INTRODUCTION

Freshwater mussels in the family Unionidae, are considered one of the most imperiled taxa in the world and a substantial proportion of freshwater mussel diversity can be found in North America (Bogan, 2008; Graf & Cummings, 2007; Haag & Williams, 2014). North America is home to about 297 native freshwater mussel species and ~70% of these species are considered endangered, threatened, or of special concern with only 24% of the species considered to be stable (Williams et al., 2011). Humans have indirectly impacted critical habitat through climate change effects (Dudgeon et al., 2005; Haag & Williams, 2014; Strayer & Dudgeon, 2010; Williams et al., 2011) and directly impacted populations by constructing dams, channelizing, and polluting rivers. (Williams et al., 1989). Mussel species declines...
are an increasing concern because of their crucial role providing ecosystem services such as environmental nutrient recycling, structural habitat for other species, food resource, and biofiltration (USFWS, 2012a; Vaughn, 2018; Vaughn et al., 2008). Their ecosystem services and intrinsic value warrant the development of comprehensive conservation strategies to preserve them. To achieve this, researchers have started to examine the ecological and genetic conditions of imperiled freshwater mussel species at the population level. Successful conservation of imperiled species, such as freshwater mussels, must include a measurement of available genetic diversity as it represents the raw material for adaptation to environmental changes. Lack of genetic diversity can cause populations to become genetically fixed and intolerant to a constantly changing environment (Frankel & Soulé, 1981). Population connectivity is important for maintaining genetic diversity and can be heavily impacted by anthropogenic activity. In this manuscript, we estimate the distribution of genetic diversity and population connectivity to inform conservation decisions of an imperiled freshwater mussel species, the Sheepnose mussel (Plethobasus cyphyus).

There are three species within the genus Plethobasus, all of which are listed as endangered. Of the three species, the Sheepnose currently occupies the broadest distribution (Hove et al., 2015; Turgeon et al., 1998). All three species of Plethobasus have exhibited range-wide population declines presumably as a result of anthropogenic changes to their habitat and it is thought that these changes are the most prominent factor in the decline of Plethobasus (Stein & Flack, 1997; USFWS Service, 1984; US ACOE, 2011). Unlike its much rarer congeners, the relatively high abundance and widespread nature of the Sheepnose allows for an opportunity to conserve this species.

Sheepnose usually occur in shallow shoals with moderate-to-swift currents over coarse gravel and sand (Oesch, 1984). However, other habitat features may include mud, cobble, and boulders in deeper large river runs (Parmalee & Bogan, 1998). In general, freshwater mussels are long-lived species with life spans ranging from two years to decades (Mutvei et al., 1994). Sheepnose are estimated to live ~20–30 years and become reproductively mature around 5 years of age (Hove et al., 2015). The Sheepnose, as well as all unionid species, utilize a mating strategy in which males expel sperm into the water column, which are then taken in by females for fertilization of their eggs which are held in modified portions of their gills. After fertilization, the mature larvae, glochidia, are released into the water and attach to the gills or fins of fish where they complete their development. After development is complete, the juvenile mussels then drop off the gills and establish themselves in the substrate (Parmalee & Bogan, 1998). Originally, the only known natural host fish for the Sheepnose was the Sauger (Sanders canadensis) (Surber, 1913), however, a more recent study has shown that the Sheepnose appears to be a cyprinid host specialist (Hove et al., 2015). Gene flow is thought to be achieved through the dispersal of sperm and the glochidia larval stage (Ferguson et al., 2013; Hove et al., 2015) so conservation efforts must include consideration of host species availability so the life cycle can remain complete.

Historically, the Sheepnose mussel occurred throughout much of the Mississippi River system (Figure 1), including The Ohio River Basin (Allegheny, Kanawah, Ohio, Tennessee, Tippecanoe rivers), Mississippi River Basin (Big Sunflower, Chippewa, Gasconade, Meramec, St. Croix, Wisconsin rivers) (USFWS, 2002). However, as of April 2012, the Sheepnose was listed as endangered under the Endangered Species Act (USFWS, 2012b), primarily because it had been extirpated from two-thirds of its former range and remaining locations appear to be spatially and genetically isolated from each other (USFWS, 2002). According to a status report conducted in 2002 by the USFWS, of the 77 streams that were historically occupied by Sheepnose populations, only 26 streams are thought to still be occupied. This decline in range and abundance has been attributed to human impacts such as land development, dams, and
pollution (Haag & Williams, 2014). However, published and unpublished records since the 1800s, indicate that although Sheepnose was historically widespread, the species was often described as uncommon (USFWS, 2002). Archaeological evidence of discovered shell fragment locations indicates that this species may have been uncommon or rare for centuries (Parmalee & Bogan, 1998). The goals of this study are to describe the genetic diversity and population structure across the majority of the extant range of the Sheepnose. The loss of Sheepnose populations across its range indicates the possibility that a loss of genetic diversity and population connectivity has also occurred. Under this scenario, populations would possibly show the signature of genetic bottlenecks. If historical evidence is true and the Sheepnose has always been rare, analysis of population genetic data would not show signs of a loss of diversity or evidence of a bottleneck, and extant populations would be more resilient to the effects of isolation. Conservation implications will differ depending on the level of current connectivity and genetic diversity available within populations. If the isolated nature of the Sheepnose populations has resulted in the erosion of genetic diversity, then reestablishing habitat connectivity and implementing translocations could be prudent. If, however, the Sheepnose appears to have maintained genetic diversity despite population loss and isolation, then maintaining available habitat and improving connectivity would likely be a higher priority management objective. This study is intended to provide knowledge about the extent of isolation and genetic diversity of the Sheepnose by estimating contemporary population connectivity and structure to inform conservation decisions.

