Phylogeography and population genetics of a headwater-stream adapted crayfish, 
*Cambarus pristinus* (Decapoda: Cambaridae), from the Cumberland Plateau in Tennessee

A thesis
Presented to
The College of Graduate Studies
Austin Peay State University
In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Biology

Brooke A. Grubb
December, 2019
Statement of Permission to Use

In presenting this thesis in partial fulfillment of the requirements for the Master of Science in Biology at Austin Peay State University, I agree that the library shall make it available to borrowers under the rules of the library. Brief quotations from this field study are allowable without special permission, provided that accurate acknowledgement of the source is made. Permissions for extensive quotations or reproduction of this field study may be granted by my major professor, or in his/her absence, by the Head of the Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Brooke A. Grubb

______________________________

Date:

______________________________
For Rose Mier
She introduced me to the
wonderful world of crayfish.
ACKNOWLEDGMENTS

First, I want to thank Dr. Rebecca Blanton Johansen for all her help, guidance, and feedback during my master’s work. Without Rebecca focusing my enthusiasm, I likely would have followed every ‘shiny object’ and my master’s work would’ve been a much longer process. I also want to thank her for all the invaluable advice related to pursuing a PhD and being available to answer any and all questions I might have had and thank her for allowing me to pursue my Ph.D. while finishing my master’s work, despite the difficulties of working long distance.

I would also like to thank Dr. John Johansen. His experience with crayfishes in Tennessee has been an invaluable source of information. I also want to thank him for always helping in the field. His field commentary always brought group morale up and his knowledge of the area proved invaluable.

I want to thank Rebecca and John both for letting me stay with them my last few months in Tennessee and for constantly being sounding boards for my thoughts.

I also want to thank Dr. Mollie Cashner who has provided invaluable advice when I was having trouble with lab work and for asking questions which would make me look at my work from a different angle. I would like to thank her and Dr. Steve Hamilton for serving on my committee and providing feedback on my thesis.

My day to day life would have been bland without all the amazing friends I’ve made in the graduate program. You all provided an amazing amount of support over the last couple of years. You pitched in with field work, lab work, watched my pets when needed, listened to my grievances, climbed with me, and ultimately were there whenever I needed you. May the pun board and bread phylogeny live on in infamy! Thank you:
Nastasia Disotell, Jacob Brumley, Ben Wilson, Brooke Washburn, Josh Stonecipher, Allison Kirby, Bailey Harris, John and Rebecca Johansen, Claire Ciafre, Mollie Cashner, Lyranda Thiem, Thomas Murphy, Sandra Bojic, Beau Rapier, Libnis Sanchez, Tammy St. James, Chelsea Blumbergs, Mariah Schlis-Elias, Caroline Wilford.

I would also like to thank Chris Skelton, Carl Williams, and Jeff Simmons for their help in finding and capturing crayfish. Your field experience allowed us to find individuals in an area that we had been unsuccessful in. I would like to thank Martha Lewis for allowing us to collect on her property and Matt Brown for aiding me in getting permits required for collecting at Fall Creek Falls State Park. I would also like to thank the Tennessee Wildlife Resource Agency and Austin Peay’s Center of Excellence for Field Biology for funding which made this project possible.
ABSTRACT

BROOKE A. GRUBB. Phylogeography and population genetics of a headwater-stream adapted crayfish, *Cambarus pristinus* (Decapoda: Cambaridae), from the Cumberland Plateau in Tennessee. (Under the direction of DR. REBECCA BLANTON JOHANSEN)

Habitat loss and fragmentation represent significant threats to North American crayfish diversity. Assessments of genetic diversity within and among populations of imperiled species can provide a baseline for determining the relative impacts of contemporary anthropogenic threats such as habitat fragmentation to population connectivity, as well as aid in identifying historical factors that contribute to population structure. *Cambarus pristinus*, a species of conservation concern and a headwater-stream adapted crayfish endemic to the Cumberland Plateau in Tennessee, exhibits a disjunct distribution within 4th order or lower tributaries of the Caney Fork and Big Brush Creek systems and is comprised of two morphologically distinct forms, the nominal Caney Fork form (Caney Fork system) and the Sequatchie form (Caney Fork and Big Brush Creek systems). Habitat degradation from activities such as silviculture and strip mining have been observed throughout the Cumberland Plateau and *Cambarus pristinus* has experienced recent local extirpations. Our objectives were to examine variation in mitochondrial DNA (COI) and nuclear alleles (microsatellites) to provide estimates of phylogeographic relationships and contemporary levels of genetic structure, respectively, within *C. pristinus*. We predicted that changes in stream hydrology, physiographic regions, and drainage divides would contribute to long-standing isolation among populations separated by these features. We also expected recent anthropogenic disturbances and population loss to have further impacted population connectivity and
lead to reduced genetic diversity and increased population isolation relative to historic levels. Assessment of variation in mitochondrial haplotypes and phylogeographic relationships found low haplotype divergences and broadly shared haplotypes within each morphological form, implying that gene flow was maintained among populations within a form at some level historically, and that geographic features and natural instream barriers did not prevent dispersal. Each form had a unique set of haplotypes and was recovered as a separate, divergent clade indicating the two forms represent distinct genetic lineages and supporting recognition of the Sequatchie form as a distinct taxon. For the Caney Fork form of *C. pristinus* our microsatellite data recovered a high degree of population isolation and support for the occurrence of six isolated populations. We also recovered several low genetic diversity metrics within each cluster and for the Caney Fork form overall. This suggests that the Caney Fork form of *C. pristinus* has a reduced adaptive potential and that historic connectivity has been lost under anthropogenic disturbance. We suggest that *Cambarus pristinus* warrants continued state protection and that future population genetic monitoring be implemented.
# TABLE OF CONTENTS

## CHAPTER

### PAGE

I. **Introduction** .......................................................................................................................... 1

II. **Methods** ............................................................................................................................ 7

   Sample collections .................................................................................................................. 7

   DNA extractions .................................................................................................................... 8

   Mitochondrial DNA data collection and analyses ................................................................. 8

   Microsatellite genotype data collection ............................................................................... 9

   Microsatellite marker validation ......................................................................................... 10

   Spatial genetic structure ...................................................................................................... 11

   Genetic diversity estimation ............................................................................................... 12

III. **Results** .......................................................................................................................... 13

   Sample collections ............................................................................................................... 13

   Mitochondrial DNA data collection and analyses ............................................................... 14

   Microsatellite marker validation ......................................................................................... 15

   Spatial genetic structure .................................................................................................... 15

   Genetic diversity estimates ............................................................................................... 17

IV. **Discussion** ..................................................................................................................... 18

   Phylogeography of *Cambarus pristinus* ........................................................................... 19

   Contemporary genetic diversity ......................................................................................... 23

   Conservation implications for *Cambarus pristinus* ......................................................... 28
TABLES...........................................................................................................................................31
FIGURES...........................................................................................................................................34
LITERATURE CITED...........................................................................................................................42
APPENDIX A......................................................................................................................................58
List of Tables

Table 1. Locality information and number of individuals of *Cambarus pristinus* captured. A “+” denotes new localities identified and sampled for *Cambarus pristinus*. Site ID corresponds to those used in all other tables and figures.

Table 2. Genetic diversity measures based on 19 microsatellite loci for 6 distinct population clusters examined for the Caney Fork form of *Cambarus pristinus*: mean alleles per locus (Na), allelic richness (AR), private allelic richness (PAR), observed (H_o) and expected (H_e) heterozygosity, inbreeding coefficient (F_{IS}), p-values for Bottleneck, p-values for Hardy-Weinberg equilibrium (HWE), and percentage of polymorphic loci. Values in bold were significant (p < 0.05 for F_{IS} and Bottleneck, and p < 0.003 for HWE after Bonferroni correction); n is the number examined per distinct population cluster.

Table 3. Pairwise measures of genetic and geographic distances used for the IBD analysis of the Caney Fork form of *Cambarus pristinus* based on 19 microsatellite loci (Figure 4). Site ID corresponds to those used in Figure 1 and defined in Table 1. a) Pairwise *F*_{ST} values (above the diagonal) and pairwise Jost’s D values (below the diagonal). Values in bold are significant at p < 0.05 and values with an asterisk (*) are significant following Bonferroni correction (p < 0.0018). b) Pairwise river distance (above the diagonal) and pairwise Euclidean distance (below the diagonal) in kilometers.
Table S1. Genetic diversity metrics summarized from the literature for several imperiled and non-imperiled crayfish and compared to those for *Cambarus pristinus* generated herein. Species are grouped by family with continent in parentheses. Metrics include IUCN conservation status (CS) with vulnerable (V), data deficient (DD), least concern (LC), and endangered (EN); mean alleles per locus ($N_a$); allelic richness (AR); observed ($H_o$) and expected ($H_E$) heterozygosity. A “-” indicates that genetic diversity metric was not reported.

Table S2. Primer sequences for 42 microsatellite loci optimized for the Caney Fork form of *Cambarus pristinus* and the Sequatchie form of *C. pristinus* including locus features: repeat motif, the number of alleles observed ($N_a$), allele size range in base pairs (bp), and annealing temperature ($T_a$). Data are summarized from the number of individuals indicated except for loci in bold which are summarized from our full data set of the Caney Fork form of *C. pristinus* with 169 individuals and an exploratory data set of the Sequatchie form with 21 individuals.

Table S3. Effective population size ($N_e$) estimates generated herein and population census size ($N_c$) estimates expressed as number of individuals/100 meters from Johansen et al. (2016). A “-” indicates that the metric was not reported. Site ID corresponds to Table 1.
LIST OF FIGURES

Figure 1. Known historical localities for *Cambarus pristinus*. The Caney Fork form is denoted by circles and the Sequatchie form is denoted by diamonds. A black “X” denotes an uncollected locality. Shapes filled with color denote sites where *C. pristinus* was detected and captured for use in our genetic assessments and unfilled shapes denote sites where *C. pristinus* was not detected. The white star denotes where the Caney Fork river becomes >4th order. The Cumberland Plateau is represented in gray and the surrounding regions are white. The thick black line represents the Western Escarpment of the Cumberland Plateau. Letters by shapes correspond to Site ID defined in Table 1……..34

Figure 2. The 95% statistical parsimony haplotype network of the 51 individuals of *C. pristinus* examined using the mitochondrial COI gene. Two networks were recovered, with each network corresponding to a single form of *C. pristinus*. Haplotypes differed by 1-4 mutations within each form. The number of mutations between haplotypes of each form exceeded the connection limit of 11 mutational steps. A circle represents a single haplotype and the “H” numbers designate different haplotypes. Size of the circles reflect the number of individuals sharing that haplotype. Colors represent individual sites and correspond to Figure 1. Open circles represent unsampled haplotypes and the line between circles denote a single mutation………………………………………………………………36

Figure 3. Maximum likelihood 50% majority rule consensus tree from 1000 bootstraps recovered for *Cambarus pristinus* from the mitochondrial COI gene. Sequence
divergences between species pairs are in parentheses. Colors identify localities and corresponds to those in Figure 1..........................37

**Figure 4.** STRUCTURE analysis based on 19 microsatellite loci from 165 individuals across eight localities representing the Caney Fork form of *Cambarus pristinus*. a) Geographic distribution of the six population clusters inferred from the hierarchical STRUCTURE analyses (by colored polygons). Colors correspond to the inferred ancestral population depicted in the STRUCTURE plots (b-d). Letters by circles represent Site ID defined in Table 1. b) Initial STRUCTURE analysis with and without use of LOCPRIOR. c) Hierarchical STRUCTURE plot using LOCPRIOR of Cluster A recovered in the initial STRUCTURE analysis. d) Hierarchical STRUCTURE plot using the LOCPRIOR for Cluster B recovered in the initial STRUCTURE analysis. b-d) Boxes designate localities and the color of each individual bar within a box denotes population ancestry of each individual. Letters below boxes and by circles correspond to site ID in Table 1. Delta K method is denoted by "ΔK"; the mean log-likelihood method is denoted by "μL(K)"; and K represents the number of clusters inferred. An “*” identifies the 6 distinct clusters recovered from all STRUCTURE runs........................................38

**Figure 5.** Discriminant analysis of principle components (DAPC) based on 19 microsatellite loci and with individuals labeled by locality for the Caney Fork form of *Cambarus pristinus*. Ellipses denote the 6 genetic clusters recovered that correspond to the 6 clusters recovered in the STRUCTURE analyses (Figure 4). The top-right inset shows retained discriminant eigenvalues of the 5 principal components in relative
magnitude. The first two principal axes are shown, with the first on the vertical
axis.................................................................40

