FINAL PERFORMANCE REPORT

As Required by

THE ENDANGERED SPECIES PROGRAM

TEXAS

Grant No. TX ET-148-R

F12AP01132

Endangered and Threatened Species Conservation

Genetic demography of endemic and endangered taxa in springs of the Edwards Plateau

Prepared by:

Dr. Chris Nice

Carter Smith
Executive Director

Clayton Wolf
Director, Wildlife

30 November 2015

INTERIM REPORT

STATE: ____Texas_____ GRANT NUMBER: ___TX E-148-R-1__

GRANT TITLE: Genetic demography of endemic and endangered taxa in springs of the Edwards Plateau

REPORTING PERIOD: ____1 September 2012 to 31 August 2015

OBJECTIVE(S). To perform a comparative population genetics study of *Eurycea* salamanders, *Heterelmis* riffle beetles, and *Stygobromus* amphipods to ask if the landscape or life history strategies of these taxa shape their dispersal patterns and to characterize the relationship between variation in spring discharge and population growth or decline.

Segment Objectives:

Task 1: Specimen collection. (Project Year1) Use DNA extracted from samples used in previous phylogeographic studies involving the spring endemic taxa of interest and supplement the population sampling such that all three taxa are sampled from springs or caves where they coexist. We will also sample previously un-sampled intervening springs or caves that potentially exchange migrants with some of the previously sampled sites.

Task 2: "Next-generation" sequencing. (Project Years 1-2) We will use thousands of short DNA sequences randomly scattered throughout the genome of the taxa of interest to make inferences about population dynamics. We will generate reduced genomic complexity libraries for each individual using a restriction fragment-based procedure, as we have successfully implemented in previous studies. We will sequence sets of 384 individuals (four sets total, one each for *Heterelmis* and *Stygobromus*, and, because of the large size of the *Eurycea* genome). DNA sequencing of the four libraries will be performed by the National Center for Genome Research (Sante Fe, NM, USA). We expect to generate approximately 160 million short DNA sequences from each of the four libraries, from which thousands of informative genetic markers will be selected for data analysis.

Task 3: Data analysis, including Approximate Bayesian Computation. (Project Years 2-3) We will use simulation methods to analyze this large amount of genetic data to make inferences about complex parameters such as dispersal rate and population growth. We will use relevant ecological and spatial information in modeling, such as spring discharge and distance among springs or caves, both of which may affect dispersal rates. We will obtain spring discharge data relevant sources.

Significant Deviations:

None.

Summary Of Progress:

Please see Attachment A.

Location: Texas State University, San Marcos, Hays County, Texas.

Cost: Costs were not available at time of this report, they will be available upon completion of the Final Report and conclusion of the project.

Prepared by: Craig Farquhar Date: 30 November 2015

Approved by: _____ Date: 30 November 2015
C. Craig Farquhar

ATTACHMENT A

Genetic Demography of Endemic and Endangered Taxa in Springs of the Edwards Plateau

Final Report - November 29, 2015

Principal Investigator: Chris Nice
Department of Biology, Texas State University
601 University Drive
San Marcos, TX 78666
Phone: 512-245-3358
Fax: 512-245-8713
Email: ccnice@txstate.edu

Co-Principal Investigator: Lauren Lucas
Department of Biology, Texas State University
601 University Drive
San Marcos, TX 78666
Phone: 512-245-3358
Fax: 512-245-8713
Email: llucas@txstate.edu

Abstract

Genotyping-by-Sequencing (GBS) methods were used to investigate the distribution of population genetic variation within and among population samples of spring-endemic organisms from the Edwards Plateau region of central Texas. Reduced-representation genomic libraries were constructed for four taxa: *Eurycea* salamanders, *Heterelmis* riffle beetles, *Stygobromus* amphipods, and *Stygoparnus* dryopid beetles. These libraries were sequenced with the Illumina HiSeq DNA sequencer and produced data for 100's to 1000's of variable loci for each taxon. For the first objective of this project, these data were used to describe genetic variation and to test model-based hypotheses about the patterns of gene exchange within Comal Springs, New Braunfels, Texas. Comal Springs was the specific focus because all four taxa co-occur there and these populations are considered to be of conservation concern. Approximate Bayesian Computation analyses indicated support for an island model of nearly equal gene flow among subpopulations at Comal Springs for all four taxa. Estimates of migration were relatively high, indicating that the entire Comal Springs complex might be considered as a single conservation unit for each taxon. However, investigations of associations of allelic variation with environmental / habitat variables also found evidence of some local-adaption in all four taxa, suggesting that the subpopulations might contain unique genetic variation despite high levels of gene flow. For the second objective, GBS data were used to examine the relationships between the Comal Springs populations of *H. comalensis* and *S. pecki* and congeneric species and populations across the Edwards Plateau. Levels of genetic diversity within *H. comalensis* and *S. pecki* are not substantially different from diversity observed in congeners. Estimates of pairwise population differentiation and genetic distance were used to illustrate relationships based on allele frequency similarities. *Heterelmis comalensis* appears to be closely related to *H. glabra*, but quite distinct from *H. vulnerata*. *Stygobromus pecki*, sampled from Comal Springs and Hueco Springs, have the least amount of differentiation among *Stygobromus* populations sampled, but samples of *S. longipes*, *S. dejectus* and an unknown or undescribed taxon from Fessenden Springs (also sometimes known as Stockman Springs) are all closely related to *S. pecki*. *Stygobromus flagellatus* sampled from San Marcos Springs is distantly related to the other taxa. These data form the basis of an improved understanding of the patterns of geographic genetic variation for these spring-endemic taxa in the Edwards Plateau.