2 MATERIALS AND METHODS

2.1 Data collection

2.1.1 Study area and sampling

A combination of microsatellite markers and mitochondrial DNA sequences was used to estimate genetic diversity, population structure, contemporary migration, and population size changes in the Sheepnose. Samples (N = 164) for DNA extraction were collected from seven different localities (Table 1). Collection efforts were focused on the Mississippi and Ohio river basins (Figure 1). Mussels were collected by snorkeling or SCUBA at various locations. Samples for DNA extraction were collected by taking a small (~1 mm) biopsy of mantle tissue (Berg et al., 1995) or by using cytology brushes that were swabbed over the mantle tissue of mussels to accumulate mucous and sloughed cells (Henley et al., 2006). Biopsy samples were stored in 95% ethanol and DNA was extracted from mantle tissue samples using the QiaGen DNeasy® Blood and Tissue Kit (Qiagen #69506) according to the kit instructions. Cytology brush samples were stored in the lysis buffer provided with the Puregene Buccal Cell Core Kit B (Qiagen) and DNA was extracted following the kit instructions. Extracted DNA was quantified using a Nanodrop ND1000 spectrophotometer and stored at 4°C.

2.1.2 Microsatellites

Sixteen species-specific polymorphic microsatellite loci were used to genotype samples. These markers were developed by Genetic Identification Services, Chatsworth, CA (Appendix S1). Polymerase chain reactions (PCR) were performed using 10 µL reactions (~2 ng of genomic DNA was used in each reaction). The standard M13 protocol (Schuelke, 2000) was used with the fluorescent dye labeled with HEX (Applied Biosystems). Reagents for a 10 µl reaction included: 6.6 µL H2O, 1 µL Biolase Buffer (10×), 0.3 µL MgCl2, 0.8 µL dNTP, 0.1 µL forward primer, 0.1 µL reverse primer, 0.05 µL M13 labeled oligo, 0.05 µL Biolase Taq polymerase, and 1 µL template DNA. Reactions were run in Eppendorf Master Cycler thermal cyclers under consistent conditions (95°C/5 min; [94°C/30 s; 62°C/60 s; 72°C/30 s] × 10 cycles; [94°C/30 s; 55–59.4°C/60 s; 72°C/30 s; 72°C/20 min] × 25 cycles; 72°C/4 min). Negative controls were performed with every reaction to detect potential contamination. PCR products were visualized on 1% agarose gels to confirm that the reaction was successful and that the negative control showed no contamination. PCR products were then sent to the Iowa State DNA Facility where they underwent capillary electrophoresis to determine allele sizes. The

| Site     | Site ID | Sample Size | H_o | H_e | Allelic Richness | Private Alleles | F   |
|----------|---------|-------------|-----|-----|------------------|-----------------|-----|
| Allegheny| ALL     | 22          | 0.772 | 0.768 | 7.031            | 6               | 0.072 |
| Chippewa | CHIP    | 21          | 0.818 | 0.813 | 7.359            | 3               | -0.012 |
| Meramec  | MER     | 26          | 0.834 | 0.844 | 8.440            | 11              | 0.010 |
| Mississippi| MISS   | 51          | 0.825 | 0.859 | 8.180            | 18              | 0.040 |
| Tippecanoe| TIPP    | 19          | 0.860 | 0.834 | 8.768            | 18              | -0.017 |
| Tennessee| TN      | 8           | 0.837 | 0.796 | 8.615            | 4               | -0.052 |
| Wisconsin| WIS     | 17          | 0.831 | 0.836 | 8.106            | 17              | 0.008 |
| Total/Avg|         | 7           | 0.825 | 0.821 | 8.071            | 11              | 0.007 |

Note: Results from analysis of microsatellite data. Observed (H_o) and expected (H_e) heterozygosity, allelic richness, private alleles, and fixation index (inbreeding) of each site across all loci.
raw data were then scored using GeneMarker® v1.85 (Hulce et al., 2011) and converted into the desired software input format using base RStudio (RStudio Team, 2016).

2.1.3 Mitochondrial sequences

Mitochondrial DNA sequences of an 883 base pair fragment of the first subunit of the NADH dehydrogenase gene (ND1) were also generated for samples. DNA sequence data for the ND1 gene were generated through PCR with Leu-uuFR (5′-TGCCAGAAAAGTGCATCAGATTAAAGC-3′) and LoGlyR (5′-CTCTGCTTTGGAAAGGCAAAGGTACT-3′) primers (Serb et al., 2003). For a 25 μl reaction: 1 μl LEU UURF primer, 1 μl LoGlyR primer, 9.5 μl H2O, 12.5 μl MyTaq polymerase (Bioline), and 1 μl template DNA. Approximately ~1 ng of genomic DNA was used for the template. Reactions were run in Eppendorf Master Cycler thermal cyclers at the following conditions: (95°C/60 s; 95°C/30 s; 50°C/60 s; 72°C/30 s) x 37 cycles; 72°C/60 s; 72°C/60 s). Negative controls were performed with every reaction to detect potential contamination. Products were then run on 1% agarose gel to assess that the reaction was successful and to verify that the negative control showed no contamination. Successful PCR products were identified for sequencing using the ExoSAP-IT (US Biochemicals). Template consisted of ~1 ng of genomic DNA. All sequencing reactions were run on an ABI 3730xl machine (Applied Bio systems #4337454) and LoGlyR and Leu-uuFR primers (Serb et al., 2003). The cycle sequencing reactions were run using Big Dye® ver. 3.1 cycle sequencing kit (Applied Biosystems #4337454) and sent to the Iowa State DNA Facility for sequencing. The raw results were then analyzed and edited using GENEIOUS v8.1.6 (Kearse et al., 2012). Results were also converted into amino acids to confirm that sequences were aligned properly before exporting the matrix for further analyses. DNA sequence data have been submitted to GenBank (Accession Number: MH853483).