**Figure 6.** Isolation by distance analysis for the eight sites of *Cambarus pristinus* showing
the relationship between geographic (log-transformed river and Euclidean kilometers)
and genetic distances ($F_{ST}$ and Jost’s D) based on 19 microsatellite loci. For
corresponding pairwise values refer to Table 3. Comparisons using log-transformed river
distances are in black and those using log-transformed Euclidean distances are in gray
(River- $F_{ST}$: $R^2=0.0776$; $p=0.1479$, Jost’s D: $R^2=0.1266$; $p=0.091$; Euclidean- $F_{ST}$:
$R^2=0.1570$; $p=0.0542$, Jost’s D: $R^2=0.1511$; $p=0.0541$).................................41

**Figure S1.** Criteria for selecting the number of retained principle components (PCs) and
clusters for DAPC analysis. Red circles denote values retained in the DAPC analysis a)
Graphic depicting recommended number of PCs to retain from the optim.a function in
Adegenet. The a-score is the difference between the proportion of successful
reassignment and values obtained using random groups retaining different numbers of
PCs. b) Graph depicting number of retained PCs and explained cumulative variance. c)
Graph depicting number of cluster and corresponding Bayesian information criterion
(BIC) score. d) Proportion of individuals of a sampling site assigned to a genetic cluster
denoted by DAPC (represented as the columns of boxes). The size of the squares
represents the proportion of the individuals with the scale in the bottom left. Cluster
letters correspond to Figure 5. e) Bar plot depicting membership probability of each
individual to a cluster. Each box denotes a preassigned grouping by Site ID and color
corresponds to cluster color used in Figure 4b-d. Each bar represents an individual and membership probability is determined by the 5 retained discriminant eigenvalues expressed as a percent.................................................................66

**Figure S2.** Maximum likelihood tree recovered for *Cambarus pristinus* from the recovered mitochondrial COI gene showing within clade phylogenetic relationships. Support values >50% bootstrap support are shown. Site ID corresponds to Table 1 and numbers in parentheses represent number of individuals from that site recovered at that node. Numbers in brackets show within clade sequence divergence.........................68

**Figure S3.** Geographic distribution of the six distinct populations of the Caney Fork form inferred from the STRUCTURE and DAPC analyses in relation to specific anthropogenic disturbances (strip mining and silviculture) in the region. Colored polygons represent a distinct population cluster and the black circle inside denotes a collected locality. The size of the black circle denotes the allelic richness recovered for each site and corresponds to the scale in the bottom left........................................69
Introduction

Crayfishes are a diverse group of benthic, freshwater crustaceans with over 600 recognized species and new species described annually (Crandall and Buhay 2008). Approximately two-thirds of this diversity is found in the southeastern United States in the family Cambaridae and many species in this family exhibit high degrees of endemism (Lodge et al. 2000; Taylor et al. 2007; Dyer et al. 2013; Crandall and De Grave 2017; Dyer and Brewer 2018; Loughman and Williams 2018). Crayfishes, like many invertebrates, are a vital part of aquatic community structure. They act as ecosystem engineers by burrowing into substrates and creating microhabitats for other invertebrates (Creed and Reed 2004; Pintor and Soluk 2006; Glon and Thoma 2017), alter trophic interactions directly and indirectly via feeding links at multiple trophic levels (Creed 1994; Lodge et al. 1994; Dorn and Wojdak 2004), and are a vital part of many fish diets (Whitledge and Rabeni 1997). Despite their known biological importance and biodiversity there is a paucity of information on the distribution, life history, and genetic diversity of many species (Taylor and Schuster 2004; Moore et al. 2013; Figiel 2016), which impairs development of effective conservation management strategies for the estimated 48% of North American crayfishes considered imperiled (Taylor et al. 2007).

Overexploitation, habitat fragmentation, pollution, and invasive species are the main drivers of crayfish imperilment (Taylor et al. 2007). Several activities that contribute to habitat fragmentation and alteration in North America crayfishes include urban development, dams/water management, and logging/silviculture practices (Richman et al. 2015). Many crayfishes also have small geographic ranges and small
populations, which makes them more vulnerable to such human-mediated threats (Gaston 1994; Angermeier 1995).

In general, population connectivity via gene flow is disrupted by human-mediated fragmentation but is also influenced by several other factors, including dispersal ability of an organism, natural physical barriers in the environment and, geographic distance between populations (Koizumi et al. 2012; Phillipsen et al. 2015; Davis et al. 2018; Serrao et al. 2018; Tonkin et al. 2018). Compared to other aquatic groups such as fishes, few studies have examined population structure or the overall genetic diversity in crayfish species to explicitly test how such factors may impact population connectivity; the majority of those that have were conducted in Europe and Australia/Oceania (Figiel 2016). Studies that have focused on North American crayfishes have been limited to understanding genetic diversity and population connectivity of invasive populations, phylogeography of widely distributed species, or testing phylogenetic relationships within Cambaridae (Barbaresi et al. 2003; Fetzner and Crandall 2003; Finlay et al. 2006; Yue et al. 2010a). Given the high diversity and level of imperilment of North American crayfishes (Taylor et al. 2007), studies that examine how specific factors may influence population connectivity, especially for species that have small native ranges, are needed. Such studies can inform management of imperiled species by providing a basis for estimating the migration or dispersal potential of species, identifying historical processes that have shaped species distributions, predicting how species may respond to disturbance, and testing hypotheses about the relationships between species traits and the degree of population connectivity observed (Lowe and Allendorf 2010; Koizumi et al. 2012; Wang 2014; Wang and Bradburd 2014; Lowe et al. 2017). For example, it is often
hypothesized that species with low dispersal ability would exhibit high levels of population structure or reduced gene flow, as in the Murray Crayfish, _Euastacus armatus_ of Australia (Whiterod et al. 2016). However, Finlay, et al. (2006) found a high degree of connectivity between surface and cave populations of the Cavespring Crayfish, _Cambarus tenebrosus_, despite a presumed limited dispersal ability in the species.

In riverine environments, dispersal is further limited due to the linear or dendritic pattern of these environments that restrict movement opportunities to narrow aquatic habitat corridors (Fagan 2002; Tonkin et al. 2018). Due to this, habitat degradation or artificial fragmentation is more likely to inhibit gene flow, resulting in population isolation, genetic drift, and reduced population persistence (Ward et al. 1994; Alp et al. 2012; Pilger et al. 2015, 2017). Given the largely linear nature of riverine systems, it is also expected that populations of species will typically display a pattern of increasing genetic isolation with increasing geographic distance among populations know as isolation-by-distance (IBD; Fetzner et al. 2003; Sexton et al. 2013).

Crayfishes, however, may be less constrained by their aquatic environments than other taxa and thus, may not show expected patterns of population isolation. For example, some species engage in limited overland dispersal (Furse et al. 2004; Oliveira et al. 2015; Lipták et al. 2016; Herrmann et al. 2018; Thomas et al. 2018) and dispersal ability is highly variable among crayfishes. Some species exhibit high degrees of dispersal moving across several aquatic systems (Barbaresi et al. 2004; Li et al. 2012; Loughman et al. 2013) and others show limited dispersal with individuals moving ≤ 1 km (Robinson et al. 2000; Bubb et al. 2006; Hurry et al. 2015; Whiterod et al. 2016). Dispersal potential may also be tied to fecundity and body size. For instance, in darters, species with higher
Fecundity exhibit a higher dispersal potential (Turner et al. 1996). Fecundity is closely tied to body size in crayfishes as the number of eggs fertilized is determined by the amount of available surface area on the underside of the mother’s abdomen where the eggs are attached (Distefano et al. 2013). Several barriers limiting crayfish dispersal have been identified including; waterfalls, dams, high gradient streams, and rapid changes in stream flow (Light 2003; Kerby et al. 2005; Bubb et al. 2006; Foster and Keller 2011) but the effects of these barriers on population connectivity vary by species.

To further examine factors that impact population connectivity and population persistence in crayfishes, we examined genetic diversity in the imperiled Pristine Crayfish, *Cambarus pristinus*. It is a headwater-stream adapted crayfish, endemic to approximately 374 km² on the Cumberland Plateau of Tennessee (Figure 1; Withers and McCoy 2005; Rohrback and Withers 2006; Johansen 2018) and occurs only in lower order stream reaches (< 4th) and tributaries of the Upper Caney Fork River, Bee Creek, Big Brush Creek, and Cane Creek systems of the Caney Fork and Sequatchie River systems (Rohrback and Withers 2006; Johansen et al. 2016). Within these tributaries, Williams et al. (2004) noted that occurrence was directly related to the presence of large flat rocks in shallow pools with low flow, leading to a patchy distribution within each tributary. It is unclear if pool habitat use is related to habitat specificity, congener competition (*C. parvoculus* and *C. sphenoides*), or some other factor. At several sites, *C. pristinus* occurs in low densities (less than 50 individuals per 100 m²; Johansen et al. 2016).

There are two distinct morphological forms of *C. pristinus*: the nominal form, which occurs only within the Caney Fork River system (referred to herein as the Caney
Fork form) and the Sequatchie form, whose range spans a drainage divide, found in the Sequatchie River system (Tennessee River drainage) and one tributary of the Caney Fork River system (Cumberland River drainage; Figure 1). The forms are allopatric and distinguishable based on differences in gonopod morphology (Rohrback and Withers 2006). Within the Caney Fork form’s range, many tributaries of the Caney Fork River system (excluding those of the Upper Caney Fork and within the Bee Creek system that are on the Cumberland Plateau) join the Caney Fork mainstem after it flows off the West Escarpment of the Cumberland Plateau (Figure 1). The West Escarpment is characterized by stark drops in elevation where many streams have, through erosion, created confined, fast-flowing gulfs or ravines (Bouchard 1975). Such drastic changes in stream gradient may have created long-standing barriers or filters to gene flow among tributary populations of *C. pristinus* from these areas of the Caney Fork. Because *C. pristinus* is a headwater-stream specialist, it is likely that the larger order (>4th order) reaches of the Caney Fork River also may limit dispersal. Headwater-stream specialists are adapted to dispersing within small stream systems and may be unable to disperse across or within streams that have shifted in topography and hydrology as the river becomes larger (Hurry et al. 2015; Paz-Vinas et al. 2015; Schmidt and Schaefer 2018b). For example, two small-stream adapted fishes, including *Etheostoma basilare*, the Corrugated Darter, and *Etheostoma akatulo*, the Bluemask Darter, show evidence of reduced population connectivity or isolation among tributaries of the Caney Fork River, suggesting the mainstem Caney Fork River may serve as a barrier to their dispersal (Hollingsworth Jr. and Near 2009; Robinson et al. 2013). Similarly, the Sequatchie form is distributed across a drainage divide (Caney Fork River flows to the Cumberland River drainage and the
Sequatchie River flows to the Tennessee River drainage. For aquatic organisms in general, but especially those with presumed low dispersal potential, drainage divides often can be barriers to gene flow as seen in the White-clawed Crayfish, *Austropotamobius pallipes*, and the Barrens Topminnows, *Fundulus julisia* (Gouin et al. 2006; Hurt et al. 2017). It is important to evaluate how historical processes such as these have shaped gene flow, population structure, and distribution of a species as these processes can leave genetic signatures that are detected by genetic markers, such as microsatellites, used to assess more contemporary levels of population connectivity or isolation and thus, lead to incorrect conclusions about how contemporary processes have impacted the populations studied (Balkenhol et al. 2009; Zellmer and Knowles 2009; Davis et al. 2014; Epps and Keyghobadi 2015). Therefore, our first objective was to determine if drainage boundaries, physiographic breaks, changes in hydrology, or other geological features have created long-standing isolation among populations of *C. pristinus*. In addition, we assessed whether the two forms distinguished by morphological differences showed genetic divergence, representing distinct genetic lineages.

Recent surveys have reported loss of *C. pristinus* at several historical localities (Williams et al. 2004; Withers and Mccoy 2005; Rohrback and Withers 2006; Johansen et al. 2016) often in conjunction with anthropogenic disturbances such as bridge construction. Population loss can contribute to increased isolation of extant populations, increasing the likelihood of further extirpation due to loss of genetic diversity through genetic drift (Frankham 1995a). Given these observations, our second objective was to identify contemporary patterns of genetic diversity and population structure in *C. pristinus*. Given the expectation of strong IBD patterns in riverine taxa, we expected
contemporary population structure would be influenced by the spatial arrangement of populations, but also by increased habitat alteration and anthropogenic disturbances leading to reduced population connectivity relative to observed historical connectivity. The small geographic range, small population size, and headwater-stream specialization, and potential habitat specificity suggests that *C. pristinus* has an increased susceptibility to isolation by habitat fragmentation. This study will provide a much-needed assessment of how this and other similar species may have responded to both historic geological processes or features of the environment and to contemporary anthropogenic activities that alter population connectivity.