Introduction

Eurycea salamanders, *Heterelmis* riffle beetles, *Stygobromus* amphipods, and *Stygoparnus* dryopid beetles have endemic and endangered members in springs of the Edwards Plateau. It is important for the U.S. Fish and Wildlife Service (USFWS) to understand the biology of these endangered taxa, especially their population dynamics such as dispersal rates and population sizes, to manage and protect them. However, direct observations (e.g., mark and recapture) are difficult in aquatic biota, particularly those that spend some or all of their time in subterranean habitats. Population genetic techniques have been used over the last two decades to overcome this challenge by providing an indirect method for estimating

population parameters. Previous population genetic assessments were undertaken for the three taxa (Gonzales, 2008; Lucas *et al.*, 2009; Ethridge *et al.*, 2013). These studies used the population genetic techniques available at the time and thereby provided important baseline information, however, they did not provide a complete picture of the evolutionary history of these taxa nor a comprehensive sampling of genetic variation at the genomic level. Several recent advances in genetic sampling ("next-generation" sequencing) and genetic analyses (Approximate Bayesian Computation, or ABC methods) now make it possible for researchers to make more accurate inferences about population dynamics by sampling much more of the variability among individuals at the genome level. Here we use genome-wide genetic data produced using "next-generation" sequencing technology to examine patterns of genetic differentiation among populations of endangered, spring-endemic organisms of the Edwards Plateau.

This project consists of two main components. First, we examined patterns of gene flow within the Comal Springs complex in New Braunfels, Texas, a site where *Eurycea*, *Heterelmis*, *Stygobromus* and *Stygoparnus* co-occur. ABC methods were used to test alternative models of gene flow in a comparative study of all four spring-endemic taxa. Summary statistics describing genetic variation and inferences of demographic history were compared in the context of variation in habitat affinities and body size among taxa, and variation in spring discharge among sites.

In the second part of the study, sampling localities for *Heterelmis* and *Stygobromus* beyond Comal Springs were added to provide a broader geographical perspective on the organization of genetic variation. This sampling included other nominal taxa (species) besides *Heterelmis comalensis* and *Stygobromus pecki*, the species found in Comal Springs. Here we used genome wide data to calculate measures of population differentiation to illustrate the patterns of genetic variation among nominal taxa and across geographic space.

Objective

To perform a comparative population genetics study of *Eurycea* salamanders, *Heterelmis* riffle beetles, *Stygobromus* amphipods, and *Stygoparnus comalensis* dryopid beetles to ask if the landscape or life history strategies of these taxa shape their dispersal patterns and to characterize the relationship between variation in spring discharge and population growth or decline.

Location

Samples of the focal taxa were collected from spring sites in central Texas. Table 1 and Fig. 1 provide sampling information.

Table 1: Details of sampling. Sampling locations are numbered as in Fig 1. Sample sizes are provided in parentheses after the name of the nominal taxon. Where there is uncertainty regarding taxonomic designation, the species epithet is "?". **E** indicates endangered taxa.

Site No.	Site Name	Nominal Taxa (n)	Aquifer	River	County	Lat. Long.
1	Caroline Spring	*H. glabra* (10)	Edwards-Trinity	Rio Grande	Terrell	30.469, -101.803
2	Dolan Springs	*H. glabra* (10)	Edwards-Trinity	Rio Grande	Terrell	29.896, -100.982
3	Indian Springs	*H. ?* (10)	Edwards-Trinity	Rio Grande	Val Verde	29.663, -100.927
4	Fessenden Springs	*H. glabra* (11) *S. ?* (11)	Edwards-Trinity	Guadalupe	Kerr	30.167, -99.343
5	Guadalupe River – Hwy 474	*H. vulnerata* (10)	Trinity	Guadalupe	Kendall	29.894, -98.670
6	Cave Without A Name	*S. longipes* (4)	Trinity	Guadalupe	Kendall	29.886, -98.618
7	Cascade Caverns	*S. dejectus* (13)	Trinity	Cibolo Creek	Kendall	29.764, -98.679
8	Stealth Cave	*S. dejectus* (5)	Trinity	Cibolo Creek	Bexar	29.660, -98.559
9	Magic Springs	*S. longipes* (1)	Trinity	Guadalupe	Comal	29.907, -98.444
10	CM Cave	*S. longipes* (2)	Trinity	Guadalupe	Comal	29.911, -98.433
11	Fern Bank	*H. glabra* (25)	Trinity	Blanco	Hays	29.994, -97.996
12	San Marcos Springs	*H. comalensis* (28)**E** *S. flagellatus* (28)	Edwards	Blanco	Hays	29.894, -97.927
13	San Marcos River	*H. vulnerata* (10)	Edwards	Blanco	Hays	29.864, -97.927
14	Hueco Springs	*S. pecki* (12)**E**	Edwards	Guadalupe	Comal	29.760, -98.141
15	Comal Springs	*H. comalensis* (70)**E** *S. pecki* (77)**E** *E. neotenes* (60) *S. comalensis* (53)	Edwards	Guadalupe	Comal	29.718, -98.132
16	Plum Creek	*H. vulnerata* (11)	Carrizo	Blanco	Caldwell	29.655, -97.600
17	Guadalupe River – Hwy 183	*H. vulnerata* (10)	Carrizo	Guadalupe	Gonzales	29.485 -97.449

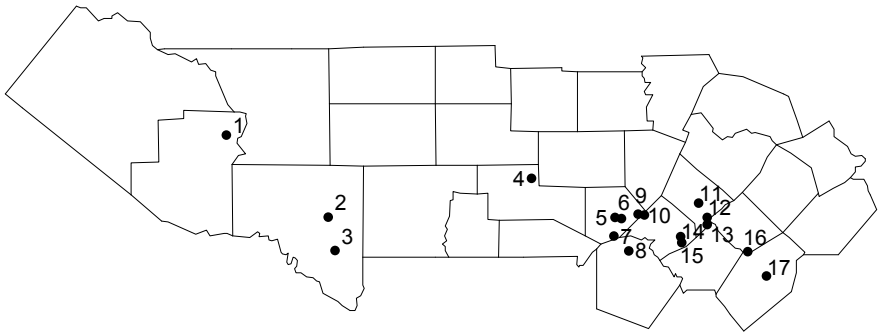

Figure 1: Map of sampling localities in the Edwards Plateau region of Texas. Numbers for the sampling sites correspond to Table 1.

