. Ecol. Evol. Syst.* **35**, 199–227.

Spielman, D., Brook, B.W. & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. USA* **101**, 15261–15264.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.

Taylor, C.A. (1997). Taxonomic status of the subgenus Erebicambarus, genus *Cambarus* (Decapoda: Cambaridae) east of the Mississippi. *J. Crustacean Biol.* **17**, 352–360.

Taylor, C.A., Warren, M.L. Jr., Fitzpatrick, J.F. Jr., Hobbs, H.H. III, Jezerinac, R.F., Pflieger, W.L. & Robison, H.W. (1996). Conservation status of crayfishes of the United States and Canada. *Fisheries* **21**, 25–38.

Templeton, A.R. (1989). The meaning of species and speciation: a genetic perspective. In *Speciation and its consequences*: 3–27. Otte, D. & Endler, J.A. (Eds). Sunderland, MA: Sinauer Associates.

Templeton, A.R. (1993). The 'eve' hypothesis: a genetic critique and reanalysis. *Am. Anthropol.* **95**, 51–72.

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.R. (2001). Using phylogeographic analyses of gene trees to test species status and processes. *Mol. Ecol.* **10**, 779–791.

Templeton, A.R. (2004). Statistical phylogeography: methods of evaluating and minimizing inference errors. *Mol. Ecol.* **13**, 789–809.

Templeton, A.R., Boerwinkle, E. & Sing, C.F. (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., Routman, E.J. & Phillips, C.A. (1995). Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in *Ambystoma tigrinum*. *Genetics* **140**, 767–782.

Templeton, A.R. & Sing, C.F. (1993). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**, 659–669.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L. & Westbrooks, R. (1996). Biological invasions as global environmental change. *Am. Sci.* **84**, 468–478.

Vrba, E.S. (1992). Mammals as key to evolutionary theory. *J. Mammal.* **73**, 1–28.

Watterson, G.A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276.

Wiens, J.J., Chippindale, P.T. & Hillis, D.M. (2003). When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. *Syst. Biol.* **52**, 501–514.

Williams, J.D., Warren, M.L., Cummings, K.S., Harris, J.L. & Neves, R.J. (1993). Conservation status of freshwater mussels of the United States and Canada. *Fisheries* **18**, 6–22.

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Kidd, J.R., Deinard, A.S., Ryder, O., Kidd, K.K. & Li, W.H. (2003). Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**, 1511–1518.