**Methods**

*Sample collections*

*Cambarus pristinus* specimen collection occurred from October, 2017 to October, 2018 using standard seining and dip net methods (Parkyn 2016) at 13 of the 30 known historical localities in the Upper Caney Fork and Sequatchie River drainages (Table 1; Figure 1); other historical localities were not sampled due to lack of landowner permission or proximity to previously collected sites (Figure 1). In addition, four non-historical localities in the Sequatchie River drainage were sampled (Table 1). Genetic samples were collected by taking a specimen’s chelae or the whole specimen in 95% ethanol for DNA preservation. Whole specimens were collected if chelae were reduced/absent or if total carapace length (TCL) was \( \leq 16 \) mm to ensure enough tissue for future genetic analyses.
DNA extraction

DNA was extracted from chela or abdominal tissue using a Qiagen DNEasy Blood and Tissue kit following standard manufacturer protocols. Because exoskeleton debris impaired our extracted DNA and resulted in PCR failure, we removed debris from all tissues under a microscope prior to extraction. Two elutions were used to generate a 180 μL and 80 μL volume of extracted DNA. DNA was quantified on a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific). Samples over 100ng/μL were diluted to a concentration of 30 ng/μL for microsatellite locus amplification. Any concentrations under 100 ng/μL were used undiluted in microsatellite locus amplification. No dilutions were created for mitochondrial gene sequencing.

Mitochondrial DNA data collection and analyses

Five individuals per locality, except for Long Fork (LF) where only a single individual was collected and used, were amplified using Polymerase Chain Reaction (PCR) and sequenced for the mitochondrial DNA (mtDNA) Cytochrome Oxidase Subunit I (COI) gene using previously published primers (Folmer et al. 1994). PCR reactions used a 25.0 μL total reaction volume with 2.00 μL of individual DNA, 0.75 μL of 25mM MgCl₂ (New England Biolabs), 2.50 μL of Standard 10X buffer (New England Biolabs), 0.50 μL of 10mM dNTP (New England Biolabs), 1.00 μL 10pM primer LCO1490, 1.00 μL 10pM primer HCO2198, 0.25 μL 5000 U/mL Taq Polymerase (New England Biolabs), and 17.00 μL of PCR water. Thermocycler conditions included 35 cycles of the following: 30s at 94°C, 30s at 50°C, and 90s at 72°C, which occurred after an initial denaturization step of 3 minutes at 94°C. These steps were followed by a final
extension of 5 minutes at 72°C. Sanger sequencing was performed by Yale’s DNA Analysis Facility and resulting sequences were aligned and edited using CodonCode Aligner v8.0.1 (Soft Genetics). To minimize the potential for pseudogenes, methods recommended by Buhay (2009) and Song et al. (2008) were followed.

To examine phylogeographic relationships among the resulting COI haplotypes, haplotype networks were constructed in TCS v1.2.1 using a statistical parsimony analysis (Templeton et al. 1992) with a 95% connection limit. Final haplotype network graphics were created in TCSBeautifier (Murias dos Santo et al. 2016). Additionally, a phylogenetic tree was constructed in MEGA v 7.0.26 (Kumar et al. 2016) using a maximum likelihood analysis with 1000 bootstraps. The Hasegawa-Kishino-Yano with gamma distribution (HKY+G) substitution model (Hasegawa et al. 1985) was used as the parameter model of sequence evolution. This was determined by the Akaike information criterion in MEGA’s model selection test consisting of 24 possible substitution models (Nei and Kumar 2000). The final graphic was created in FigTree v 1.4.4 (Rambaut 2012).

Microsatellite genotype data collection

Eighty-two species-specific primers were examined for successful amplification and variation in the Caney Fork Form of C. pristinus. Forty-two loci successfully amplified but only 19 were polymorphic and scoreable (Table S2). These 19 loci were placed into five panels, consisting of four loci with each locus assigned a unique M13 dye tag for the entirety of the study. PCR reactions consisted of 10.00 µL total volume including 1.00 µL of 10X concentrate standard Taq reaction buffer - Mg free (New England Biolabs), 1.20 µL 25mM MgCl₂ (New England Biolabs), 0.20 µL 10 mM dNTPs (New England Biolabs), 0.25 µL 10pM forward primer, 0.50 µL 10pM reverse primer, 0.10 µL 5000 U/mL Taq
Polymerase, 0.10 µL M13-labeled dye (Applied Biosystems, Inc.), and 1.00 µL DNA. Thermocycler conditions consisted of 35 cycles of the following: 30s at 94°C, 30s at the primer specific annealing temperature, and 90s at 72°C, which occurred after an initial denaturation step of 1 minute at 94°C. These steps were followed by a final extension of 5 minutes at 72°C and held at 12°C until the PCR product could be retrieved.

The PCR products were multiplexed by panel and genotyped by the University of Florida Interdisciplinary Center for Biotechnology Research Genotyping and Gene Expression Core using the Liz 600 size standard and an AB3730 sequencer. The resulting genotypes were automatically scored in GeneMarker v1.6 (SoftGenetics), followed by manual confirmation and edits as needed. Microsatellite data were not generated for the Sequatchie form due to small sample sizes.

Microsatellite marker validation

Potential scoring errors of microsatellite loci due to null alleles was evaluated using PopGenReport v3.0.4 (Adamack and Gruber 2014) in R; scoring errors due to large allele dropout and stutter were evaluated with MICRO-CHECKER v2.2.3 (Oosterhout et al. 2004). Null allele validation was conducted following methods of Waples and Manangwa et al. (2018, 2019). Linkage disequilibrium for all locus pairs and departures from Hardy-Weinberg equilibrium (HWE) per locus were tested with GENEPOP v1.1.2 (Rousset 2008) in R. Parameters for HWE and linkage disequilibrium were evaluated with the default Markov chain parameter settings of 10,000 dememorization steps, 1,000 batches, 10,000 iterations per batch and p-values were adjusted following Bonferroni corrections to reduce Type I errors (Rice 1989).
Spatial genetic structure

Genetic differentiation was assessed using two pairwise metrics as recommended by Meirmans and Hedrick (2011). Pairwise $F_{ST}$ (Weir and Cockerham 1984) values were generated in GENETIX v4.05 (Belkhir et al.) and tested for significance using 10,000 permutations of the data. Pairwise Jost’s D (Jost 2008; Jost et al. 2018) values were generated in PopGenReport v3.0.4 in R and tested for significance in GenAlEx v6.503 (Peakall and Smouse 2006, 2012).

Pairwise $F_{ST}$ and Jost’s D values were used to test for a signature of isolation by distance (IBD) using two pairwise geographic distance matrixes measured in Google Earth. One matrix consisted of pairwise log-transformed riverine distances (km). The second matrix was based on pairwise log-transformed Euclidean distance (km) to account for potential underground water flow between localities due to the Cumberland Plateau’s karst topography (Buhay and Crandall 2005; Finlay et al. 2006) or direct overland movements (Oliveira et al. 2015; Thomas et al. 2018). Significance was tested between geographic distance and genetic distance using a Mantel test with 10,000 randomizations of the data using IBD v1.52 (Bohonak 2002). Partial Mantel tests were used to assess if the physiographic break of the Cumberland Plateau escarpment or shift to a large mainstem river ($>4^{\text{th}}$ order; Figure 1) had any effect on genetic distance. Partial mantel tests were conducted using only the pairwise river distance measures.

STRUCTURE v2.3.4 (Pritchard et al. 2000) was used to further examine population sub-division. The initial run consisted of the following parameters: no \textit{a priori} population assignment, a generalized admixture model, correlated allele frequencies, 5 iterations for each value of $K (K=1-9)$, and 10,000 burn-in Markov chain Monte Carlo
(MCMC) steps followed by 100,000 MCMC steps. The most likely number of population clusters was determined by using the mean log-likelihood (\( \text{Ln}[\text{Pr}(X|K)] \)) (Pritchard et al. 2000) and \( \Delta K \) (Evanno et al. 2005) methods in the program STRUCTURE HARVESTER web v0.6.94 (Earl and VonHoldt 2012).

Hierarchical analyses (Vähä et al. 2007; Janes et al. 2017) of initial clusters were run in STRUCTURE using the same parameters as the initial run. Subsequent analyses were run in STRUCTURE using sampling localities as priors in the LOCPRIOR model as recommended by Hubisz et al. (2009) to assess potential weak population structure. The program CLUMPAK v1.1 (Kopelman et al. 2015) was used to summarize and present the output of independent runs of each \( K \).

Discriminant analysis of principle components (DAPC) can describe between-subpopulation variation while minimizing within-subpopulation variation noise. DAPC was conducted in Adegenet v2.1.1 in R (Jombart 2008). The optimal number of clusters \( K \) was determined using the k-means procedure of the function \textit{find.cluster} and inferred from the Bayesian Information Criterion (BIC; Figure S1). Parameters were established using 10,000 iterations and 100 randomly chosen starting centroids to allow the algorithm to converge. Five Principal Components (PC) were retained as determined by the function \textit{optim.a.score} with 10 simulations.

\textit{Genetic diversity estimation}

Genetic diversity metrics for each distinct population cluster recovered from our structure analyses of the Caney Fork form were calculated in the R package PopGenReport v3.0.4. Metrics calculated for each cluster included mean number of alleles per locus (\( N_a \)), percent of polymorphic loci, allelic richness (AR), private allelic
richness (PAR), and observed ($H_o$) and expected ($H_e$) heterozygosity. The inbreeding coefficient ($F_{IS}$) was determined using GENETIX v4.05 (Belkhir et al. 2004) with 10,000 permutations to test for significance. All metrics calculated, except for PAR, were also estimated for the Caney Fork form overall.

Effective population size ($N_e$) was estimated using the linkage disequilibrium method in NeESTIMATOR v2.01 (Waples and Do 2010; Do et al. 2014) to account for a single year dataset. Effective population size was estimated per cluster and for the Caney Fork form with the upper and lower bounds determined by 95% CI and alleles with a frequency of $<0.02$ excluded to prevent upward bias of $N_e$ (Waples 2006).

The program BOTTLENECK v1.2.02 (Piry et al. 1999) was used to test for recent population declines in each cluster and for the Caney Fork form. An excess of heterozygosity was tested using a two-phase model (TPM) with 0, 10, and 20% multistep mutation rates (Zachariah Peery et al. 2012) with 36% variation (Di Rienzo et al. 1994) via a Wilcoxon sign-rank test for significance.

**Results**

*Sample collections*

We obtained tissue samples from 10 of the 13 historical localities and one non-historical locality. A total of 226 individuals were collected. At least 20 individuals were collected per site except for Little Laurel Creek (*LLC*), Spring Creek (*SC*), Flatrock Branch (*FB*), and Long Fork (*LF*; Table 1). All localities were included in our mitochondrial analyses to assess phylogeographic relationships. Only localities containing the Caney Fork form were used for our microsatellite analyses as explained in our methods.
**Mitochondrial DNA data collection and analyses**

We recovered 12 unique haplotypes from the 51 individuals of *C. pristinus* sequenced for the mitochondrial COI gene (Figure 2). Eight haplotypes were recovered for the 40 Caney Fork form individuals and four haplotypes were recovered for the 11 Sequatchie form individuals. Two unique haplotype networks were recovered from the statistical parsimony analysis with a maximum connection limit of 11 steps (95%). One network consisted of the Caney Fork form and the second was comprised of the Sequatchie form (Figure 2). One common haplotype was shared among all but one locality, SC, of the Caney Fork form (H1 in Figure 2). A common haplotype was shared among 2 localities of the Sequatchie Form (H11 in Figure 2). Haplotypes within the Caney Fork form network differed by one to four mutations and those within the Sequatchie form by one to five mutations; between networks haplotypes differed by 11 or more mutations (Figure 2).

Similarly, two distinct clades, each representing a form, were recovered with >95% bootstrap support from our maximum likelihood analysis (Figure 3). There was low divergence and no definable geographic structure within forms (Figure S1). Sequence divergence estimates from uncorrected p-distances was greatest between forms at 2.3% (Figure 3). Sequence divergence within the Caney Fork form was lowest at 0.12% and sequence divergence within the Sequatchie form was similarly low at 0.18% (Figure S2).
Microsatellite marker validation

A total of 169 individuals were genotyped for the Caney Fork form for the 19 loci examined. Of these, 165 individuals were successfully genotyped at 16 or more loci and were retained for subsequent analyses (53/3135 missing genotypes; 1.8%). No evidence of scoring error, stutter, or allele dropout was found in any locus. Evidence of null alleles was found in 5 loci but at only 1–2 localities. All 19 loci were retained for subsequent analyses due to a lack of linkage disequilibrium across sites, no departures from HWE among loci, and the inconsistency of null allele among sites.