Part 1: Comal Springs:

Pervasive gene flow across critical habitat for four narrowly endemic, sympatric taxa

Lauren K. Lucas[1,2], Zachariah Gompert[2], J. Randy Gibson[3], Katherine L. Bell[1], C. Alex Buerkle[4] and Chris C. Nice[1]

[1]Department of Biology, Texas State University, San Marcos, TX 78666, USA
[2]Department of Biology, Utah State University, Logan, UT 84322, USA
[3]U.S. Fish & Wildlife Service, San Marcos Aquatic Resources Center, San Marcos, TX 78666 USA
[4]Department of Botany, University of Wyoming, Laramie, WY 82071, USA

Summary

1. We studied genetic variation in four endangered animal taxa in the largest freshwater spring complex in the southwestern USA, Comal Springs (TX): *Eurycea* salamanders, *Heterelmis* riffle beetles, *Stygobromus* amphipods and *Stygoparnus* dryopid beetles. They inhabit a spring complex with stable conditions, which is threatened by climate change and aquifer withdrawals. The four taxa vary in their habitat affinities and body sizes.
2. We used genotyping-by-sequencing to obtain hundreds to thousands of genetic markers to accurately infer the evolutionary history of the taxa. We used approximate Bayesian computation to test models of gene flow and compare the results among taxa. We also looked for evidence that would suggest local adaptation within the spring complex.
3. The island model (equal gene flow among all subpopulations) was the most probable of the five models tested, and all four taxa had high migration rate estimates.
4. Small numbers of single nucleotide polymorphisms (SNPs) in each taxon were associated with environmental parameters and provide some evidence for potential local adaptation to variable conditions within Comal Springs.
5. We discuss how the results of this study can add to the habitat conservation plan for Comal Springs. If part of the spring system dries, migrants may recolonize from elsewhere within the spring complex. However, genetic variants affecting survival in particular habitats could be lost during such droughts.

Introduction

Comparative phylogeography can help explain the mechanisms responsible for the distributions of different taxa on a landscape (Knowles & Maddison, 2002). Taxa with similar histories might exhibit similar patterns of genetic variation because of the landscape they share in common. Barriers in the landscape, such as impermeable layers in an aquifer, may

prevent dispersal among habitats for all freshwater taxa sharing the landscape (Whitaker *et al.*, 2003; Marten *et al.*, 2006). However, greater dispersal ability may promote more dispersal among habitats for some taxa. For example, freshwater invertebrates with active dispersal such as adult beetles with flight capabilities may have more widespread distributions (Bilton *et al.*, 2001) than those with passive dispersal such as freshwater amphipods, which typically drift in the water column to disperse and consequently display substantial population structure within small geographic ranges (Murphy *et al.*, 2010; Robertson *et al.*, 2014). The relative importance of landscape and dispersal ability in shaping biogeographical patterns depends on the specific landscape and its inhabitants (Page & Hughes, 2014).

The results of a comparative phylogeographic analysis can facilitate the development of conservation management plans for threatened or endangered species to maintain or restrict gene flow, and thereby manage the genetic diversity of populations (Slatkin, 1987; Hermoso *et al.*, 2011). If all taxa in a common habitat show similar patterns of differentiation and gene flow, the entire habitat including conduits to gene flow could be conserved and populations could be managed together. If the taxa have different patterns of differentiation and gene flow, it might be important to manage the taxa separately. For example, if all populations of a taxon are isolated from one another, all populations could be conserved and managed as separate units to maintain biodiversity. On the opposite spectrum if all individuals can disperse across the taxon's range, we might only need to conserve and manage a subset of all populations to maintain biodiversity (Hughes *et al.*, 2013). In addition to understanding gene flow patterns among populations, we need to understand local adaptation, or ecological differences among populations, when making conservation management plans. Whereas gene flow can maintain genetic variation and combat inbreeding depression, the genetic mixture of populations that are adapted to different environmental conditions can lead to fitness reductions (outbreeding depression; Slatkin, 1987; Lenormand, 2002). Knowledge of local adaptation is useful for managing gene flow to prevent inbreeding and outbreeding depression and effectively conserving locally adaptive variation (Storfer, 1999).

Comparative phylogeography and conservation studies have typically focused on a relatively few genetic markers in line with limited financial time and resources. However, a specific region of the genome might tell only a small piece of the organism's evolutionary history. For example, the evolutionary history of the mitochondrial genome may not be indicative of the history of the nuclear genome, due to mitochondrial introgression (e.g., Gompert *et al.*, 2006) or stochastic variation. Similarly, different regions of DNA with different functional constraints might evolve at different rates (Patterson, 1999). Genotyping-by-sequencing is a cost-effective means to sample thousands of markers across the genome to provide a more accurate representation of evolutionary history (Elshire *et al.*, 2011). Plus, many markers are needed to search out genetic regions that underly local adaptation (Knowles & Maddison, 2002). Comparative phylogeography and conservation studies also traditionally have based inferences and decisions on an inferred gene tree. However, there may be multiple demographic models that fit a gene tree equally well (Nielsen & Beaumont, 2009). Instead, it is possible to determine which population histories are more compatible with the data and which are less by considering explicit demographic models. Approximate Bayesian computation (ABC, Beaumont *et al.*, 2002; Csilléry *et al.*, 2010) allows model choice by calculating the relative posterior probabilities of many different models representing any number of de-

mographic scenarios that take aspects of the system into account (Bertorelle *et al.*, 2010). Furthermore, traditional methods for estimating gene flow rates assume that the populations have reached equilibrium between genetic drift and gene flow (Slatkin, 1987), but ABC, like other more recent methods (Hey & Nielsen, 2004), decomposes inter-population genetic similarity due to recent divergence from genetic similarity due to ongoing gene flow. Herein we use genotyping-by-sequencing and ABC to discern the patterns of gene flow among subpopulations, and we investigate evidence supporting local adaptation in a comparison of four spring-endemic taxa in their critical habitat.