2.2 Data analysis

2.2.1 Microsatellites

Genetic diversity

Each locus was tested for deviations from Hardy–Weinberg expectations, potential genotyping errors, and linkage disequilibrium at each collection site using MICROCHECKER (Van Oosterhout et al., 2004). The average number of alleles (A), expected heterozygosity ($H_e$), observed heterozygosity ($H_o$), and fixation index ($F_{ST}$) were calculated using GenAIEx v6.5 (Peakall & Smouse, 2012). Rarefied allelic richness was calculated using the package hierfstat (Goudet & Thibaut, 2015) in RStudio. Private alleles were also determined by site using the package poppr (Kamvar et al., 2014) in RStudio. Pairwise $F_{ST}$ (Wright, 1969) values were calculated using GenAIEx. Adjusted $F_{ST}$ ($G_{ST}$) (Hedrick, 2005) and Jost’s D (Jost, 2008) values were calculated using the package DEMEtics (Gerlach et al., 2010) in RStudio in order to account for a potential depression of the standard $F_{ST}$ measure due to high allelic diversity (many loci had between 15 and 45 alleles). A permutation test was performed on the $F_{ST}$ values using base RStudio to determine degree of genetic differentiation found among sampling locations that were located within versus between drainage basins. This test was permuted 100,000 times and $p$-values were assessed at a 0.05 significance level. Microsatellite genotypes have been submitted to DRYAD (https://doi.org/10.5061/dryad.gxd2547mm).

Population structure

Clustering of Sheepnose collection sites into distinct genetic groups was conducted using the program STRUCTURE v2.3.4 (Pritchard et al., 2000). STRUCTURE analyses consisted of a burn-in of 100,000 Markov chain Monte Carlo (MCMC) iterations followed by 1,000,000 iterations using the admixture model and correlated allele frequencies. Each run had 1–8 possible K values (n collection sites –1) and 10 replicates of each run. After an initial analysis detected two populations (K = 2), subsequent analyses to detect substructure were also conducted. STRUCTURE runs were conducted on the two populations identified in the initial run using the same procedures as previously, but with possible K values of 1–4 and 1–5. The web application POPHELPER (Francis, 2016) was used to determine the most probable value of K utilizing the Evanno method (Evanno et al., 2005) to determine the second-order rate of change in the distribution of L(K). POPHELPER was also used to merge the 10 replicates together and to graphically display results. An analysis of molecular variance (AMOVA) with 999 permutations was conducted using GenAIEx to further examine Sheepnose population structure.

Estimation of migration

The program BAYESASS (Wilson & Rannala, 2003) was used to estimate asymmetrical migration rates. Rates were estimated among collection sites because of the high degree of genetic differentiation observed between sites based on the $F_{ST}$ values. BAYESASS estimates genetic flow among sites as a migration rate ($m$) which can be interpreted as the fraction of migrants per generation in one population that is derived from a source population. These estimations are calculated using a Bayesian approach and MCMC sampling to generate values for $m$ over the last few (<5) generations (Wilson & Rannala, 2003). Given a generation time of approximately 5 years for the Sheepnose (Hove et al., 2015), BAYESASS estimated $m$ values over the past ~25 years. Run lengths and parameters were optimized to ensure convergence and delta parameters were adjusted to accommodate 40–60% acceptance. BAYESASS appeared to reach convergence after five runs with a different initial seed and a Bayesian deviance metric (Spiegelhalter, 2002) was used to select the run that best fit the dataset. TRACER v1.7 (Ramault et al., 2018) was also used to visualize mixing, suitable burn-in values, and convergence problems. The final run consisted of the parameters from the best selected run with a run length of 5x10⁷ iterations and sampling every 100 iterations. The burn-in period consisted of 2x10⁷ iterations.
TABLE 2  Pairwise adjusted $G_{ST}$ (below diagonal) and Jost’s $D$ (above diagonal) values for microsatellite data among all sites

|         | ALL   | CHIP  | MER   | MISS  | TIPP  | TN    | WIS   |
|---------|-------|-------|-------|-------|-------|-------|-------|
| ALL     | 0.3680| 0.5726| 0.3688| 0.1047| 0.5703| 0.4150|
| CHIP    | 0.4168|       | 0.2062| 0.1011| 0.2539| 0.4535| 0.0533|
| MER     | 0.6181| 0.2300|       | 0.1868| 0.4336| 0.4674| 0.1610|
| MISS    | 0.4122| 0.1114| 0.2045|       | 0.2820| 0.3827| 0.0580|
| TIPP    | 0.1223| 0.2831| 0.4679| 0.3086|       | 0.4129| 0.3183|
| TN      | 0.6190| 0.4953| 0.5048| 0.4174| 0.4504|       | 0.4199|
| WIS     | 0.4621| 0.0587| 0.1784| 0.0623| 0.3495| 0.4578|       |

Changes in population size

Considering apparent declines in the number of Sheepnose populations and the potential genetic bottlenecks associated with habitat fragmentation and isolation (Andersen et al., 2004), a test for genetic bottlenecks at each site was conducted using BOTTLENECK (Piry et al., 1999). Both tests implemented by BOTTLENECK were used for the analysis. A Wilcoxon’s sign rank test was conducted to determine whether the heterozygosity of a site was less than predicted under mutation-drift equilibrium. This test can detect bottlenecks over the last 2–4Ne generations. The second test uses a mode-shift test of allele proportions over the last few dozen generations (Cornuet & Luikart, 1996; Luikart et al., 1998). This analysis was conducted with 10,000 replications under the stepwise mutation (SMM) and two-phase model (TPM) that included 95% single-step mutations and 5% multi-step mutations and a variance of 12 as recommended by Piry et al. (1999). All collection sites were tested separately for a bottleneck and the $p$-values estimated by the Wilcoxon’s sign rank test were assessed at a 0.05 significance level.