Spatial genetic structure

Both the initial STRUCTURE analysis without a priori population assignments and the run using our eight Caney Fork form sites as priors in a LOCPRIOR model recovered K=3 as the most likely number of clusters using the mean log-likelihood method (Figure 4b). The ΔK method for the initial STRUCTURE analysis also indicated K=3 but recovered K=2 under our LOCPRIOR model. Due to the congruence of the initial analysis under both methods and the mean log-likelihood method under the LOCPRIOR model, a K=3 including Cluster A, Cluster B, and Cluster SC (Figure 4b) was used for the subsequent hierarchical analyses. Little to no admixture was observed among these three clusters (Figure 4b).

Potential hidden structure was further examined using hierarchical STRUCTURE analyses for clusters recovered from our initial STRUCTURE output that included multiple sites. Cluster A, which included sites CF, LLC, PoC, MC and WF was examined using the LOCPRIOR model in which each site was used as a prior. The ΔK method indicated K=2 and the mean log-likelihood method indicated K=3, but both
recovered the West Fork (WF) as a distinct cluster. An additional STRUCTURE analysis excluding WF from Cluster A was conducted to determine if there was additional hidden structure in Cluster A; both the ΔK and mean log-likelihood methods recovered K=2, corresponding to site Meadow Creek (MC) and the Upper Caney Fork clusters (Figure 4c). Cluster A was ultimately split into three clusters with sites MC and WF as distinct clusters and sites Caney Fork (CF), Pokepatch Creek (PoC), and Little Laurel Creek (LLC) within the third cluster (Upper Caney Fork cluster), with evidence of only low levels of admixture among these three clusters (Figure 4c). The hierarchical STRUCTURE analysis for Cluster B had congruence between the ΔK and mean log-likelihood methods in that both recovered K=2. These two clusters represented sites WFLCC and PuC which had little to no admixture (Figure 4d). Ultimately, we recovered six distinct populations clusters denoted as: Upper Caney Fork Cluster, MC, WF, PuC, WFLCC, and SC clusters (Figure 4); most clusters were comprised of a single locality examined (Figure 4a).

Six distinct clusters, using 5 principle components (PCs), also were inferred from the discriminant analysis of principle components (DAPC; Figure S1a; b; c). These six clusters were similar to the six distinct population clusters recovered from our STRUCTURE analyses except that some individuals from the Upper Caney Cluster were assigned to either the MC or WF cluster in our DAPC analysis; this discrepancy is likely due to the limits of the membership assignment algorithm (Figure S1d; e).

Pairwise $F_{ST}$ (range: 0.002–0.511) and Jost’s D values (range: 0.002–0.442) were similar across several comparisons and the majority were significant after Bonferroni correction. Spring Creek (SC) consistently had the highest values under both metrics.
indicating that SC is the most genetically differentiated site (Table 3). The few sites where pairwise comparisons were not significant were between sites found within the Upper Caney Cluster recovered from our STRUCTURE and DAPC analyses providing support for connectivity among the sites within the Upper Caney Fork Cluster and isolation of all other sites, resulting in the 6 distinct population clusters recovered.

We found no significant pattern of IBD using river distance ($F_{ST}$: $R^2=0.0776$; $p=0.1479$, Jost’s D: $R^2=0.1266$; $p=0.091$) and Euclidean distance ($F_{ST}$: $R^2=0.1570$; $p=0.0542$, Jost’s D: $R^2=0.1511$; $p=0.0541$). In all analyses, <16% of the total genetic variation was explained by geographic distance between sites (Figure 6). Our partial mantel tests showed a significant correlation between the West Escarpment of the Cumberland Plateau and our pairwise genetic metrics (Jost’s D: $R^2=0.7688$, $p=0.0001$; $F_{ST}$: $R^2=0.6281$, $p=0.0001$) and a non-significant correlation between the larger order of the Caney Fork River and our pairwise genetic metrics (Jost’s D: $R^2=0.0568$, $p=0.1462$; $F_{ST}$: $R^2=0.0459$, $p=0.1080$). However, our relationship between geographic and genetic distance was still non-significant after accounting for the effect of the West Escarpment of the Cumberland Plateau ($F_{ST}$: $R^2=0.0776$; $p=0.1582$, Jost’s D: $R^2=0.0874$; $p=0.1494$).

*Genetic diversity estimates*

A total of 115 alleles were amplified across all loci with an average of 6.1 alleles per locus across all clusters (Range: 2–12 alleles). Allelic diversity was low overall, but relatively uniform among clusters, with $N_a$ ranging from 2.16–4.84 alleles and AR ranging from 1.89–4.03 alleles (Table 2). Observed ($H_o$) and expected heterozygosity ($H_e$) were similar across all clusters (range: 0.33–0.51 $H_o$; 0.38–0.54 $H_e$) except for PuC ($H_o=0.16$; $H_e=0.20$; Table 2), which had relatively lower heterozygosity. Private allelic
richness (PAR) was <0.10 except for WFLCC and SC. Half of the clusters did not have $F_{IS}$ values that differed significantly from zero but the other half had significant positive $F_{IS}$ values indicating the presence of inbreeding for those clusters (Table 2). A signature of a bottleneck event was detected for PuC with a 20% two phase model (TPM). There was no evidence for deviations from HWE after Bonferroni correction within any cluster (Table 2). $N_e$ varied among clusters (64.5–1347.5) but was low overall for the Caney Fork form with an estimate of 15.5 individuals (95% CI: 14.2–16.9).

**Discussion**

Benthic, headwater-adapted species often have relatively low dispersal potential and exhibit hierarchical population structure due to natural instream conditions such as shifts in stream topography and hydrology (Hurry et al. 2015; Schmidt and Schaefer 2018b). Additionally, for such species, human-mediated habitat fragmentation often has a greater impact on the persistence of populations, and thus the species overall, due to their already restricted dispersal potential and limited dispersal pathways (Ward et al. 1994; Alp et al. 2012; Paz-Vinas et al. 2015).

For conservation management, it is important to disentangle population structure generated by geological features from structure created by anthropogenic disturbance. This is because structure related to anthropogenic habitat degradation or fragmentation can be confounded by or confused with signals from long-standing isolation from natural instream filters or barriers to gene flow (Zellmer and Knowles 2009; Davis et al. 2014; Epps and Keyghobadi 2015). Thus, we examined the potential for historic long-standing isolation within *C. pristinus* due to natural instream factors such as physiographic breaks and changes in river size, or past vicariance, and the natural degradation of genetic
similarity due to geographic distance known as isolation-by-distance (IBD). Given human-mediated habitat degradation and population extirpation observed in *C. pristinus*, we also evaluated contemporary population connectivity. We found that historical connectivity was maintained within the nominate form of *C. pristinus* (Caney Fork form) despite potential geological and hydrological barriers but found signatures of increased population isolation under contemporary processes. Contemporary signatures of increased isolation were not attributed to IBD. We also found evidence of divergence between the nominal Caney Fork form and the Sequatchie form, supporting the hypothesis that these represent distinct taxa.

*Phylogeography of Cambarus pristinus*

Previous studies have shown that large mainstem portions of rivers, changes in geology through physiographic breaks, and drainage boundaries serve as filters or barriers to dispersal of various benthic, aquatic taxa and that these barriers lead to long-standing isolation (Hughes et al. 1995; Fetzner and Crandall 2003; Nguyen et al. 2004; Hollingsworth Jr. and Near 2009; Kanno et al. 2011; Lamphere and Blum 2011; Hurry et al. 2015). Therefore, we expected to find signatures of long-standing isolation within *C. pristinus* attributed to reduced dispersal across a drainage divide, across the larger river portions of the mainstem Caney Fork River, and changes in stream gradient associated with the physiographic break of the Western Escarpment of the Cumberland Plateau. Although we found divergence between the two forms, suggesting their long-standing isolation, we found no evidence to suggest long-standing isolation within either form due to these presumed barriers. Each form had a common haplotype that was shared among the majority of sampled localities and haplotype divergence within each form was low;
together these data indicate that gene flow was likely maintained among populations at some level historically. This contrasts with the high degree of divergence exhibited by *E. basilare*, a Caney Fork River endemic. Like *C. pristinus*, *E. basilare* are benthic, and small-stream adapted with a presumed limited dispersal potential, particularly across large river habitats. Unlike *C. pristinus*, *E. basilare* displays a high degree of micro-endemism in which populations show evidence of isolation among tributary systems of the Caney Fork River. For example, each of several tributary systems to the Caney Fork (where *E. basilare* occurs) contains a different cryptic species of *E. basilare*; diverging between 8-2 mya. The authors attributed isolation of populations among tributaries to the strict breeding habitat requirements and reduced larvae dispersal potential of these species (Hollingsworth Jr. and Near 2009; Robinson et al. 2013; Fluker et al. 2014).

Crayfishes, however, may have less stringent habitat requirements for successful breeding. Ovigerous females of *Cambarus friaufi*, the Hairy Crayfish, opportunistically find brooding refuge (Black et al. 2015) and *Procambarus clarkii*, the Red Swamp Crayfish, readily disperses when brooding or caring for young (Oliveira et al. 2015). In addition, juvenile crayfish have been observed to disperse after becoming independent from the mother (Miller et al. 2014b; Glon et al. 2019). The relaxed breeding requirements and increased juvenile dispersal may explain why *C. pristinus* does not exhibit the same degree of isolation exhibited in darters of the Caney Fork River system. However, additional work is needed to test this explicitly. Sites SC of the Caney Fork form and LF of the Sequatchie Form each exhibited only a single haplotype unique to that site. For site LF this may suggest that the mainstem Big Brush Creek limits dispersal within the Sequatchie form although it is more likely a result of low sample size and
additional sampling would be needed to provide resolution. Site SC lies at the periphery of the geographic range of the Caney Fork form of *C. pristinus*. Populations found at the edge of a species’ range typically are constrained by less favorable habitat that may reduce migration rates, increase isolation, and reduce genetic diversity (Eckert et al. 2008; Micheletti and Storfer 2017).

Our study is not alone in showing historic connectivity within crayfish on the Cumberland Plateau. The broadly distributed crayfishes, *Cambarus parvoculus* (Mountain Midget Crayfish), *C. jezerinaci* (Spiny Scale Crayfish), and *C. distans* (Boxclaw Crayfish), all exhibit similarly low haplotype diversity and divergence (Thoma and Fetzner, Jr 2008), however these species appear to be more generalist in their habitat requirements. Historic connectivity has been maintained in other headwater stream-adapted crayfish as well. *Euastacus bispinosus*, the Glenelg Spiny Crayfish, was shown to maintain historic connectivity across drainages in Australia (Miller et al. 2014a). Additionally, all headwater-stream specialists species may not exhibit the same dispersal limitations and additional factors such as density, drainage shape, and variation in large river distance separating headwater habitats should be considered (Schmidt and Schaefer 2018a).

The recovery of unique, divergent haplotypes that were sorted into two divergent clades, suggests long-standing isolation of the two morphologically distinct forms of *C. pristinus*. The distribution of these two forms, however, does not correspond to contemporary geographic breaks examined, given that the Sequatchie River form spans the Cumberland-Tennessee River drainage divide. This pattern of discord between divergence and contemporary geographic breaks have been observed in several other
aquatic species (Buhay et al. 2007; Berendzen et al. 2008; Wagner and Blanton 2017). This phylogeographic pattern may reflect one of several alternative scenarios. For example, ancestral populations within the Caney Fork River may have experienced past vicariance resulting in two forms in this system with subsequent invasion of the Sequatchie system through underground, overland, or headwater transfer. Alternatively, it may reflect isolation of a widespread ancestral species between the two drainages with subsequent transfer of the Sequatchie form into the Caney Fork River. Given the distance that separates the confluence of the Tennessee and Cumberland Rivers (in western Kentucky) and the absence of any records of the species in streams > 4th order in size, it seems less plausible that the current distribution of the Sequatchie form is due to long-distance dispersal through these two drainages. Estimation of divergence times from taxa closely related to *C. pristinus* and also found on the Cumberland Plateau would suggest the two forms of *C. pristinus* diverged sometime during or after the Pleistocene (Crandall et al. 2015). It is likely divergence between the two forms is related to climate oscillations or shifts in stream gradient and drainage patterns, which have contributed to speciation of several lineages of freshwater taxa (Thornbury 1965; Near et al. 2001; Berendzen et al. 2003, 2008; Kozak et al. 2006). Estimation of divergence times for the two lineages would help link the observed patterns to specific geological events in the area and to evaluate and test for potential alternative scenarios that may have led to divergence of the two forms.
Contemporary genetic diversity

Although not directly comparable due to variation in alleles examined, heterozygosity and allelic richness of the Caney Fork form of *C. pristinus* were similar to values observed for other imperiled crayfishes (Azuma et al. 2011; Gouin et al. 2011; Li et al. 2012; Gross et al. 2013; Miller et al. 2014a). Heterozygosity and allelic richness of several imperiled crayfishes, including *C. pristinus*, were relatively lower when compared with non-imperiled crayfishes such as *Procambarus clarkii*, the Red Swamp Crayfish (Li et al. 2012; Table S4). This suggests that *C. pristinus* has reduced adaptive potential when compared to non-imperiled crayfishes (Frankham 1995).