The aquifer systems of the Edwards Plateau in Texas, USA, are biotically diverse and home to a large number of locally endemic species (Longley, 1981; Brune, 2002, e.g., Figure 2). This endemism is a product of dissolution of limestone through time that has created numerous caves and springs. Both vicariant and dispersal events may have shaped the patterns of endemism in the Edwards Plateau, but the relative importance of each is unknown. Despite their limitations (namely the use of few genetic markers), previous phylogeography studies have quantified population genetic structure or the amount of gene flow among populations of some endemic members of the Edwards Plateau aquifer system. For example, **?** found a pattern of isolation by distance (IBD) and no recent gene flow among populations of neotenic *Eurycea* salamanders spread across two aquifers and two river drainages in the Edwards Plateau. Whereas across the same landscape, T. Gonzales and colleagues (Gonzales, 2008) found no pattern of IBD among sampled populations of the riffle beetles *Heterelmis comalensis* and *Heterelmis glabra*, perhaps suggesting different dispersal capabilities of *Heterelmis* and *Eurycea*. Here we revisit phylogeography studies like these but at a smaller scale, within an environmentally-sensitive freshwater spring complex, Comal Springs in south-central Texas. Identifying the scale at which a population becomes structured is important for management. After all, endemic species in very small areas can have population differentiation and low levels of gene flow, such as the case with the desert spring amphipod, *Wangiannachiltonia guzikae*, found in less than a one km^2 area in the Great Artesian Basin of central Australia (Robertson *et al.*, 2014).

Comal Springs consists of 425 spring openings that feed into Landa Lake and six spring runs (C. Norris et al., unpubl. data). The spring system covers a distance of 1,300 meters and is the largest spring system in the southwestern United States. The average flow between 1993 and 2008 was 291 cfs (http://www.eahcp.org/). Although Comal Springs reportedly has the greatest discharge of any springs in the southwestern USA, the flows can diminish rapidly during drought conditions. In the most extreme example, the springs completely ceased to flow from June 13 to November 3, 1956 (Brune, 2002). Spring runs R4 and R5 (Figure 3 B) are the most susceptible to the the cessation of flow during droughts (C. Norris et al., unpubl. data). Spring runs R1, R2 and R3 (Figure 3 B) discharge from the upthrown Comal Springs fault block, and these springs stop flowing when the water levels in the upthrown block drop below the elevation of the individual springs (189.9 meters above mean sea level (mamsl), S. Johnson and G. Schindel, unpubl. data). Approximately 75% of the total spring flow from Comal Springs is from the downthrown Artesian fault block in the bottom of Landa Lake (UP, WS, SI and BW, Figure 3 B), and during periods of low flow, Comal Springs is entirely fed by water from the Artesian fault block. The Artesian fault block stops flowing when groundwater elevation drops below 188.7 mamsl (S. Johnson and

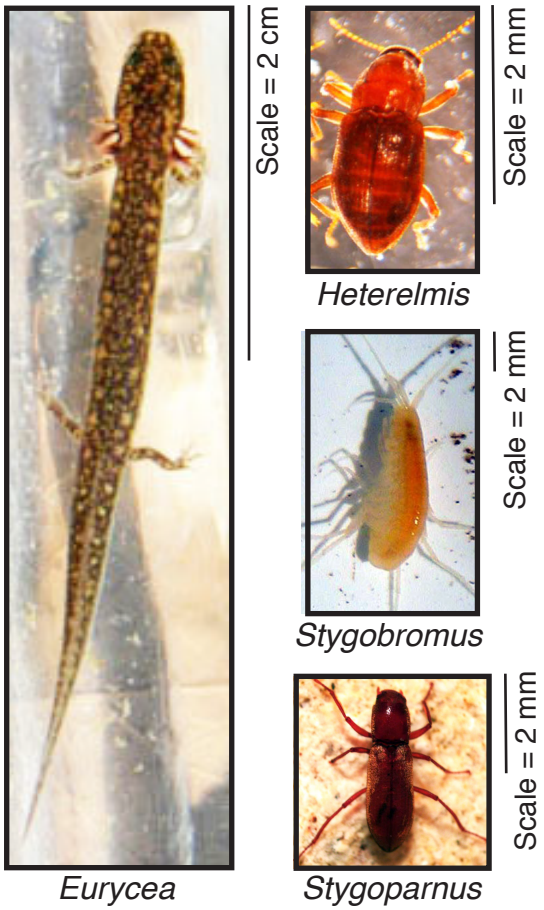

Figure 2: Spring-endemic taxa in Comal Springs (New Braunfels, TX): Comal Springs salamander (*Eurycea* sp.), Comal Springs riffle beetle (*Heterelmis comalensis*), Peck's cave amphipod (*Stygobromus pecki*) and Comal Springs dryopid beetle (*Stygoparnus comalensis*). Scale bars illustrate their true lengths: *Eurycea* is 46 mm, *Heterelmis* is 2 mm, *Stygobromus* is 10 mm, *Stygoparnus* is 4 mm.

G. Schindel, unpubl. data). The flow decreases because of climate variation and general warming (Loáiciga *et al.*, 2000) and withdrawals from the Edwards Aquifer via wells for municipal, irrigation, livestock, and industrial or commercial purposes (i.e., two million water consumers, http://www.eahcp.org/). Despite the intentional aquifer withdrawls, Comal Springs is home to several endemic and federally-listed species. They are found in several localities of suitable habitat within the Comal Springs complex (referred to as subpopulations herein; each colored two-letter code in Figure 3 B-D is placed at a subpopulation). All subpopulations have relatively stable temperatures of about 23.3 °C, nearly neutral pH and rocky substrate. Some fluctuation in habitat conditions can be beneficial for maintaining genetic variation, but aquifer pumping reduces the flux of oxygen and dissolved organic carbon downstream, which alters redox reactions and pH, respectively (Humphreys, 2009), and severe decreases in spring discharge may reduce habitat suitability in the long term. Similarly, temporal variation in discharge within an aquifer can complicate patterns of dispersal routes for spring- or cave-endemic taxa. Floods may raise the water table sufficiently to open new subterranean conduits or carry organisms via aboveground rivers. Droughts may lower the water table such that previously used conduits are no longer accessible.