3  | RESULTS
3.1  | Microsatellites
3.1.1  | Genetic diversity

Two of the microsatellite loci (C109A and C115) amplified poorly and were subsequently dropped from the analysis. Only one locus (D113) exhibited an excess of homozygosity and potential null alleles at multiple sampling locations, and it was also dropped from analysis. Microsatellite analysis results described below are from the remaining 13 microsatellite loci (Appendix S1). MICROCHECKER analysis revealed that 5 loci were out of Hardy–Weinberg proportions at 1–2 sampling sites. There were 295 alleles across the 13 loci with 11 to 47 alleles per locus (Appendix S1). Observed heterozygosity was high (>0.70) at all sites (Table 1). Allelic richness averaged 8.071 alleles over all sampling locations. The number of private alleles ranged from 3 (CHIP) to 18 (MISS and TIPP) (Table 1). The fixation index ($F_{IS}$) ranged from −0.052 at the TN site to 0.072 at the ALL site (Table 1). Standard $F_{ST}$ values (Wright, 1969), $G_{ST}$, and Jost’s $D$ metrics were also calculated. Since the standard $F_{ST}$ values were likely depressed due to the high number of alleles at several loci, only $G_{ST}$ and Jost’s $D$ values were reported (Table 2). The $G_{ST}$ values indicated a high degree of population differentiation at all sampling sites except MISS and WIS (0.0623). The permutation test indicated a significant difference ($p = .005$) between samples found in the Upper Mississippi River Basin (CHIP, MISS, MER, and WIS) and the Ohio River Basin (ALL, TIPP, and TN).

3.1.2  | Population structure

The STRUCTURE analysis including all sampling locations indicated that the most likely value of $K$ was 2 (Evanno et al., 2005). Sampling sites in the Upper Mississippi River Basin (CHIP, MER, MISS, and
WIS) strongly clustered together and represent a different cluster from sites in the Ohio River Basin (ALL, TIPP, and TN) (Figure 2a). An AMOVA analysis also indicated significant genetic differentiation between the Mississippi and Ohio river basins \( (p \leq 0.0001) \). Despite sampling sites clustering into major river basins, with few exceptions, each sampling location showed a high degree of differentiation from other sites (Table 2). Because the sample sites exhibited substantial genetic differentiation, and populations may be structured hierarchically, with coarser structure obscuring more fine-scale structure, additional STRUCTURE analyses were conducted separately on the two initial clusters. STRUCTURE analysis of the Upper Mississippi River Basin indicated a \( K \) value of 3 (Figure 2b), clustering the MER and CHIP sites by themselves and the WIS, and MISS sites clustered together. Analysis of the Ohio River Basin indicated a \( K \) value of 2 with the ALL site clustered separately from the TIPP and TN sites (Figure 2c).

### 3.1.3 Estimation of migration

Contemporary migration rates estimated by BAYESASS indicated that migration between most sample sites was low. Only five of the 42 pairwise comparisons exhibited rates greater than 0.1 (Table 3). The permutation test indicated that there was significantly \( (p = 0.008) \) more migration occurring among sampling sites within the same drainage basin (Mississippi and Ohio basins) than sites found in different drainage basins. All instances of migration rates >0.1 were asymmetrical, with migrants moving in only one direction. Three of these were in the Upper Mississippi River Basin and two were in the Ohio River Basin (Table 3, Figure 3).

### 3.1.4 Changes in population size

No significant values indicating bottlenecks were obtained based on the Wilcoxon’s Sign Test. All sampling sites also exhibited an L-shaped distribution characterized by a high proportion of low-frequency alleles and a smaller proportion of alleles of intermediate frequencies indicating no recent bottlenecks had occurred.

### 3.2 Mitochondrial sequences

#### 3.2.1 Sequence diversity

Analyses were performed on 157 aligned DNA sequences of 883 nucleotide base pairs that had no missing data or ambiguous sites. Thirty-nine mtDNA haplotypes were detected across all sample sites. The number of haplotypes ranged from 5 to 12 per sampling site (Table 4) with the MISS site having the most \( (n = 12) \) and the TIPP site having the fewest \( (n = 5) \). The number of unique haplotypes ranged from 3 to 9 per site and accounted for about 85% of detected haplotypes overall (33 of 39). Haplotype diversity \( (H_d) \) ranged from 0.4620 (TIPP) to 0.9286 (TN) and the nucleotide diversity \( (\pi) \) ranged from 0.0008 (TIPP) to 0.0039 (WIS). A comparison of pairwise \( F_{ST} \) values indicated that all sample sites were significantly different from each other except the TIPP and ALL sites and the WIS and ALL sites (Table 5). A minimum spanning network indicated that only three haplotypes were shared across most sampling sites with a star-like pattern stemming from these common haplotypes consistent with population expansion (Slatkin & Hudson, 1991) (Figure 4).

#### 3.2.2 Population expansion

Evidence for a population expansion was found for all sites except WIS. The sum of square deviations and raggedness index indicated that these sites did not significantly differ from the population expansion model \( (p \geq 0.05) \) (Table 4). The detected population expansion is also indicated by the star-like pattern seen in the minimum spanning network (Slatkin & Hudson, 1991) (Figure 4). However, the WIS site indicated a significant deviation from the population expansion model in both the sum of square deviations \( (0.044, p = 0.01) \) and

![Figure 2](image-url)  
**Figure 2** (a) STRUCTURE clustering of the seven sampled Sheepnose populations into \( K = 2 \) groups corresponding to Upper Mississippi (CHIP, MER, MISS, and WIS) and Ohio River (ALL, TIPP, and TN) basins. (b) The Mississippi River Basin group clustered into \( K = 3 \) groups with CHIP and WIS clustering together and MER and MISS clustering on their own. (c) Ohio River Basin group clustered into \( K = 2 \) with TIPP and TN clustering together and ALL clustering on its own.
raggedness index (0.129, $p = .02$) (Table 4). Convergence could not be reached for the TN site, so it was dropped from the mismatch distribution analysis.