The Caney Fork form overall did not exhibit evidence of a bottleneck, however, site PuC (Puncheoncamp Creek), showed signatures of a bottleneck, suggesting a recent drastic decline in population size at this site. The Cumberland Plateau has experienced extensive sandstone and coal mining activity until the Surface Mining Control and Reclamation Act of 1977 (Schorr et al. 2006) and Puncheoncamp Creek historically had mining practices in its headwaters directly upstream of our sampling site (Figure S3; Moore 1985). Additionally, several abandoned mining sites have been repurposed for silvicultural practice. Withers and McCoy (2005) expressed that the majority of sub-watersheds within the range of *C. pristinus* drain large tracts of timber management land and several streams were not protected from the deleterious effects of sediment transport to the stream, which may result in loss of interstitial space. Crayfish density has been reported to decline in the presence of mining and silviculture activities due to the loss of interstitial space, increase of metal pollutants, and reduced canopy cover (Allert et al. 2012, 2013; Welsh and Loughman 2015). Allert (et al. 2012, 2013) found that the
greatest impacts to crayfish density occurred directly downstream of mining sites with reduced effects as distance from mines increased. Surface and deep anthracite coal-mining have been designated as major factors in the decline of the federally listed Guyandotte Crayfish, *Cambarus veteranus* (Loughman et al. 2016) and logging practices have been a concern for several other crayfishes, including the Piedmont Blue Burrowing Crayfish, *Cambarus harti* and the vulnerable Kiamichi Crayfish, *Faxonius saxatilis* (Jones and Bergey 2007; Helms et al. 2013).

Other *C. pristinus* sites may have experienced past declines in population size that were not detected by our bottleneck analysis, which relies on measures of heterozygote excess. This method requires a reduction of 50 – 80% of the effective population to detect bottlenecks two-thirds of the time, even when additional markers are included (Hoban et al. 2013). In addition, other factors such as pre-bottleneck genetic diversity, bottleneck persistence, when the bottleneck occurred, and population growth can influence or obscure genetic signals (Williamson-Natesan 2005; Zachariah Peery et al. 2012).

We found evidence of inbreeding at several sites (*WFLCC, PoC, SC,* and *PuC;* Table 2). This was not surprising for *PuC*, because it showed signatures of a bottleneck and inbreeding often occurs after a bottleneck event (Hedrick and Kalinowski 2000). However, inbreeding can occur in small population without a prior bottleneck event. All clusters except for *SC*, had inbreeding coefficients with 95% confidence intervals that spanned zero. Thus, we could not confidently conclude the occurrence of inbreeding at the population level (Table 2; Colegrave and Ruxton 2003). However, we recovered a significant inbreeding coefficient for the Caney Fork form (*FIS* = 0.22620) with a 95% confidence interval that did not span zero. This result is concerning because, in many
species, inbreeding often leads to reduced fitness and an increased extinction risk (Frankham 2005; Wright et al. 2008). However, crayfishes may have an inherent degree of inbreeding due to poor dispersal ability (Miller et al. 2014a) and dominant hierarchical social structure, where dominant male crayfishes have additional mating opportunities (Villanelli and Cherardi 1998; Tierney et al. 2000; Moore and Bergman 2005). However, several crayfishes of Cambaridae exhibit polyandry with 2-3 males siring a single clutch (Walker et al. 2002; Yue et al. 2010b; Kahrl et al. 2014) and polyandry is thought to offset inbreeding depression by increasing half-sib progeny and decreasing full-sib progeny (Tregenza and Wedell 2000; Cornell and Tregenza 2007). It is unlikely that the significant inbreeding coefficient for the Caney Fork form is only due to mating strategies since we observed low effective population size estimates, low allelic richness, and localized bottleneck events in the species as well.

In wild populations of animals, the breeding population ($N_e$) is on average approximately 10% of the population census size (Frankham 1995b). Johansen et al. (2016) estimated a census size for several populations of the Caney Fork form of *C. pristinus*; estimates ranged from 45–168 individuals (Table S3). Our point estimates of effective population size ranged from 64.5–1347.5 but the upper bounds of our confidence intervals were infinite for many clusters reducing our confidence in the $N_e$ estimates (Table 2). However, the Caney Fork form of *C. pristinus* had low point estimates of $N_e$ for most clusters examined and for the species overall. Individual cluster estimates are likely impacted by low sample sizes, but the estimate for the species overall had narrow confidence intervals, around the point estimate of only 15.5 individuals. This low estimate likely reflects both a low number of individuals and low genetic diversity in
the breeding population. The 100/1000 rule recommends that a $N_e$ of 100 be maintained to prevent inbreeding depression and an $N_e$ of 1000 to maintain evolutionary potential (Frankham et al. 2014). Following this rule, our data suggests that the Caney Fork form of *C. pristinus* lacks a sufficient breeding population size or has insufficient genetic diversity in the breeding population to prevent inbreeding and has reduced evolutionary potential.

A common pattern observed in aquatic systems is an increase in genetic distance as geographic distance increases or IBD (Wright 1943). We did not find support for IBD when using river distance in the Caney Fork form of *C. pristinus* (Figure 6). It is possible that IBD is masked at the small spatial scale that is the extent of our study, especially if low levels of gene flow are maintained as suggested by the few non-significant pairwise $F_{ST}$ and Jost’s D values we observed (Table 3; Phillipsen et al. 2015; Menger et al. 2017). We continued to find a non-significant IBD even after accounting for the significant correlation between genetic distance and the West Escarpment of the Cumberland Plateau suggesting that something other than geographic distance, such as habitat degradation, is generating the observed population structure.

It is unlikely that *C. pristinus* disperses across land. We recovered a non-significant IBD relationship when using Euclidean distance and a lack of population connectivity across headwater systems from our STRUCTURE analysis. If aquatic organisms maintain geneflow via overland dispersal, then populations across headwater systems in close proximity would exhibit higher degrees of genetic similarity than population spaced far apart within a single headwater system. (Finn et al. 2007). We found significant pairwise $F_{ST}$-values among our populations and a high degree of
dissimilar genetic diversity metrics between our two nearest sites (<2km) across land, **PuC** and **WFLCC**. In addition, we found greater similarity across several genetic 
diversity metric within the Upper Caney Fork headwater system than among the Upper 
Caney Fork and Bee Creek headwater systems. Even though overland dispersal can occur 
within crayfishes it likely differs among species. For instance, even in dry periods, 
several species of stream-dwelling crayfish choose to burrow into the hyporheic zone 
instead of dispersing across land to available water sources (Jones and Bergey 2007; 

The congruence between our pairwise $F_{ST}$ values and the STRUCTURE and 
DAPC analyses provide support for six isolated populations of the Caney Fork form of *C. pristinus*: (1) Upper Caney, (2) MC, (3) WF, (4) PuC, (5) WFLCC, and (6) SC. This high degree of contemporary population isolation contrasts with our phylogeographic analysis 
that showed connectivity was maintained throughout the Caney Fork form’s range historically. One explanation is that the retreating edge of the Western Escarpment has 
resulted in a loss of underground stream connectivity (Anthony and Granger 2007a) as 
indicated by cave abandonment along the tributaries of the Caney Fork River (Anthony 
and Granger 2004). Cave abandonment is estimated to have begun ~5Mya with 
subsequent abandonment occurring into the early Holocene (Anthony and Granger 
2007b). Loss of underground stream connectivity has been documented in several cave-
oblige fauna, including crayfishes along the escarpment lines of the Cumberland 
Plateau (Buhay and Crandall 2005; Finlay et al. 2006; Buhay et al. 2007; Niemiller et al. 
2008; Niemiller and Zigler 2013).
However, given that our data suggests populations maintained some level of gene flow historically and cave abandonment occurred over a long geologic time span, it is unlikely loss of underground connectivity has contributed to contemporary population isolation. Instead, the high degree of contemporary population isolation observed most likely reflects human-mediated habitat alterations and subsequent population loss resulting in reduced extant population size and connectivity. Anthropogenic disturbance such as mining and silviculture practice have been reported throughout the range of *C. pristinus* with several anthropogenic sites near the six distinct population clusters we recovered (Figure S3). Johansen (2018) found that the likelihood of *C. pristinus* occupancy would decrease as evergreen forest cover or conductivity increased. Previous studies have shown that acid mine drainage from mining can increase conductivity and the bulk of silviculture practice on the Cumberland Plateau is from loblolly pine, an evergreen, plantations (McGrath et al. 2004; Schorr et al. 2013). Coal mining activity on the Cumberland Plateau has been reported to reduce salamander richness (Schorr et al. 2013; Muncy et al. 2014), alter macroinvertebrate assemblages (Gangloff et al. 2015), create local extirpation and species replacement in native fishes (Schorr et al. 2006), and reduce population density or create local extirpation in crayfishes (Allert et al. 2012, 2013; Welsh and Loughman 2015).

*Conservation implications for Cambarus pristinus*

The observed geographic separation, phylogenetic divergence, and morphological differences between the two forms of *C. pristinus* support recognition of the Sequatchie form as a distinct species and we recommend it be provisionally referred to as *Cambarus aff. pristinus*. Due to this, the geographic range of nominal *C. pristinus* (herein the Caney
Fork form) is restricted to the Upper Caney Fork and Bee Creek system of the Cumberland Plateau.

Although, a population genetics analysis is needed to assess genetic diversity and gene flow within *Cambarus aff. pristinus*, repeated failures to detect individuals at several localities is concerning and may indicate local extirpation. Several individuals we collected had symptoms of the lethal Porcelain Disease, Thelohaniasis. Porcelain Disease occurs in several decapod crustaceans and is caused by the microsporidian parasites of the genus *Thelohanania* (El-Matbouli and Soliman 2006). Disease prevalence is thought to be linked to increased stress and habitat alteration (Imhoff et al. 2009). Loblolly pine plantations are found throughout the range of *Cambarus aff. pristinus*; these plantations can alter water chemistry and stream habitat and thus, the benthic community structure (McGrath et al. 2004). We propose that additional surveys be conducted to determine the extent of the geographic range of *Cambarus aff. pristinus* and that future monitoring to detect Porcelain Disease be incorporated. In addition, a thorough assessment of genetic diversity within *Cambarus aff. pristinus* is warranted to determine population persistence.

Low genetic diversity, $N_e$ estimates, presence of inbreeding, and population structure exhibited within *Cambarus pristinus sensu stricto* (Caney Fork form) suggests that they have a reduced evolutionary potential and ability to weather future stochastic events (Frankham 1995a; Lowe and Allendorf 2010), warranting continued state protection. In addition, we observed a high degree of contemporary population isolation. This is in stark contrast to historic connectivity as discussed and implies human-mediated habitat alternations have likely impacted this species. However, additional work from a landscape genetics approach would provide a more robust test for effects of specific
landscape and environmental factors have contributed to the observed patterns of genetic diversity observed. *Cambarus pristinus sensu stricto* will also benefit from habitat and occupancy modeling, which would help to understand how abiotic factors have influenced population growth and persistence and outline favorable habitat conditions. Continued population genetic monitoring would establish trends in genetic diversity in the species and populations, which can be used to assess the actual impact of conservation actions. Additionally, because our analyses only provide a single snapshot of genetic diversity, adding time-series sampling would lead to more robust estimates of $N_e$ (Gilbert and Whitlock 2015).
Table 1. Locality information and number of individuals of *Cambarus pristinus* captured. A “+” denotes new localities identified and sampled for *Cambarus pristinus*. Site ID corresponds to those used in all other tables and figures.