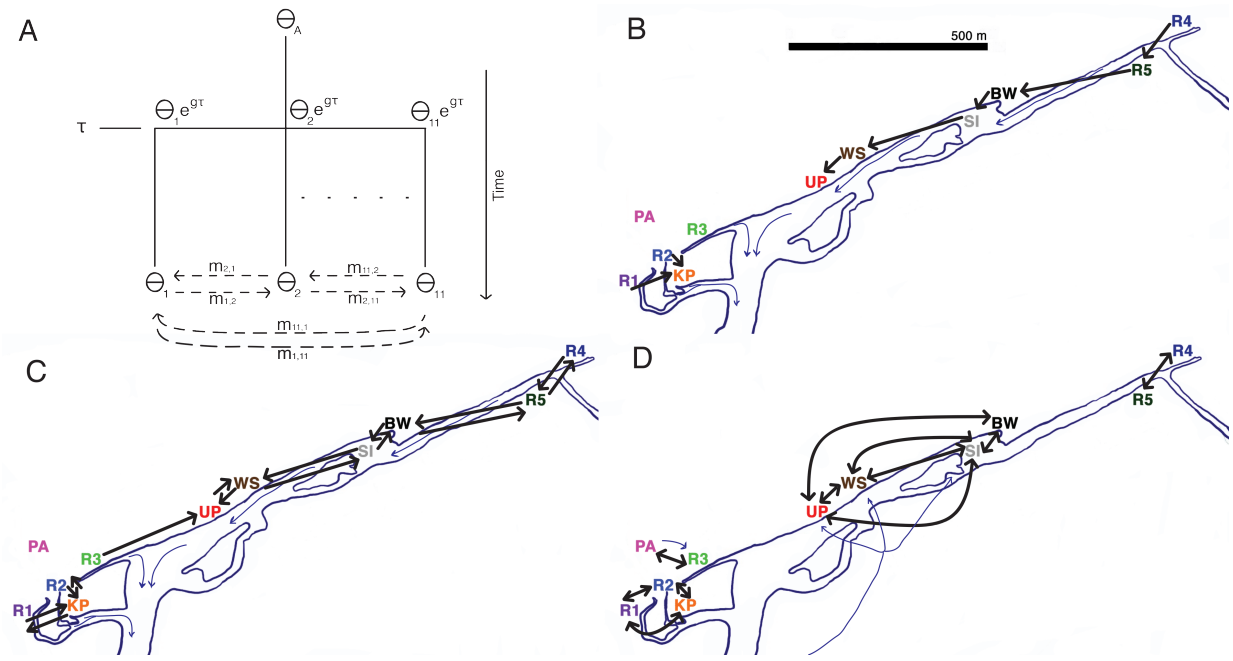

Figure 3: Model testing using ABC. Pane A is the genealogy we simulated at each locus. Panes B-D are maps of Comal Springs showing the localities of the eleven subpopulations (colored two-letter codes). We tested five models of gene flow for each taxon: no migration among subpopulations, an island model, a stepping stone with unidirectional gene flow along surface stream flow (B), a stepping stone model with bidirectional gene flow (C), and gene flow among subpopulations fed by the same groundwater sources (D). Thin blue arrows indicate water flow and thick black arrows represent gene flow.

Here we focus on the comparative population genetics of four of the several spring-endemic taxa of the Edwards Plateau whose ranges overlap in Comal Springs: Comal Springs salamander, *Eurycea* sp. (Plethodontidae: Hemidactyliini), Comal Springs riffle beetle, *Heterelmis*

comalensis (Coleoptera: Elmidae), Peck's cave amphipod, *Stygobromus pecki* (flagellatus species group, Amphipoda: Crangonyctidae) and Comal Springs dryopid beetle, *Stygoparnus comalensis* (Coleoptera: Dryopidae; Figure 2). From here on we will refer to these taxa by their generic names: *Eurycea*, *Heterelmis*, *Stygobromus* and *Stygoparnus*. The latter three of these are recognized as federally endangered species because they have restricted ranges; they are only found in one other spring complex (*Heterelmis* is in San Marcos Springs, *Stygobromus* is in Hueco Springs, and *Stygoparnus* also lives in Fern Bank Springs). Also, they depend on stable habitat conditions threatened by climate change and human use of the aquifers. Additionally, Lucas *et al.* (2009) found that the *Eurycea* in Comal Springs are likely on an independent evolutionary trajectory, and they currently have no federal or state conservation status. The three invertebrate taxa are bred captively by the U.S. Fish and Wildlife Service (USFWS) for restocking in the event that all or part of Comal Springs dries. All four taxa are generally restricted to water their entire lives, as the *Eurycea* are neotenic, *Heterelmis* larvae are aquatic and adults have vestigial hind wings and therefore cannot fly, *Stygobromus* complete their entire life cycle in water, and *Stygoparnus* larvae are thought to be terrestrial (living on the ceilings of spring orifices) and adults are aquatic (Barr & Spangler, 1992). *Eurycea* are found under rocks with minimal embeddedness at or near spring openings. However, in ephemeral springs of the Edwards Plateau *Eurycea* species are thought to lay eggs and seek refuge within subterranean habitats when aboveground conditions are unfavorable (Chippindale *et al.*, 2000; Fries, 2002; Bendik & Gluesenkamp, 2013). *Heterelmis* can be found attached to rocks, roots, and leafy and woody debris at or near springs and seeps. Later-instar larvae drift, perhaps to locate favorable habitat for pupation (C. Norris, unpubl. data). *Stygobromus* are found on rock and associated debris in or near spring sources and in one shallow well within 110 m from Comal Springs (subpopulation PA in Figure 3 B; Gibson JR, 2008). Little is known about the natural history of *Stygoparnus*, but they are found at spring orifices and subterranean habitat (specifically, subpopulation PA in Figure 3 B). These differences in affinity to particular habitats within Comal Springs across the four taxa might affect their patterns of gene flow, as might their physical sizes. *Eurycea* are an order of magnitude larger than *Heterelmis* and *Stygoparnus* adults, and about five times larger than the *Stygobromus* adults (Figure 2).

We asked two main questions. First, within Comal Springs, how do patterns of gene flow of the four endemic taxa compare to one another? We answered this question with summaries of genetic variation and structure but most importantly by explicitly testing hypotheses of patterns of gene flow using ABC. Second, is there evidence of local adaptation to the subpopulations? We answered this question by examining associations between highly differentiated single nucleotide polymorphisms (SNPs) and environmental variables (e.g., temperature). We use this information to discuss management priorities and how individuals could be pooled in captivity. The results from this comparative phylogeographic study add to the existing habitat conservation plan.