### DISCUSSION

This study found that despite the lack of gene flow and population isolation, there is still a high degree of genetic diversity among sampling sites at both the microsatellite and mitochondrial loci. Similar levels of genetic diversity have been found in other studies of both rare and common freshwater mussel species (Elderkin et al., 2007; Geist & Kuehn, 2005; Inoue et al., 2014; King et al., 1999; Zanatta & Murphy, 2007, 2008), possibly indicating abilities of some freshwater mussels to maintain genetic diversity despite isolation. Isolated Sheepnose populations may be large enough to maintain high level of genetic diversity and buffer populations against the erosive effects of genetic drift (Elderkin et al., 2007; Lande & Barrowclough, 1987). Additionally, Sheepnose are estimated to have lifespans of up to 30 years (Hove et al., 2015), such long-lived species can also buffer populations from the loss of genetic diversity due to drift (Hoffman et al., 2017). If this is true, efforts and resources aimed at conservation strategies such as propagation and translocations would be better directed toward regaining habitat suitability and connectivity (Olson & Vaughn, 2020).

### TABLE 3  Asymmetrical pairwise contemporary migration rates and associated 95% confidence intervals generated by BAYESASS

|        | ALL | CHIP | MER | MISS | TIPP | TN | WIS |
|--------|-----|------|-----|------|------|----|-----|
| ALL    | 0.012 | 0.012 | 0.012 | 0.012 | 0.011 | 0.012 | 0.012 |
| C.I. (95%) | (−0.010, 0.033) | (−0.0010, 0.033) | (−0.010, 0.034) | (−0.011, 0.034) | (−0.010, 0.033) | (−0.011, 0.034) |
| CHIP   | 0.012 | 0.0119 | 0.260* | 0.012 | 0.012 | 0.012 | 0.012 |
| C.I. (95%) | (−0.011, 0.034) | (−0.010, 0.034) | (0.240, 0.284) | (−0.038, 0.062) | (−0.011, 0.035) | (−0.011, 0.034) |
| MER    | 0.011 | 0.010 | 0.245* | 0.010 | 0.010 | 0.0101 |
| C.I. (95%) | (−0.010, 0.032) | (−0.009, 0.029) | (0.196, 0.294) | (−0.050, 0.071) | (−0.009, 0.029) | (−0.009, 0.029) |
| MISS   | 0.019 | 0.006 | 0.016 | 0.006 | 0.0057 | 0.006 |
| C.I. (95%) | (−0.006, 0.044) | (−0.005, 0.017) | (−0.005, 0.027) | (−0.032, 0.043) | (−0.005, 0.017) | (−0.005, 0.017) |
| TIPP   | 0.255* | 0.013 | 0.013 | 0.015 | 0.013 | 0.013 |
| C.I. (95%) | (0.200, 0.310) | (−0.011, 0.037) | (−0.011, 0.037) | (−0.009, 0.039) | (−0.011, 0.037) | (−0.011, 0.037) |
| TN     | 0.174* | 0.022 | 0.034 | 0.036 | 0.022 | 0.022 |
| C.I. (95%) | (0.087, 0.260) | (−0.019, 0.063) | (−0.006, 0.075) | (−0.018, 0.091) | (−0.039, 0.083) | (−0.019, 0.063) |
| WIS    | 0.014 | 0.014 | 0.016 | 0.248* | 0.014 | 0.014 |
| C.I. (95%) | (−0.012, 0.040) | (−0.012, 0.040) | (−0.010, 0.042) | (0.218, 0.278) | (−0.044, 0.071) | (−0.012, 0.040) |

*Indicates estimated rates >0.1.
Analysis of genetic structure indicates that currently, the Sheepnose consists of five genetically distinct populations. In the Mississippi River Basin, the populations consist of the Chippewa and the Wisconsin rivers, the Meramec River, and the Mississippi River. In the Ohio River Basin, the Allegheny River population is distinct from the Tennessee and Tippecanoe rivers. These populations are in turn composed of a number of distinct demes—isolated subpopulations that have a distinct gene pool. Both mitochondrial and microsatellite markers indicate a lack of connectivity between demes as evidenced by the unique mitochondrial haplotypes and private alleles and high values for pairwise $F_{ST}$, adjusted $G_{ST}$, and Jost’s $D$ (Tables 2 and 5). It is tempting to attribute the genetic differentiation detected between Sheepnose demes to be the result of habitat degradation that has fragmented the species range and isolated these groups from each other. Dams are considered highly detrimental to unionoid populations because they disrupt dispersal of host fishes (Watters, 1995), feeding ability (Bates, 1962; Negus, 1966), and alter stream flow and depth (Salmon & Green, 1983). The United States has about 75,000 dams and almost half (~30,000) can be found in the Mississippi River system (Graf, 1999). The direct impacts of dams on mussel health and their indirect impacts on dispersal may be contributing factors to the isolation between Sheepnose demes detected at the contemporary timescale. The construction of dams, river channelization, increased pollution, and invasive species are collectively contributing to the loss of habitat and changes in the distribution of freshwater mussels (Williams et al., 1989). However, the contemporary fragmentation of populations and demes due to anthropogenic barriers and habitat loss are not likely what is being detected in the analysis of the data for the Sheepnose. For long-lived species with long generation times, and large $N_e$ like most unionids, it would most likely take centuries for any genetic signature of these anthropogenic issues to be detected using microsatellites (Haag, 2012; Hoffman et al., 2017). Instead, we propose that the differentiation observed is the result of changes in climate and on the landscape that occurred during and subsequent to the Pleistocene.