<table>
<thead>
<tr>
<th>Form</th>
<th>Site ID</th>
<th>Stream</th>
<th>County</th>
<th>Date Collected</th>
<th>Latitude</th>
<th>Longitude</th>
<th>No. Captured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caney Fork</td>
<td><em>CF</em></td>
<td>Caney Fork River</td>
<td>Cumberland</td>
<td>14-Oct-2017; 08-May-2018</td>
<td>35.99108</td>
<td>-85.18052</td>
<td>24</td>
</tr>
<tr>
<td>Caney Fork</td>
<td><em>WF</em></td>
<td>West Fork</td>
<td>Cumberland</td>
<td>14-Oct-2017; 08-May-2018</td>
<td>35.92435</td>
<td>-85.21365</td>
<td>29</td>
</tr>
<tr>
<td>Caney Fork</td>
<td><em>MC</em></td>
<td>Meadow Creek</td>
<td>Cumberland</td>
<td>14-Oct-2017; 08-Apr-2018</td>
<td>35.88828</td>
<td>-85.17790</td>
<td>27</td>
</tr>
<tr>
<td>Caney Fork</td>
<td><em>WFLCC</em></td>
<td>West Fork Little Cane Creek</td>
<td>Cumberland</td>
<td>14-Oct-2017; 08-May-2018</td>
<td>35.79953</td>
<td>-85.20750</td>
<td>23</td>
</tr>
<tr>
<td>Caney Fork</td>
<td><em>PoC</em></td>
<td>Pokepatch Creek</td>
<td>Cumberland</td>
<td>08-Apr-2018; 08-May-2018</td>
<td>35.94527</td>
<td>-85.18546</td>
<td>30</td>
</tr>
<tr>
<td>Caney Fork</td>
<td><em>LLC</em></td>
<td>Little Laurel Creek</td>
<td>Cumberland</td>
<td>23-Jun-2018</td>
<td>35.84049</td>
<td>-85.12305</td>
<td>11</td>
</tr>
<tr>
<td>Caney Fork</td>
<td><em>SC</em></td>
<td>Spring Creek</td>
<td>Van Buren</td>
<td>23-Jun-2018</td>
<td>35.77052</td>
<td>-85.28789</td>
<td>19</td>
</tr>
<tr>
<td>Caney Fork</td>
<td><em>PuC</em></td>
<td>Puncheon Camp Creek</td>
<td>Cumberland</td>
<td>23-Jun-2018</td>
<td>35.83650</td>
<td>-85.22590</td>
<td>31</td>
</tr>
<tr>
<td>Caney Fork</td>
<td></td>
<td>Henderson Branch</td>
<td>Cumberland</td>
<td>23-Jun-2018</td>
<td>35.82225</td>
<td>-85.16087</td>
<td>0</td>
</tr>
<tr>
<td>Sequatchie</td>
<td><em>CC</em></td>
<td>Camp Creek</td>
<td>Van Buren</td>
<td>11-Jan-2018</td>
<td>35.65099</td>
<td>-85.33764</td>
<td>21</td>
</tr>
<tr>
<td>Sequatchie</td>
<td><em>FB</em></td>
<td>Flatrock Branch +</td>
<td>Sequatchie</td>
<td>09-Oct-2018</td>
<td>35.49406</td>
<td>-85.40233</td>
<td>12</td>
</tr>
<tr>
<td>Sequatchie</td>
<td><em>LF</em></td>
<td>Long Fork 2</td>
<td>Sequatchie</td>
<td>09-Oct-2018</td>
<td>35.50113</td>
<td>-85.40877</td>
<td>1</td>
</tr>
<tr>
<td>Sequatchie</td>
<td></td>
<td>Big Brush Creek +</td>
<td>Sequatchie/ Bledsoe</td>
<td>08-Sep-2018; 08-Oct-2018</td>
<td>35.50017</td>
<td>-85.40623</td>
<td>0</td>
</tr>
<tr>
<td>Sequatchie</td>
<td></td>
<td>Laurel Fork +</td>
<td>Sequatchie</td>
<td>24-Jun-2018</td>
<td>35.52947</td>
<td>-85.43130</td>
<td>0</td>
</tr>
<tr>
<td>Sequatchie</td>
<td></td>
<td>Unnamed Tributary to Bird Fork +</td>
<td>Sequatchie</td>
<td>08-Sep-2018; 08-Oct-2018</td>
<td>35.48935</td>
<td>-85.43797</td>
<td>0</td>
</tr>
<tr>
<td>Sequatchie</td>
<td></td>
<td>Glady Fork</td>
<td>Sequatchie</td>
<td>09-Sep-2018; 09-Oct-2018</td>
<td>35.52543</td>
<td>-85.46857</td>
<td>0</td>
</tr>
<tr>
<td>Sequatchie</td>
<td></td>
<td>Long Fork 1</td>
<td>Sequatchie</td>
<td>09-Sep-2018; 09-Oct-2018</td>
<td>35.50195</td>
<td>-85.45370</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2. Genetic diversity measures based on 19 microsatellite loci for 6 distinct population clusters examined for the Caney Fork form of *Cambarus pristinus*: mean alleles per locus ($N_a$), allelic richness (AR), private allelic richness (PAR), observed ($H_o$) and expected ($H_e$) heterozygosity, inbreeding coefficient ($F_{IS}$), p-values for Bottleneck, p-values for Hardy-Weinberg equilibrium (HWE), and percentage of polymorphic loci. Values in bold were significant (p < 0.05 for $F_{IS}$ and Bottleneck, and p < 0.003 for HWE after Bonferroni correction); n is the number examined per distinct population cluster.

<table>
<thead>
<tr>
<th>Cluster ID</th>
<th>n</th>
<th>$N_a$ (95% CI)</th>
<th>$N_e$ (95% CI)</th>
<th>AR</th>
<th>PAR</th>
<th>$H_o$</th>
<th>$H_e$</th>
<th>$F_{IS}$ (95% CI)</th>
<th>Bottleneck</th>
<th>HWE</th>
<th>% Polymorphic loci</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upper Caney</strong></td>
<td>58</td>
<td>3.95 ($\infty$)</td>
<td>3.14 (827–$\infty$)</td>
<td>0.08</td>
<td>0.46</td>
<td>0.48</td>
<td>0.03237 (-0.025–0.074)</td>
<td>0.369</td>
<td>0.011</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td><strong>WF</strong></td>
<td>25</td>
<td>3.21 (58.3–$\infty$)</td>
<td>2.79 (1046)</td>
<td>0.05</td>
<td>0.45</td>
<td>0.44</td>
<td>0.00318 (-0.100–0.059)</td>
<td>0.340</td>
<td>0.671</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td><strong>MC</strong></td>
<td>25</td>
<td>3.05 (53.1–$\infty$)</td>
<td>2.56 (785.3)</td>
<td>0.00</td>
<td>0.44</td>
<td>0.43</td>
<td>-0.01705 (-0.124–0.045)</td>
<td>0.325</td>
<td>0.705</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td><strong>WFLCC</strong></td>
<td>22</td>
<td>4.84 (52.7–$\infty$)</td>
<td>4.03 (136.6)</td>
<td>0.22</td>
<td>0.51</td>
<td>0.54</td>
<td>0.08564 (-0.007–0.128)</td>
<td>0.384</td>
<td>0.007</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td><strong>SC</strong></td>
<td>19</td>
<td>2.16 (19.3–$\infty$)</td>
<td>1.89 (1347.5)</td>
<td>0.12</td>
<td>0.16</td>
<td>0.20</td>
<td>0.22550 (0.069–0.300)</td>
<td>0.892</td>
<td>0.009</td>
<td>68.42%</td>
<td></td>
</tr>
<tr>
<td><strong>PuC</strong></td>
<td>16</td>
<td>2.58 ($\infty$) (26.9–$\infty$)</td>
<td>2.34 ($\infty$)</td>
<td>0.06</td>
<td>0.33</td>
<td>0.38</td>
<td>0.14904 (-0.008–0.223)</td>
<td>0.039</td>
<td>0.155</td>
<td>73.68%</td>
<td></td>
</tr>
<tr>
<td><strong>Species Overall</strong></td>
<td>165</td>
<td>6.10 (14.2–16.9)</td>
<td>2.66 (15.5)</td>
<td>-</td>
<td>0.41</td>
<td>0.42</td>
<td>0.22620 (0.179–0.265)</td>
<td>0.964</td>
<td>0.197</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Pairwise measures of genetic and geographic distances used for the IBD analysis of the Caney Fork form of *Cambarus pristinus* based on 19 microsatellite loci (Figure 4). Site ID corresponds to those used in Figure 1 and defined in Table 1. a) Pairwise $F_{ST}$ values (above the diagonal) and pairwise Jost’s D values (below the diagonal). Values in bold are significant at $p < 0.05$ and values with an asterisk (*) are significant following Bonferroni correction ($p < 0.0018$). b) Pairwise river distance (above the diagonal) and pairwise Euclidean distance (below the diagonal) in kilometers.