Methods

Molecular methods

There are eleven localities with suitable habitat within Comal Springs where the four focal taxa occur (i.e., subpopulations). The individuals genotyped for this project mainly were collected previously for other projects during 2005-2013, and were collected from a subset of the eleven subpopulations. We genotyped 60 *Eurycea* from three subpopulations, 70 *Heterelmis* from seven subpopulations, 68 *Stygobromus* from six subpopulations, and 53 *Stygoparnus* from four subpopulations (Table 1). All individuals were collected in accordance with USFWS (TE676811-2) and Texas Parks and Wildlife (SPR-0390-045) permits. We used DNA previously extracted from *Eurycea*, *Heterelmis* and *Stygobromus* (Lucas *et al.*, 2009; Gonzales, 2008; Ethridge *et al.*, 2013, respectively). We used the DNeasy 96 Blood and Tissue Kit (QIAGEN Sciences, Germantown, MD, USA) to extract DNA from entire *Stygoparnus* individuals.

Table 2: Number of individuals sampled from each subpopulation of four sympatric animal taxa in Comal Springs.

	PA	R1	R2	KP	R3	UP	WS	SI	BW	R5	R4
Eurycea		14			24			22			
Heterelmis		10	10	10	10		10	10	10		
Stygobromus	5	14			16	9		18		6	
Stygoparnus	9	15	8					21			

We obtained DNA sequence data from many loci to obtain accurate estimates of the evolutionary history of each taxon. Genotyping-by-sequencing is a cost-effective way to do this that does not require a reference genome and works by sequencing DNA near restriction sites. We followed the protocol described by Gompert *et al.* (2012) and Parchman *et al.* (2012) for preparing reduced genomic complexity libraries for each individual; here we briefly describe the protocol and highlight details in which our protocol differed. We first used restriction enzymes, EcoRI and MseI, to fragment individuals' genomes and thereby reduce genome complexity. We ligated Illumina sequencing adapters onto each DNA fragment and labeled the fragments of each individual with 8-base pair (bp), 9-bp, or 10-bp barcodes (individual identification sequences) to allow for multiplexing hundreds of individuals in one sequencing lane. These barcodes came from a library of 768 barcodes, each of which differs by four bases from any other sequences in the library to ensure barcode sequencing error recognition and correction (Meyer & Kircher, 2010). We amplified fragments with PCR and size selected 250-350 bp fragments with gel electrophoresis. We purified the gel excisions with QiaQuick gel extraction kits (QIAGEN Sciences, Germantown, MD, USA). The quality and concentration of libraries was assessed with a NanoDrop spectrophotometer (NanoDrop products, Wilmington, DE, USA) and quantitative electrophoresis in a Bioanalyzer (Agilent, Inc., Santa Clara, CA, USA). The National Center for Genome Research (NCGR, Santa Fe, NM) used the Illumina HiSeq platform to sequence the *Eurycea*, *Heterelmis* and *Stygobromus* libraries. After removing sequences that contained exclusively nucleotides used in library preparation

and other contaminants, we received 41.8 million filtered *Eurycea* sequences, 9.7 million filtered *Heterelmis* sequences, and 24 million filtered *Stygobromus* sequences from NCGR. These filtered sequences were 90-92 bp after barcodes were removed. The *Stygoparnus* library was prepared after modifications were made to this protocol; namely, after PCR, we added additional dNTPs and primers and ran the reaction for an additional cycle (98 °C for 3 minutes, 60 °C for 2 minutes, and 72 °C for 10 minutes) to ensure the PCR product would be dominated by double-stranded fragments. The University of Texas Genomic Sequencing and Analysis Facility (Austin, TX) sequenced the *Stygoparnus* library on the Illumina HiSeq 2500 platform, and we received 674 million filtered sequences. These filtered sequences were 86-88 bp after barcodes were removed.

We generated a set of reference sequences (i.e., a pseudo-reference genome) for each taxon because we do not have reference genomes for the four taxa with which to align our 86-92 bp sequences. We took a maximum of 15 million sequences per taxon from the millions of filtered sequences and used SeqMan NGen smng version 4.0.0.116 (DNASTAR, Inc., Madison, WI, USA) to perform a de novo assembly. The resulting assembled reference sequences were made of contigs with a minimum coverage depth of 5x, a minimum length of 80 bp, and a maximum length of 96 bp. We found 6,204 contigs in the *Eurycea* dataset, 494 contigs in the *Heterelmis* dataset, 4,980 contigs in the *Stygobromus* dataset, and 226,532 contigs in the *Stygoparnus* dataset. We performed reference-based assemblies by aligning each full set of sequences to its reference sequence using SeqMan NGen xng version 4.0.0.116 (DNASTAR, Inc., Madison, WI, USA). We used a minimum match percentage of 90% and a match size of 60 bp for both the de novo and reference-based assemblies. We removed contigs that matched more than one place in the reference sequence. The full list of parameters used in the assemblies is available from the authors by request. To identify single nucleotide polymorphisms (SNPs) in the assembled contigs and determine the number of sequences of each alternative nucleotide state for each individual and locus, we used custom Perl scripts in conjunction with samtools and bcftools (Li *et al.*, 2009). We identified the SNPs with: only two alleles to exclude paralogs, allele counts that do not violate the assumption of a binomial distribution, a low probability of the observed data if the SNPs were invariant, the posterior probability of the sequence data under a null model that the nucleotide was invariant was less than 0.01, and coverage of five or more sequences for each subpopulation sampled per taxon to ensure sufficient coverage. We have high confidence in the SNPs identified using these strict criteria, but we have likely failed to identify those SNPs with rare alleles. We found 7,035 SNPs in the *Eurycea* dataset, 545 SNPs in the *Heterelmis* dataset, 5,432 SNPs in the *Stygobromus* dataset, and 191,678 SNPs in the *Stygoparnus* dataset. The mean number of sequences per SNP per individual (i.e., coverage) was 5.78 for the identified *Eurycea* SNPs, 4.30 for the *Heterelmis* SNPs, 3.33 for the *Stygobromus* SNPs and 2.80 for the *Stygoparnus* SNPs. Due to this relatively low and variable sequence coverage across individuals, we incorporated genotype uncertainty in our genetic variation and population structure analyses instead of calling genotypes (Buerkle & Gompert, 2013).