| Site ID | Samples | Number of haplotypes | Number of unique haplotypes | Haplotype Diversity ($H_d$) | Nucleotide Diversity ($\pi$) | Sum of Square Deviations (SSD) | Raggedness Index |
|---------|---------|----------------------|-----------------------------|----------------------------|----------------------------|-----------------------------|-----------------|
| ALL     | 22      | 8                    | 3                           | 0.7922                     | 0.0034                     | 0.045 (0.19)              | 0.124 (0.14)    |
| C.I. (95%) |       |                      |                             | (0.7235, 0.8609)           | (0.0013, 0.0054)           |                             |                 |
| CHIP    | 19      | 7                    | 6                           | 0.7135                     | 0.0020                     | 0.025 (0.55)              | 0.079 (0.61)    |
| C.I. (95%) |       |                      |                             | (0.6110, 0.8160)           | (0.0006, 0.0033)           |                             |                 |
| MER     | 26      | 8                    | 4                           | 0.7231                     | 0.0023                     | 0.021 (0.68)              | 0.058 (0.73)    |
| C.I. (95%) |       |                      |                             | (0.6424, 0.8038)           | (0.0008, 0.0038)           |                             |                 |
| MISS    | 46      | 12                   | 9                           | 0.5749                     | 0.0019                     | 0.014 (0.12)              | 0.083 (0.56)    |
| C.I. (95%) |       |                      |                             | (0.4878, 0.6620)           | (0.0006, 0.0031)           |                             |                 |
| TIPP    | 19      | 5                    | 3                           | 0.4620                     | 0.0008                     | 0.004 (0.81)              | 0.112 (0.64)    |
| C.I. (95%) |       |                      |                             | (0.3262, 0.5978)           | (0.0001, 0.0015)           |                             |                 |
| TN      | 8       | 6                    | 4                           | 0.9286                     | 0.0024                     | -                           | -               |
| C.I. (95%) |       |                      |                             | (0.8442, 1.013)            | (0.0007, 0.0044)           |                             |                 |
| WIS     | 17      | 8                    | 4                           | 0.8750                     | 0.0039                     | 0.044 (0.01*)             | 0.129 (0.02*)   |
| C.I. (95%) |       |                      |                             | (0.8223, 0.9277)           | (0.0016, 0.0063)           |                             |                 |
| Total   | 157     | 39                   | 33                          | 0.7242                     | 0.0024                     |                             |                 |

Note: 95% confidence intervals are given for the haplotype and nucleotide diversity values. Sum of Square Deviations (SSD) and Raggedness Index results of the mismatch distribution analysis across all sites. Values and associated p-values are listed.

*Indicates significance.

|              | ALL | CHIP | MER | MISS | TIPP | TN |
|--------------|-----|------|-----|------|------|----|
| ALL          | 0.2921* |     |     |      |      |    |
| CHIP         | 0.4721* | 0.1543* |    |      |      |    |
| MER          | 0.3370* | 0.6357* | 0.7506* |      |      |    |
| MISS         | 0.0268 | 0.5008* | 0.6396* | 0.3669* |      |    |
| TIPP         | 0.1666* | 0.6401* | 0.7172* | 0.4299* | 0.0627* |    |
| TN           | 0.0607 | 0.3146* | 0.3146* | 0.3557* | 0.1748* | 0.3588* |

Note: *Indicates significant differentiation at a 0.05 significance level.
and before European contact in North America. The high number of private alleles and unique haplotypes found at each sample location indicate that these locations have been isolated for a substantial amount of time. Similar results have been found in other studies of freshwater mussels (Elderkin et al., 2007; Hewitt, 1996; Inoue et al., 2014; Jones et al., 2015; Tomilova et al., 2020). Together these studies indicate that for many freshwater mussels, population genetic structure is more reflective of long-term factors related to changes since the Pleistocene era and not from recent anthropogenic causes.

The amount of migration estimated between sampling locations was low overall and very low rates were estimated between the Ohio and Mississippi river basins (Table 3). In the Mississippi River Basin, the migration appears to be unidirectional—from tributaries to the Mississippi River, which could be due to several factors including dispersal via fishes (glochidia), sperm gene flow, or downstream displacement during floods. In the Ohio River Basin, the Allegheny River appears to be receiving migrants from the Tennessee and Tippecanoe rivers. This result is harder to interpret, as it seems unlikely that host fishes carrying glochidia larvae could effectively travel between these sites. These results in the Ohio basin may instead be due to the absence of samples from the mainstem Ohio River. Host fish vagility may also be an important factor on the population structure observed in the Sheepnose. Previous studies on freshwater mussels have invoked host fishes as contributing to or reinforcing the population structure (Chong & Roe, 2018; Zanatta & Wilson, 2011). The most recent information indicates that hosts of the Sheepnose include cyprinids (Hove et al., 2015) which may lack the dispersal capabilities of larger riverine fishes (Comte & Olden, 2018) and therefore contribute to the isolation between populations and demes.

Despite the detected isolation, high levels of diversity were still observed within in demes. With the exception of the WIS deme, the analysis of mitochondrial data revealed a pattern consistent with expanding populations. Such a pattern may be a result of responses of the Sheepnose to climate change oscillations during the Pleistocene (Alberdi et al., 2015). North American freshwater systems were heavily impacted by the expansion and contraction of Pleistocene glaciers, and the population structure of freshwater organisms often reflect these events (Berendzen et al., 2010; Inoue & Berg, 2017; Inoue et al., 2014; Jones et al., 2015; Mathias et al., 2016; Pielou, 2008). Since the retreat of glaciers following the last glacial maxima, aquatic species in the Upper Mississippi River Basin have had opportunity to expand their ranges northward into previously glaciated
regions (Hewitt, 1996, 2000; Inoue et al., 2014; Pielou, 2008; Stewart & Lister, 2001).

Our results indicate that the contemporary pattern of low gene flow and isolation occurring among populations is the result of prehistoric changes to the landscape that have eliminated populations and introduced barriers to gene flow. Although anthropogenic influences may be too recent to explain the observed patterns, they may be reinforcing existing genetic differences. It appears that the long lifespan of the Sheepnose may be delaying the reduction in genetic diversity typically associated with isolation. However, if isolation persists, it is possible that genetic diversity in these demes will start to erode. Our results suggest that efforts should be made to re-establish gene flow among demes to support the maintenance of genetic diversity. For example, Sheepnose could be re-established within basins to facilitate connectivity between demes and within populations. The isolation of mussels into demes makes them more sensitive to stochastic events (Fagan et al., 2005), however, managers should be cautious about disrupting any local adaptations that may have been acquired by these demes (Fitzpatrick et al., 2015; Lean et al., 2017). Propagation and re-introduction operations for re-establishing Sheepnose within their historical ranges (Geist, 2010; Jones et al., 2006; Minckley, 1995) should aim to avoid disrupting localized adaptations; therefore, translocations between different populations are not recommended (Fitzpatrick et al., 2015; Lean et al., 2017).