<table>
<thead>
<tr>
<th>Site ID</th>
<th>CF</th>
<th>WF</th>
<th>MC</th>
<th>WFLCC</th>
<th>PoC</th>
<th>LLC</th>
<th>SC</th>
<th>PuC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>-</td>
<td>0.042*</td>
<td>0.062*</td>
<td>0.145*</td>
<td>0.010</td>
<td>0.002</td>
<td><strong>0.412</strong></td>
<td>0.257*</td>
</tr>
<tr>
<td>WF</td>
<td>0.038*</td>
<td>-</td>
<td>0.097*</td>
<td>0.171*</td>
<td>0.065*</td>
<td>0.046</td>
<td><strong>0.420</strong></td>
<td>0.313*</td>
</tr>
<tr>
<td>MC</td>
<td>0.054*</td>
<td>0.086*</td>
<td>-</td>
<td>0.181*</td>
<td>0.052*</td>
<td>0.034</td>
<td><strong>0.426</strong></td>
<td>0.330*</td>
</tr>
<tr>
<td>WFLCC</td>
<td>0.180*</td>
<td>0.209*</td>
<td>0.214*</td>
<td>-</td>
<td>0.143*</td>
<td>0.131*</td>
<td><strong>0.333</strong></td>
<td>0.229*</td>
</tr>
<tr>
<td>PoC</td>
<td>0.010</td>
<td>0.061*</td>
<td>0.046*</td>
<td>0.180*</td>
<td>-</td>
<td>0.009</td>
<td><strong>0.441</strong></td>
<td>0.258*</td>
</tr>
<tr>
<td>LLC</td>
<td>0.002</td>
<td>0.042*</td>
<td>0.029</td>
<td>0.170*</td>
<td>0.010</td>
<td>-</td>
<td><strong>0.453</strong></td>
<td>0.283*</td>
</tr>
<tr>
<td>SC</td>
<td>0.375*</td>
<td>0.379*</td>
<td>0.372*</td>
<td>0.322*</td>
<td><strong>0.442</strong></td>
<td>0.394*</td>
<td>-</td>
<td>0.511*</td>
</tr>
<tr>
<td>PuC</td>
<td>0.266*</td>
<td>0.336*</td>
<td>0.347*</td>
<td>0.279*</td>
<td><strong>0.275</strong></td>
<td>0.303*</td>
<td>0.433*</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site ID</th>
<th>CF</th>
<th>WF</th>
<th>MC</th>
<th>WFLCC</th>
<th>PoC</th>
<th>LLC</th>
<th>SC</th>
<th>PuC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>-</td>
<td>23.82</td>
<td>21.28</td>
<td>55.14</td>
<td>10.09</td>
<td>35.01</td>
<td>44.81</td>
<td>35.91</td>
</tr>
<tr>
<td>WF</td>
<td>7.94</td>
<td>-</td>
<td>5.94</td>
<td>40.86</td>
<td>14.85</td>
<td>19.75</td>
<td>30.53</td>
<td>21.63</td>
</tr>
<tr>
<td>MC</td>
<td>11.30</td>
<td>5.16</td>
<td>-</td>
<td>37.26</td>
<td>12.31</td>
<td>13.73</td>
<td>26.93</td>
<td>18.03</td>
</tr>
<tr>
<td>WFLCC</td>
<td>21.30</td>
<td>13.90</td>
<td>10.10</td>
<td>-</td>
<td>46.14</td>
<td>51.07</td>
<td>15.07</td>
<td>26.95</td>
</tr>
<tr>
<td>PoC</td>
<td>5.00</td>
<td>3.38</td>
<td>6.48</td>
<td>16.30</td>
<td>-</td>
<td>24.38</td>
<td>35.81</td>
<td>26.94</td>
</tr>
<tr>
<td>LLC</td>
<td>17.40</td>
<td>12.40</td>
<td>7.16</td>
<td>8.88</td>
<td>12.80</td>
<td>-</td>
<td>40.74</td>
<td>31.84</td>
</tr>
<tr>
<td>PuC</td>
<td>17.50</td>
<td>9.84</td>
<td>7.27</td>
<td>4.36</td>
<td>12.60</td>
<td>9.25</td>
<td>9.30</td>
<td>-</td>
</tr>
</tbody>
</table>
**Figure 1.** Known historical localities for *Cambarus pristinus*. The Caney Fork form is denoted by circles and the Sequatchie form is denoted by diamonds. A black “X” denotes an uncollected locality. Shapes filled with color denote sites where *C. pristinus* was detected and captured for use in our genetic assessments and unfilled shapes denote sites where *C. pristinus* was not detected. The white star denotes where the Caney Fork river becomes $>4^{th}$ order. The Cumberland Plateau is represented in gray and the surrounding regions are white. The thick black line represents the Western Escarpment of the Cumberland Plateau. Letters by shapes correspond to Site ID defined in Table 1.
**Figure 2.** The 95% statistical parsimony haplotype network of the 51 individuals of *C. pristinus* examined using the mitochondrial COI gene. Two networks were recovered, with each network corresponding to a single form of *C. pristinus*. Haplotypes differed by 1-4 mutations within each form. The number of mutations between haplotypes of each form exceeded the connection limit of 11 mutational steps. A circle represents a single haplotype and the “H” numbers designate different haplotypes. Size of the circles reflect the number of individuals sharing that haplotype. Colors represent individual sites and correspond to Figure 1. Open circles represent unsampled haplotypes and the line between circles denote a single mutation.
Figure 3. Maximum likelihood 50% majority rule consensus tree from 1000 bootstraps recovered for *Cambarus pristinus* from the mitochondrial COI gene. Sequence divergences between species pairs are in parentheses. Colors identify localities and corresponds to those in Figure 1.
Figure 4. STRUCTURE analysis based on 19 microsatellite loci from 165 individuals across eight localities representing the Caney Fork form of *Cambarus pristinus*. a) Geographic distribution of the six population clusters inferred from the hierarchical STRUCTURE analyses (by colored polygons). Colors correspond to the inferred ancestral population depicted in the STRUCTURE plots (b-d). Letters by circles represent Site ID defined in Table 1. b) Initial STRUCTURE analysis with and without use of LOCPRIOR. c) Hierarchical STRUCTURE plot using LOCPRIOR of Cluster A recovered in the initial STRUCTURE analysis. d) Hierarchical STRUCTURE plot using the LOCPRIOR for Cluster B recovered in the initial STRUCTURE analysis. b-d) Boxes designate localities and the color of each individual bar within a box denotes population ancestry of each individual. Letters below boxes and by circles correspond to site ID in Table 1. Delta K method is denoted by "ΔK"; the mean log-likelihood method is denoted by "μL(K)"; and K represents the number of clusters inferred. An "*" identifies the 6 distinct clusters recovered from all STRUCTURE runs.
**Figure 5.** Discriminant analysis of principle components (DAPC) based on 19 microsatellite loci and with individuals labeled by locality for the Caney Fork form of *Cambarus pristinus*. Ellipses denote the 6 genetic clusters recovered that correspond to the 6 clusters recovered in the STRUCTURE analyses (Figure 4). The top-right inset shows retained discriminant eigenvalues of the 5 principal components in relative magnitude. The first two principal axes are shown, with the first on the vertical axis.
Figure 6. Isolation by distance analysis for the eight sites of *Cambarus pristinus* showing the relationship between geographic (log-transformed river and Euclidean kilometers) and genetic distances ($F_{ST}$ and Jost’s D) based on 19 microsatellite loci. For corresponding pairwise values refer to Table 3. Comparisons using log-transformed river distances are in black and those using log-transformed Euclidean distances are in gray (River- $F_{ST}$: $R^2$=0.0776; $p$=0.1479, Jost’s D: $R^2$=0.1266; $p$=0.091; Euclidean- $F_{ST}$: $R^2$=0.1570; $p$=0.0542, Jost’s D: $R^2$=0.1511; $p$=0.0541).
Literature Cited


Belkhir K, Borsa P, Chikhi L, et al GENETIX 4.05, logiciel sous Windows TM pour la genetique des populations


Colegrave N, Ruxton GD (2003) Confidence intervals are a more useful complement to nonsignificant tests than are power calculations. Behav Ecol 14:446–447. doi: 10.1093/beheco/14.3.446


Miller JM, Niraula BB, Reategui-Zirena EG, Stewart PM (2014b) Life history and physical observations of primary burrowing crayfish (Decapoda: Cambaridae) *Cambarus (Lacunicambarus) diogenes* and *Cambarus (Tubericambarus) polychromatus*. J Crustac Biol 34:15–24. doi: 10.1163/1937240X-00002199


Rambaut A (2012) FigTree v1.4


Rohrback G, Withers D (2006a) A Status Survey of the Caney Fork Crayfish (Cambarus pristinus) and Hardin County Crayfish (Orconectes wrighti). 1–81


Withers DI, Mccoy RA (2005a) Distributional surveys for *Cambarus pristinus* and *Cambarus williami*, two endangered crayfish in Tennessee Final Report Contract # ID-05-0828-00 Tennessee Division of Natural Heritage Nashville, Tennessee. 1–57


Appendix A

Supplementary Materials
Table S1. Genetic diversity metrics summarized from the literature for several imperiled and non-imperiled crayfish and compared to those for *Cambarus pristinus* generated herein. Species are grouped by family with continent in parentheses. Metrics include IUCN conservation status (CS) with vulnerable (V), data deficient (DD), least concern (LC), and endangered (EN); mean alleles per locus ($N_a$); allelic richness (AR); observed ($H_o$) and expected ($H_E$) heterozygosity. A “-” indicates that genetic diversity metric was not reported.