Statistical analysis: tests for patterns of gene flow

We first described genetic variation within each taxon's subpopulations. We used a hierarchical Bayesian model that jointly estimated individuals' genotypes, subpopulation allele frequencies and genetic diversity, while accounting for genotype uncertainty in the data (Gompert *et al.*, 2012). This genetic diversity estimate is based on the distribution of allele frequencies across SNPs. If we assume drift and mutation are the only processes that affect diversity and they are constant, then allele frequencies will equilibrate to a beta distribution with a genetic diversity parameter. Conditional on our ascertainment of variable sites, genetic diversity is analogous to θ, which under these circumstances equals $4N_e\mu$. We placed a conditional beta (θ, θ) prior on the allele frequencies. We placed an uninformative uniform (0.001, 10000) hyper prior on θ. We obtained posterior parameter estimates for allele frequencies using MCMC. Each subpopulation analysis consisted of two chains iterated for 100,000 steps after a 10,000 step burn-in. We saved every 45th step. We assessed convergence and mixing with sample history plots in R (R Development Core Team, 2012).

We described subpopulation genetic structure by calculating Nei's G_{ST} (a multiallelic analogue of Wright's F_{ST}, Nei, 1973, herein called F_{ST}) with allele frequencies estimated from the previously described hierarchical Bayesian model and the equation $(H_T - H_S)/H_T$ for each SNP at each MCMC step. We then averaged F_{ST} across MCMC steps to get a point estimate for each SNP. We also took the mean of F_{ST}s across SNPs for each pair of subpopulations within taxa, which we refer to as genome-average pairwise F_{ST}s. We used an ordination method, non-metric multidimensional scaling (NMDS), to visualize subpopulation differentiation using genome-average pairwise F_{ST}s. NMDS does not force birfurcating relationships among subpopulations. We used the isoMDS function in the MASS package in R to conduct Kruskal's NMDS with the *Eurycea* data with one dimension, the *Heterelmis* data with three dimensions, the *Stygobromus* data with two dimensions and the *Stygoparnus* data with three dimensions. We also used the Mantel.rtest function in the ade4 package in R to conduct Mantel tests with genome-average pairwise F_{ST}s and straight-line geographic distances between pairs of subpopulations to test the significance of the association between F_{ST} and distance (i.e., isolation by distance or IBD) in each dataset. Distance matrices were based on Euclidean distances. Each test was based on 9999 randomizations.

We used approximate Bayesian computation (ABC) to test which models (hypotheses) of gene flow best explained patterns of genetic variation in the data. We tested five competing models: 1) no gene flow among subpopulations, 2) an island model with equal or constant gene flow among all subpopulations, 3) a stepping stone model with unidirectional gene flow along surface stream flow (Figure 3 B), 4) a stepping stone model with bidirectional gene flow with and against surface stream flow (Figure 3 C), and 5) gene flow among subpopulations fed by the same groundwater sources (Figure 3 D). We developed the two surface stream models (models 3 and 4) based on current stream flow paths. We developed the groundwater model (model 5) based on our current understanding of groundwater flow at Comal Springs based on dye trace studies (S. Johnson and G. Schindel, unpubl. data). Water feeding Comal Springs comes from two major sources. The spring runs, R1, R2 and R3, are from one flow path. Most of Landa Lake is from another, presumably deeper, source, as it has consistently higher temperature by 0.5 °C. We cannot calculate the likelihood of each hypothesized

gene flow model (that is, the probability of obtaining the data given the model and specific parameter values), so we approximate the model likelihood using ABC (Nielsen & Beaumont, 2009). Specifically, we simulated data given parameter values taken from prior distributions. We calculated summary statistics on the simulated data and the observed data. We then calculated the probability of each hypothesized gene flow model by finding the number of simulated datasets yielding summary statistics that lie within a small distance of the summary statistic computed from the observed data.

First, we specified the data for ABC. We identified our 86-92 bp loci that were infinite sites compatible (i.e., every mutation occurs at a unique nucleotide, verified by the four gamete test within a contig). We assumed there is no recombination within a locus but free recombination between loci. Because it is time intensive to simulate enough demographic histories and calculate the corresponding summary statistics to find enough simulations similar to the true demographic histories, we did not use our full locus datasets for ABC; instead, we used a subset of our full datasets to include one locus per individual for higher coverage loci. Specifically, in the *Eurycea* dataset, we identified 475 variable loci with data for ten or more individuals in each sampled subpopulation. We used the 174 variable loci with data for five or more *Heterelmis* individuals in each sampled subpopulation, and the 496 variable loci with data for five or more *Stygobromus* individuals in each sampled subpopulation. There were 61,180 variable loci with data for five or more individuals per subpopulation in the *Stygoparnus* dataset, which was too many to run ABC practically, so we randomly sampled 500 of the 61,180 loci using R and a custom Perl script.

Second, we used a custom Perl script and the software ms (Hudson, 2002) to simulate demographic histories (Figure 3 A) and calculate the corresponding summary statistics. We simulated the genealogy at each locus, where subpopulation θs were a fraction of the ancestral θ after their simultaneous split from the common ancestor. After splitting, subpopulations were allowed to grow or decline (+/- g) to reach a new θ. Subpopulations diverged with or without migration among subpopulations as dictated by the migration model (Figure 3 A). We placed priors on the raw parameters based on the available information we have about the taxa and the history of Comal Springs. For example, we allowed subpopulation growth or decline because it is a realistic way to represent the effect spring flow variability and habitat modification may have on subpopulation sizes. We drew the following parameters from uninformative prior distributions (prior distributions were the same for each taxon): 1) we placed a prior of 1/5 on each of the five gene flow models; 2) effective population size, N_e, was drawn from a log uniform distribution between 50 and 10,000; 3) migration rate (m), the fraction of each subpopulation made up of new migrants each generation, was drawn from a uniform distribution between 0.00005 and (1-(1/number of subpopulations)); 4) time since divergence, τ, was drawn from a log uniform distribution between 10 and 50,000 generations; 5) mutation rate per fragment, μ, was drawn from a uniform distribution between 1×10^{-7} to 8×10^{-6}; and 6) growth rate, g, was drawn from a uniform distribution between -2 and 2. We also estimated mean θ, ancestral θ, subpopulation θ, and the number of migrants per generation per subpopulation ($4N_e^{anc}m$). We performed a large number of time-intensive simulations, roughly one million per taxon (1,234,020 *Eurycea* datasets, 1,072,002 *Heterelmis* datasets, 1,280,398 *Stygobromus* datasets, and 1,164,000 *Stygoparnus* datasets), to ensure the observed summary statistics were similar to a large number of

simulated summary statistics. We simulated data for eleven subpopulations for each taxon to include all potential subpopulations in Comal Springs (Figure 3 B-D). We used the same sample sizes of the observed datasets.