Overall, the Sheepnose appear to have maintained a surprising amount of genetic diversity within populations despite the lack of connectivity detected among demes. Mechanisms that may be buffering the erosion of genetic diversity such as their long lifespan and potentially high effective population sizes may not continue to withstand the negative effects of prolonged isolation and result in the eventual disappearance of the Sheepnose from portions of its current range. To effectively conserve this species, managers should create objectives focused on re-establishing and maintaining enough suitable habitat so Sheepnose can naturally re-establish populations and maintain connectivity between demes. Improved connectivity can facilitate the maintenance of genetic diversity and adaptive potential in the face of climate change and other emerging stressors.

ACKNOWLEDGMENTS
Numerous individuals provided samples for this study, and thanks are extended (in no particular order) to D. Heath, J. Clayton, G. Zimmerman, D. Hubbs, S. Ahlstedt, C. Barnhart, T. Brady, N. Eckert, J. Faiman, J. Hundley, S. McMurray, R. Villela, B. Fisher, B. Seitman, G. Wege, M. Hove, A. Fullenkamp, D. Helms, S. Gritters, K. Cummings, J. Tiemann, A. Price, S. Bales, D. Shasteen, A. Berger, and J. Schwab. R. Creasey and A. Casey assisted with laboratory work. This project was funded by the US Fish and Wildlife Service and the Iowa Department of Natural Resources, Conservation and Recreation Divisions, and fulfills part of the requirement for a M.S. degree at Iowa State University. We also thank John Nason and Julie Blanchong for direction and advice during data analysis and for their input on this project and the anonymous reviewers for their helpful comments.

CONFLICT OF INTEREST
The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

AUTHOR CONTRIBUTIONS
Sara Schwarz: Data curation (equal); Formal analysis (lead); Methodology (equal); Writing – original draft (lead); Writing – review & editing (equal). Kevin Roe: Conceptualization (lead); Funding acquisition (lead); Project administration (lead); Writing – review & editing (equal).

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://doi.org/10.5061/dryad.gxd2547mm.

DATA AVAILABILITY STATEMENT
Mitochondrial DNA sequence data have been submitted to GenBank (Accession Number: MH853483). Microsatellite genotypes have been submitted to DRYAD (https://doi.org/10.5061/dryad.gxd2547mm).

ORCID
Kevin J. Roe https://orcid.org/0000-0002-2228-3019

REFERENCES
Alberdi, A., Gilbert, M. T. P., Razgour, O., Aizpurua, O., Aihartza, J., & Garin, I. (2015). Contrasting population-level responses to Pleistocene climatic oscillations in an alpine bat revealed by complete mitochondrial genomes and evolutionary history inference. Journal of Biogeography, 42(9), 1689–1700.
Andersen, L. W., Fog, K., & Damgaard, C. (2004). Habitat fragmentation causes bottlenecks and inbreeding in the European tree frog (Hyla arborea). Proceedings of the Royal Society B: Biological Sciences, 271(1545), 1293.
Bandelt, H., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, 16(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036
Bates, J. M. (1962). Impact of impoundment on the mussel fauna of Kentucky Reservoir, Tennessee River. The American Midland Naturalist, 68, 222–236.
Berendzen, P. B., Dugan, J. F., & Gamble, T. (2010). Post-glacial expansion into the paleozoic plateau: evidence of an ozarkian refugium for the Ozark minnow Notropis nubilus (Teleostei: Cypriniformes). Journal of Fish Biology, 77, 1114–1136.
Berg, D. J., Haag, W. R., Guttmann, S. I., & Sickel, J. B. (1995). Mantle biopsy: a technique for non-destructive tissue-sampling of
Fagan, W. A., Aumann, C., Kennedy, M., & Unmack, P. (2005). Rarity, friction.

Ferguson, C. D., Blum, M. J., Raymer, M. L., Eackles, M. S., & Krane, D. E. (2013). Population structure, multiple paternity, and long-distance transport of spermatozoa in the freshwater mussel, Cumberlandia monodonta (Bivalvia: Unionidae): Evidence of high frequency data. The Cambridge Philosophical Society, 181(2), 163–182.

Elderkin, C. L., Christian, A. D., Vaughn, C. C., Metcalfe-Smith, J. L., & Berg, D. J. (2007). Population genetics of the freshwater mussel, Amblyma plicata (Say 1817) (Bivalvia: Unionidae): Evidence of high dispersal and post-glacial colonization. Conservation Genetics, 8, 355–372. https://doi.org/10.1007/s10592-006-9175-0

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analysis under Linux and Windows. Molecular Ecology Resources, 10, 564–567.

Fagan, W., Aumann, C., Kennedy, M., & Unmack, P. (2005). Rarity, fragmentation and the scale dependence of extinction risk in desert fishes. Ecology, 86, 34–41. https://doi.org/10.1890/04-0491

Ferguson, C. D., Blum, M. J., Raymer, M. L., Eackles, M. S., & Krane, D. E. (2013). Population structure, multiple paternity, and long-distance transport of spermatozoa in the freshwater mussel Lampsilis cardium (Bivalvia: Unionidae). Freshwater Science, 32(1), 267–282.

Fitzpatrick, S. W., Gerberich, J. C., Kronenberger, J. A., Angeloni, L. M., & Funk, W. C. (2015). Locally adapted traits maintained in the face of high gene flow. Ecology Letters, 18(1), 37–47. https://doi.org/10.1111/ele.12388

Francis, R. M. (2016). POPHELPER: an R package and web app to analyze and visualize population structure. Molecular Ecology Resources, 17(1), 27–32.

Frankel, O. H., & Soulé, M. E. (1981). Conservation and evolution. Cambridge University Press.

Geist, J. (2010). Strategies for the conservation of endangered freshwater pearl mussels (Margaritifera L.); a synthesis of Conservation Genetics and Ecology. Hydrobiologia, 644(1), 69–88.