<table>
<thead>
<tr>
<th>Species</th>
<th>CS</th>
<th>$N_a$</th>
<th>AR</th>
<th>$H_o$</th>
<th>$H_E$</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cambaridae (North America)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pristine Crayfish (<em>Cambarus pristinus</em>)</td>
<td>DD</td>
<td>6.10</td>
<td>2.66</td>
<td>0.41</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Red Swamp Crayfish (<em>Procambarus clarkii</em>)</td>
<td>LC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invasive populations</td>
<td></td>
<td>6.42–14.33</td>
<td>-</td>
<td>0.56–0.81</td>
<td>0.70–0.89</td>
<td>(Li et al. 2012)</td>
</tr>
<tr>
<td>Native populations</td>
<td></td>
<td>14.58</td>
<td>-</td>
<td>0.65</td>
<td>0.88</td>
<td>(Li et al. 2012)</td>
</tr>
<tr>
<td><strong>Parastacidae (South America / Australia)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glenelg Spiny Freshwater Crayfish (<em>Euastacus bispinosus</em>)</td>
<td>V</td>
<td>1.0–2.3</td>
<td>1.07–1.75</td>
<td>0.004–0.42</td>
<td>0.03–0.36</td>
<td>(Miller et al. 2014a)</td>
</tr>
<tr>
<td><strong>Astacidae (Europe / North America)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noble Crayfish (<em>Astacus astacus</em>)</td>
<td>V</td>
<td>2.2–4.8</td>
<td>2.0–4.2</td>
<td>0.21–0.53</td>
<td>0.23–0.60</td>
<td>(Gross et al. 2013)</td>
</tr>
<tr>
<td>White-clawed Crayfish (<em>Austropotamobius pallipes</em>)</td>
<td>EN</td>
<td>2.02</td>
<td>2.02</td>
<td>-</td>
<td>0.295</td>
<td>(Gouin et al. 2011)</td>
</tr>
<tr>
<td>Signal Crayfish (<em>Pacifastacus leniusculus</em>)</td>
<td>LC</td>
<td>4 - 33</td>
<td>-</td>
<td>0.25–0.84</td>
<td>0.30–0.94</td>
<td>(Azuma et al. 2011)</td>
</tr>
<tr>
<td><strong>Cambaroideidae (Asia)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korean Crayfish (<em>Cambaroides similis</em>)</td>
<td>DD</td>
<td>4.3–8.1</td>
<td>-</td>
<td>0.33–0.47</td>
<td>0.63–0.73</td>
<td>(Ahn et al. 2011)</td>
</tr>
</tbody>
</table>
**Table S2.** Primer sequences for 42 microsatellite loci optimized for the Caney Fork form of *Cambarus pristinus* and the Sequatchie form of *C. pristinus* including locus features: repeat motif, the number of alleles observed (N_a), allele size range in base pairs (bp), and annealing temperature (T_a). Data are summarized from the number of individuals indicated except for loci in bold which are summarized from our full data set of the Caney Fork form of *C. pristinus* with 169 individuals and an exploratory data set of the Sequatchie form with 21 individuals.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer sequence (5’–3’)</th>
<th>Repeat Motif</th>
<th>N_a</th>
<th>Allelic Range (bp)</th>
<th>T_a (°C)</th>
<th>N_a</th>
<th>Allelic Range (bp)</th>
<th>T_a (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cpris001</td>
<td>F: CACGACGTTGTTAAAAACGACGGGCTCCAACCTTTTAAGAGAC&lt;br&gt;R: TCAACTCCCGTCCTCCTCTCAC</td>
<td>AAAC</td>
<td>1</td>
<td>176</td>
<td>61</td>
<td>1</td>
<td>165</td>
<td>61</td>
</tr>
<tr>
<td>Cpris002</td>
<td>F: CACGACGTTGTTAAAAACGACGGGCTCCAACCTTTTAAGAGAC&lt;br&gt;R: TCTCCACCTCTCACTGTAACC</td>
<td>AAAC</td>
<td>1</td>
<td>226</td>
<td>61</td>
<td>1</td>
<td>217</td>
<td>61</td>
</tr>
<tr>
<td>Cpris004</td>
<td>F: CACGACGTTGTTAAAAACGACGGGCTCCAACCTTTTAAGAGAC&lt;br&gt;R: AAAGCAATGACTCTCCTGCTG</td>
<td>AAAC</td>
<td>1</td>
<td>285</td>
<td>63</td>
<td>1</td>
<td>285</td>
<td>63</td>
</tr>
<tr>
<td>Cpris007</td>
<td>F: CACGACGTTGTTAAAAACGACGAGGCTGACAAC&lt;br&gt;R: GGCTTTCTCTGCTCCCTATC</td>
<td>AAAG</td>
<td>5</td>
<td>328–344</td>
<td>61</td>
<td>1</td>
<td>332</td>
<td>61</td>
</tr>
<tr>
<td>Cpris008</td>
<td>F: CACGACGTTGTTAAAAACGACGAGGCTGACAAC&lt;br&gt;R: AAGTCAGTGTGCACCTATG</td>
<td>AAAG</td>
<td>1</td>
<td>374</td>
<td>62</td>
<td>1</td>
<td>378</td>
<td>62</td>
</tr>
<tr>
<td>Cpris013</td>
<td>F: CACGACGTTGTTAAAAACGACGAGGCTGACAAC&lt;br&gt;R: GACACGTTGTCACACATC</td>
<td>AAGT</td>
<td>1</td>
<td>309</td>
<td>62</td>
<td>1</td>
<td>309</td>
<td>62</td>
</tr>
<tr>
<td>Cpris016</td>
<td>F: CACGACGTTGTTAAAAACGACGAGGCTGACAAC&lt;br&gt;R: TGCAGGGTTGACACGAGTAAGG</td>
<td>ACGA</td>
<td>5</td>
<td>206–220</td>
<td>63</td>
<td>3</td>
<td>210–220</td>
<td>63</td>
</tr>
<tr>
<td>Primer</td>
<td>Primer sequence (5’ – 3’)</td>
<td>Repeat Motif</td>
<td>Primer sequence (5’ – 3’)</td>
<td>Repeat Motif</td>
<td>Allelic Range (bp)</td>
<td>T₂ (°C)</td>
<td>Allelic Range (bp)</td>
<td>T₂ (°C)</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------</td>
<td>-------------</td>
<td>----------------------------------------------------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>---------</td>
<td>--------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Cpris017</td>
<td>F: CACGACGTTTGAACGACAGGATGCTGAGAGTCTCTCGTGG</td>
<td>ACAG 1</td>
<td>244</td>
<td>62</td>
<td>1</td>
<td>240</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: GCAGCTCAAACCTTTGAGATAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris018</td>
<td>F: CACGACGTTTGAACGACTACGTCGGCTCCACTACCC</td>
<td>ACAG 1</td>
<td>263</td>
<td>61</td>
<td>2</td>
<td>263–267</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: CGTTCCCGTGAGTGCTACCTGAAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris019</td>
<td>F: CACGACGTTTGAACGACATGTTAGAGTAGGAGGCTTC</td>
<td>ACAG 2</td>
<td>273–277</td>
<td>62</td>
<td>1</td>
<td>271</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: TCCTGTGACTGCTCTCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris021</td>
<td>F: CACGACGTTTGTTGAACGACTACTAGCTGGAGAGGAGAAGG</td>
<td>ACAG 3</td>
<td>317–329</td>
<td>60</td>
<td>4</td>
<td>317–325</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: TGCTCTAAATACCTGCAAGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris022</td>
<td>F: CACGACGTTTGAACGACTACTCATTAGACTAGCTCGCG</td>
<td>ACAG 1</td>
<td>361</td>
<td>61</td>
<td>1</td>
<td>361</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: ACCTGTGTAGCTCCATTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris026</td>
<td>F: CACGACGTTTGTAAAGCACTCTCAGCCATTCTCCGTCGTCGTCG</td>
<td>ACAT 7</td>
<td>232–256</td>
<td>62</td>
<td>6</td>
<td>236–264</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: GTGTGAGCTTAAAGAGGAGCCTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: GACAGACTTCCCATGCAAGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris034</td>
<td>F: CACGACGTTTGTAAACGACGCCAAACAGTCAGTCCAAAC</td>
<td>ACAT 7</td>
<td>383–397</td>
<td>62</td>
<td>3</td>
<td>377–381</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: ACTGGCAAGCAACATAGTCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris037</td>
<td>F: CACGACGTTTGTAAACGACTCATACCCATCTGTGAAGGTG</td>
<td>ACAT 7</td>
<td>422–454</td>
<td>63</td>
<td>3</td>
<td>426–438</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: TGCTCTGTCACCTGACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer</td>
<td>Primer sequence (5’ → 3’)</td>
<td>Repeat Motif</td>
<td>Allelic Range (bp)</td>
<td>T&lt;sub&gt;a&lt;/sub&gt; (°C)</td>
<td>Allelic Range (bp)</td>
<td>T&lt;sub&gt;a&lt;/sub&gt; (°C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris039</td>
<td>F: CACGACGGTTGTAACACGACACGACGCTATGTCATCTCG</td>
<td>ACGG</td>
<td>1</td>
<td>333</td>
<td>1</td>
<td>333</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: CGCCTGATCCGATAACTTTG</td>
<td></td>
<td></td>
<td>62</td>
<td></td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris040</td>
<td>F: CACGACGTGTAACACGACACACGACTCGCTCG</td>
<td>AGAT</td>
<td>2</td>
<td>141–149</td>
<td>2</td>
<td>141–149</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: GTACACTGCTAGGACAAAC</td>
<td></td>
<td></td>
<td>63</td>
<td></td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris043</td>
<td>F: CACGACGTGTAACACGACACCCAGGTTGCTTGCAG</td>
<td>AGAT</td>
<td>1</td>
<td>170</td>
<td>1</td>
<td>170</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: ACAGTCCAGATTCATCATC</td>
<td></td>
<td></td>
<td>61</td>
<td></td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris044</td>
<td>F: CACGACGTGTAACACGACGACTCGCAGGG</td>
<td>AGAT</td>
<td>1</td>
<td>255</td>
<td>2</td>
<td>267–271</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: TAGGACATGATGTGGTGGAG</td>
<td></td>
<td></td>
<td>61</td>
<td></td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris045</td>
<td>F: CACGACGTGTAACACGACGAGTTGAGGGATGGAGGG</td>
<td>ATAGGAGTAA</td>
<td>1</td>
<td>297</td>
<td>1</td>
<td>304</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: TCGTAGAATTGTAATTACGGC</td>
<td></td>
<td></td>
<td>62</td>
<td></td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris046</td>
<td>F: CACGACGTGTAACACGACGATGGAGACCAGCTGGCTTAC</td>
<td>AGAT</td>
<td>12</td>
<td>303–347</td>
<td>9</td>
<td>295–339</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: CGCCTACATGTTGCTGGG</td>
<td></td>
<td></td>
<td>61</td>
<td></td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris047</td>
<td>F: CACGACGTGTAACACGACGGCATCCCTCACTCATTAC</td>
<td>AGAT</td>
<td>2</td>
<td>331–335</td>
<td>1</td>
<td>335</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: ACGGAAAGCCGACCTCTTACG</td>
<td></td>
<td></td>
<td>62</td>
<td></td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris048</td>
<td>F: CACGACGTGTAACACGACGACTCGGAGCCACTCGACTTACG</td>
<td>AGAT</td>
<td>4</td>
<td>374–382</td>
<td>2</td>
<td>374–378</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: GGGTGGATGGATAGATGGGAGGAC</td>
<td></td>
<td></td>
<td>61</td>
<td></td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris049</td>
<td>F: CACGACGTGTAACACGACGACTCGGACTGCTGGAACATTTTC</td>
<td>AGAT</td>
<td>7</td>
<td>440–468</td>
<td>4</td>
<td>448–464</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: TTGACACGATGGGAGGGGAGG</td>
<td></td>
<td></td>
<td>62</td>
<td></td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris055</td>
<td>F: CACGACGTGTAACACGACGACTCGGACTCAGTACG</td>
<td>ATCC</td>
<td>1</td>
<td>188</td>
<td>1</td>
<td>188</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R: TCTAAACAGGAGGATTGGG</td>
<td></td>
<td></td>
<td>60</td>
<td></td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer</td>
<td>Primer sequence (5’ – 3’)</td>
<td>Repeat Motif</td>
<td>Allelic Range (bp)</td>
<td>Tₐ (°C)</td>
<td>Allelic Range (bp)</td>
<td>Tₐ (°C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>--------------------</td>
<td>---------</td>
<td>--------------------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris057</td>
<td>F: CACGACGTTGTAACGACACTCTCTACCATCCATCTGCG R: ACGCTAAAGCTCTCTGACTG</td>
<td>ATCC</td>
<td>1</td>
<td>348</td>
<td>62</td>
<td>1</td>
<td>348</td>
<td>62</td>
</tr>
<tr>
<td>Cpris061</td>
<td>F: CACGACGTTGTAACGACACACCTCCATACAACCAAC R: TCACCACCTCTCTCATCCC</td>
<td>AAAG</td>
<td>3</td>
<td>203–229</td>
<td>63</td>
<td>2</td>
<td>203–205</td>
<td>63</td>
</tr>
<tr>
<td>Cpris062</td>
<td>F: CACGACGTTGTAACGACAGCCCTGTAACGCTTCTCCAC R: CCATGGCTGTGGTTGTAAGAC</td>
<td>AAAG</td>
<td>1</td>
<td>312</td>
<td>64</td>
<td>1</td>
<td>312</td>
<td>64</td>
</tr>
<tr>
<td>Cpris063</td>
<td>F: CACGACGTTGTAACGACACTCTCTCATTACCC R: TCAGGTGGTGAGGAAATAGG</td>
<td>ACAG</td>
<td>10</td>
<td>213–245</td>
<td>64</td>
<td>6</td>
<td>208–232</td>
<td>64</td>
</tr>
<tr>
<td>Cpris064</td>
<td>CACGACGTTGTAACGACACCCTGGCAAGCACAACCTAGG R: ACTACCAAGACCTGAGATAG</td>
<td>ACAG</td>
<td>8</td>
<td>243–291</td>
<td>64</td>
<td>4</td>
<td>243–277</td>
<td>64</td>
</tr>
<tr>
<td>Cpris066</td>
<td>F: CACGACGTTGTAACGACATGAGGACACCACCTGAGCTG R: GATTAACGACACCAACCAACGG</td>
<td>ACAG</td>
<td>6</td>
<td>350–374</td>
<td>63</td>
<td>2</td>
<td>358–366</td>
<td>63</td>
</tr>
<tr>
<td>Cpris068</td>
<td>F: CACGACGTTGTAACGACAGGGAGCTCGATGTCGTG R: GGATTCACCTCCAGATGG</td>
<td>ACAT</td>
<td>8</td>
<td>183–211</td>
<td>62</td>
<td>6</td>
<td>164–195</td>
<td>62</td>
</tr>
<tr>
<td>Cpris071</td>
<td>F: CACGACGTTGTAACGACAACTCAAGGATGGGCTGG R: GCCAGTTGATGGTCTACCTTCC</td>
<td>ACAT</td>
<td>1</td>
<td>333</td>
<td>62</td>
<td>1</td>
<td>317</td>
<td>62</td>
</tr>
<tr>
<td>Primer</td>
<td>Primer sequence (5’ – 3’)</td>
<td>Repeat Motif</td>
<td>C. pristinus Caney Fork form (n=6)</td>
<td>C. pristinus Sequatchie form (n=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris074</td>
<td>F: CACGACGTGGTAAAAACGACGACCAGCTCCACGTATCC R: CCCATATTGACTGGCTTTCGC</td>
<td>ACAT</td>
<td>2 362–368</td>
<td>64 362–368</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris078</td>
<td>F: CACGACGTGGTAAAAACGACACTGACAAACACCATTCTC R: GACCTTCTTCCACCCACCC</td>
<td>AGAT</td>
<td>4 322–338</td>
<td>61 318–338</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris080</td>
<td>F: CACGACGTGGTAAAAACGACGCCCATCAGTTGAACCATCG R: TGTAACGTGGTGAGGTGACCTCC</td>
<td>ATCC</td>
<td>1 183</td>
<td>60 183</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpris082</td>
<td>F: CACGACGTGGTAAAAACGACACCAGCAGCTATGTCTC R: AGACAGATGGGTGGATGAGAC</td>
<td>ATCC</td>
<td>1 380</td>
<td>62 380</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table S3.** Effective population size ($N_e$) estimates generated herein and population census size ($N_c$) estimates expressed as number of individuals/100 meters from Johansen et al. (2016). A “-” indicates that the metric was not reported. Site ID corresponds to Table 1.

<table>
<thead>
<tr>
<th>Site ID</th>
<th>$N_e$ (CI)</th>
<th>$N_c$ (Regression method)</th>
<th>$N_c$ (mark-recapture method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>64.5 (27.7–∞)</td>
<td>36</td>
<td>45</td>
</tr>
<tr>
<td>LLC</td>
<td>∞ (22.6–∞)</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>MC</td>
<td>785.3 (53.1–∞)</td>
<td>36</td>
<td>-</td>
</tr>
<tr>
<td>PoC</td>
<td>913 (56.0–∞)</td>
<td>205</td>
<td>168</td>
</tr>
<tr>
<td>WF</td>
<td>1046 (58.3–∞)</td>
<td>50</td>
<td>62</td>
</tr>
<tr>
<td>WFLCC</td>
<td>136.6 (52.7–∞)</td>
<td>132</td>
<td>163</td>
</tr>
<tr>
<td>Species overall</td>
<td>15.5 (14.2–16.9)</td>
<td>49 ± 59</td>
<td>-</td>
</tr>
</tbody>
</table>
**Figure S1.** Criteria for selecting the number of retained principle components (PCs) and clusters for DAPC analysis. Red circles denote values retained in the DAPC analysis a) Graphic depicting recommended number of PCs to retain from the `optim.a` function in Adegenet. The a-score is the difference between the proportion of successful reassignment and values obtained using random groups retaining different numbers of PCs. b) Graph depicting number of retained PCs and explained cumulative variance. c) Graph depicting number of cluster and corresponding Bayesian information criterion (BIC) score. d) Proportion of individuals of a sampling site assigned to a genetic cluster denoted by DAPC (represented as the columns of boxes). The size of the squares represents the proportion of the individuals with the scale in the bottom left. Cluster letters correspond to Figure 5. e) Bar plot depicting membership probability of each individual to a cluster. Each box denotes a preassigned grouping by Site ID and color corresponds to cluster color used in Figure 4b-d. Each bar represents an individual and membership probability is determined by the 5 retained discriminant eigenvalues expressed as a percent.
Figure S2. Maximum likelihood tree recovered for *Cambarus pristinus* from the recovered mitochondrial COI gene showing within clade phylogenetic relationships. Support values >50% bootstrap support are shown. Site ID corresponds to Table 1 and numbers in parentheses represent number of individuals from that site recovered at that node. Numbers in brackets show within clade sequence divergence.
Figure S3. Geographic distribution of the six distinct populations of the Caney Fork form inferred from the STRUCTURE and DAPC analyses in relation to specific anthropogenic disturbances (strip mining and silviculture) in the region. Colored polygons represent a distinct population cluster and the black circle inside denotes a collected locality. The size of the black circle denotes the allelic richness recovered for each site and corresponds to the scale in the bottom left.