Geist, J., & Kuehn, R. (2005). Genetic diversity and differentiation of central European freshwater pearl mussel (Margaritifera L.) populations: implications for conservation and management. Molecular Ecology, 14(2), 425–439.

Gerlach, G., Jueterbock, A., Kraemer, P., Depperman, J., & Harmand, P. (2010). Calculations of population differentiation based on Gst and D: forget Gst not all of statistics! Molecular Ecology, 19, 3845–3852.

Goudet, J., & Thibout, J. (2015). Package hierfstat. R package version 0.04-22. http://www.r-project.org

Graf, D. L., & Cummings, K. S. (2007). Review of the systematic and global diversity of freshwater mussel species (Bivalvia: Unionoida). Journal of Molluscan Studies, 73, 291–314. https://doi.org/10.1093/mollus/eym029

Graf, W. L. (1999). Dam nation: A geographic census of American dams and their large-scale hydrologic impacts. Water Resources Research, 35(4), 1305–1311. https://doi.org/10.1029/1999WR900016

Haag, W. R. (2012). North American freshwater mussels. Cambridge Press.

Haag, W. R., & Williams, J. D. (2014). Biodiversity on the brink: an assessment of conservation strategies for North American freshwater mussels. Hydrobiologia, 735, 45–60. https://doi.org/10.1007/s10750-013-1524-7

Harpending, H. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Human Biology, 66(4), 591–600.

Harpending, H. C., Sherry, S. T., Rogers, A. R., & Stoneking, M. (1993). Genetic structure of ancient human populations. Current Anthropology, 34(4), 483–496.

Hedrick, P. W. (2005). A standardized genetic differentiation measure. Evolution, 59(8), 1633–1638. https://doi.org/10.1111/j.0014-3820.2005.tb01814.x

Henley, W. F., Grobler, P. J., & Neves, R. J. (2006). Non-invasive method to obtain DNA from freshwater mussels (Bivalvia: Unionidae). Journal of Shellfish Research, 25, 975–977.

Hewitt, G. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society, 58(3), 247–276. https://doi.org/10.1006/bijl.1996.0035

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. Nature, 405, 907–913. https://doi.org/10.1038/35016000

Hoffman, J. R., Willoughby, J. R., Swanson, B. J., Pangle, K. L., & Zanatta, D. T. (2017). Detection of barriers to dispersal is masked by long lifespans and large population sizes. Ecology and Evolution, 7, 9613–9623. https://doi.org/10.1002/ece3.3470

Hove, M. C., Sietsman, B. E., Berg, M. S., Frost, E. C., Wolf, K., Brady, T. R., Boyer, S. L., & Hornbach, D. J. (2015). Early life history of the sheep-nose (Plethobasus cyphus) (Mollusca: Bivalvia: Unionoida). Journal of Natural History, 50(9–10), 523–542.

Hulce, D., Li, X., & Snyder-Leiby, T. (2011). GeneMarker® Genotyping Software: Tools to Increase the Statistical Power of DNA Fragment Analysis. Journal of Biomolecular Techniques, 22(Supp), S53–S36.

Inoue, K., & Berg, D. J. (2017). Predicting the effects of climate change on population connectivity and genetic diversity of an imperiled freshwater mussel, Cumberlandia monodonta (Bivalvia: Margaritiferae), in riverine systems. Global Change Biology, 23(1), 94–107.

Inoue, K., Monroe, E. M., Elderkin, C. L., & Berg, D. J. (2014). Phylogeographic and population genetic analyses reveal Pleistocene isolation followed by high gene flow in a wide ranging, but endangered, freshwater mussel. Heredity, 112, 282–290. https://doi.org/10.1038/hdy.2013.104

Jones, J. W., Hallerman, E. M., & Neves, R. J. (2006). Genetic management guidelines for captive propagation of freshwater mussels (Unionidae). Journal of Shellfish Research, 25(2), 527–535.

Jones, J. W., Neves, R. J., & Hallerman, E. M. (2015). Historical demography of freshwater mussels (Bivalvia: Unionidae): genetic evidence for population expansion and contraction during the late Pleistocene and Holocene. Biological Journal of the Linnean Society, 114(2), 376–397. https://doi.org/10.1111/bij.12437

Jost, L. (2008). GS, and its relatives do not measure differentiation. Molecular Ecology, 17(18), 4015–4026.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially, clonal, and/or sexual reproduction. Peer J, 2, e281.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.
J. E. (1989). Fishes of North America endangered, threatened, or of special concern: 1989. *Fisheries*, 14(6), 2–20.

Williams, J. D., Warren, M. L. Jr, Cummings, K. S., Harris, J. L., & Neves, R. J. (2011). Conservation Status of Freshwater Mussels of the United States and Canada. *Fisheries*, 18(9), 6–22.

Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163(3), 1177–1191.

Wright, S. (1969). *Evolution and the Genetics of Populations*, Vol. 2. University of Chicago Press.

Zanatta, D. T., & Murphy, R. W. (2007). Range-wide population genetic analysis of the endangered northern riffleshell mussel, *Epioblasma torulosa rangiana* (Bivalvia: Unionoida). *Conservation Genetics*, 8, 1393–1404.

Zanatta, D. T., & Murphy, R. W. (2008). The phylogeographical and management implications of genetic population structure in the imperiled snuffbox mussel, *Epioblasma triquetra* (Bivalvia: Unionidae). *Biological Journal of the Linnean Society*, 93, 371–384.

Zanatta, D. T., & Wilson, C. C. (2011). Testing congruency of geographic and genetic population structure for a freshwater mussel (Bivalvia: Unionoida) and its host fish. *Biological Journal of the Linnean Society*, 103(3), 669–685.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

**How to cite this article:** Schwarz, S., & Roe, K. J. (2022). Population structure and gene flow in the Sheepnose mussel (*Plethobasus cyphyus*) and their implications for conservation. *Ecology and Evolution*, 12, e8630. [https://doi.org/10.1002/ece3.8630](https://doi.org/10.1002/ece3.8630)