We then calculated the mean, variance, and skew of five locus-specific summary statistics that describe genetic diversity within each subpopulation: expected heterozygosity (2pq), the average number of nucleotide differences between pairs of loci in the sample (π, Tajima, 1983), the number of segregating or polymorphic sites within a locus (S, Watterson, 1975), the number of private haplotypes (unique haplotypes in one subpopulation and no other), and the proportion of loci in which the rarer allele has a frequency less than 0.1 (low allele frequency). We chose these statistics because they capture different aspects of the information in the data about the genealogical history of the samples. For example, S counts each mutation once, whereas π weights sites depending on the frequency of the mutation as well (Wakeley, 2009). Importantly, our chosen statistics may also be informative of different models of migration. For example, we would expect small variance in S in a model of no gene flow and large variance in S in a model of subdivision with gene flow (Wakeley, 2009). Similarly, we would expect an excess of low frequency haplotypes in a growing subpopulation and an excess of moderate frequency haplotypes in a declining subpopulation (Wakeley, 2009). We also calculated the mean, variance, and skew of π and F_{ST} (Nei, 1973) for all pairs of subpopulations for which we had data for each taxon (three pairs of subpopulations in the *Eurycea* dataset, 21 *Heterelmis* pairs of subpopulations, 15 *Stygobromus* subpopulation pairs and six subpopulation pairs in the *Stygoparnus* dataset).

A key to successful application of these ABC methods is how well the summary statistics capture the relevant properties of the data (**?**). After running approximately 20% of the total number of simulations for each dataset, we ran diagnostic tests to ensure: 1) parameters were correlated with summary statistics and 2) summary statistics were not redundant. We used the cor function in R to estimate these correlations. We also made sure observed summary statistics fell within the distribution of simulated summary statistics. We used the hist function in R to place observed summary statistics on the distribution of summary statistics from simulated data.

Last, we based our inference on the 5000 simulations that gave summary statistics most similar to the observed summary statistics. We then performed generalized linear regressions with multinomial error functions to estimate posterior probabilities for each gene flow model for each taxon. We performed local linear regression and model averaging to estimate our parameter of interest, migration rate (m), while integrating over uncertainty in our other parameters (e.g., g, τ). We used the functions abc and postpr in the abc package in R (Csilléry *et al.*, 2010).

Statistical analysis: tests for local adaptation

To look for evidence of local adaptation within Comal Springs, we: 1) chose highly differentiated SNPs based on F_{ST}, 2) examined whether these SNPs exhibit patterns of IBD to control for IBD when testing for local adaptation, and 3) tested for correlations between genetic and environmental differences for these SNPs, which would support the hypothesis

of local adaptation. We identified SNPs with F_{ST} greater than 0.3 in at least one subpopulation pair in the *Eurycea* dataset, F_{ST} greater than 0.3 in at least one subpopulation pair in the *Heterelmis* dataset, and F_{ST} greater than 0.33 in at least one subpopulation pair in the *Stygobromus* dataset. We identified 19, 24, and 24 SNPs, respectively. For the *Sytgoparnus* dataset, we identified SNPs with a mean F_{ST} greater than 0.4 across all pairwise F_{ST}s in order to make the number of high F_{ST} SNPs comparable to the other datasets; we identified 21 SNPs. We first conducted Mantel tests with pairwise straight-line geographic distance and pairwise F_{ST} of each one of these highly differentiated SNPs to test for IBD.

We had the following environmental data available to us recorded from spring openings in most of the subpopulations in our datasets (C. Norris et al., unpub. data): elevation, maximum water depth, temperature, pH, dissolved oxygen (DO), specific conductivity, total dissolved solids (TDS) and primary and secondary substrate size (based on the Wentworth scale in which substrate codes are higher for larger substrates; Wentworth, 1922). Some of the variables are invariant across certain subpopulations. We took the median of each variable for each subpopulation. We realize variables measured at one time point are not representative of a dynamic spring system, but long-term environmental data, for a limited number of environmental variables, was only available for springs at three subpopulations. We conducted a principal component analysis (PCA) using subpopulation medians of the environmental variables to distill down the number of variables and visualize overall environmental similarities among subpopulations. Then, for each taxon, we performed a PCA with the medians of the environmental variables from the relevant subpopulations for the respective taxon. We used the prcomp function in R to perform the PCAs. We then performed partial Mantel tests to explore the association between PC scores and pairwise F_{ST} of highly differentiated SNPs while controlling for geographic distance for each dataset. We used the R package ecodist to perform each partial Mantel test. Distance matrices were based on Euclidean distances. A significant relationship between differentiation for these SNPs and the potential environmental correlates would be consistent with the hypothesis that those variants (or linked variants) are involved in local adaptation.

Results

We made comparisons of θ, an estimate of genetic diversity based on the allele frequency distribution, across subpopulations within each of the four taxa (Figure 4). *Eurycea* subpopulation θs ranged from 0.26-0.29. Subpopulation θs ranged from 0.26-0.35 and 0.33-0.45 for *Heterelmis* and *Stygobromus*, respectively. *Stygoparnus* subpopulation θs had the widest range, from 0.59-0.75. Subpopulation R1 had a lower θ for both *Stygobromus* and *Stygoparnus*. However, in general, θ was similar across subpopulations within each taxon, suggesting evolutionary processes affecting diversity (including population size and genetic drift) are similar to one another within each taxon, perhaps because all subpopulations should be thought of as one population. It is not appropriate to compare θs among taxa in this case, because the sequence coverage varied among datasets which affected the ascertainment of SNPs.

Most SNPs offered little evidence of subpopulation structure, but there were a few SNPs with higher pairwise F_{ST}s, particularly in the *Heterelmis* dataset (Figure S1 in Supporting

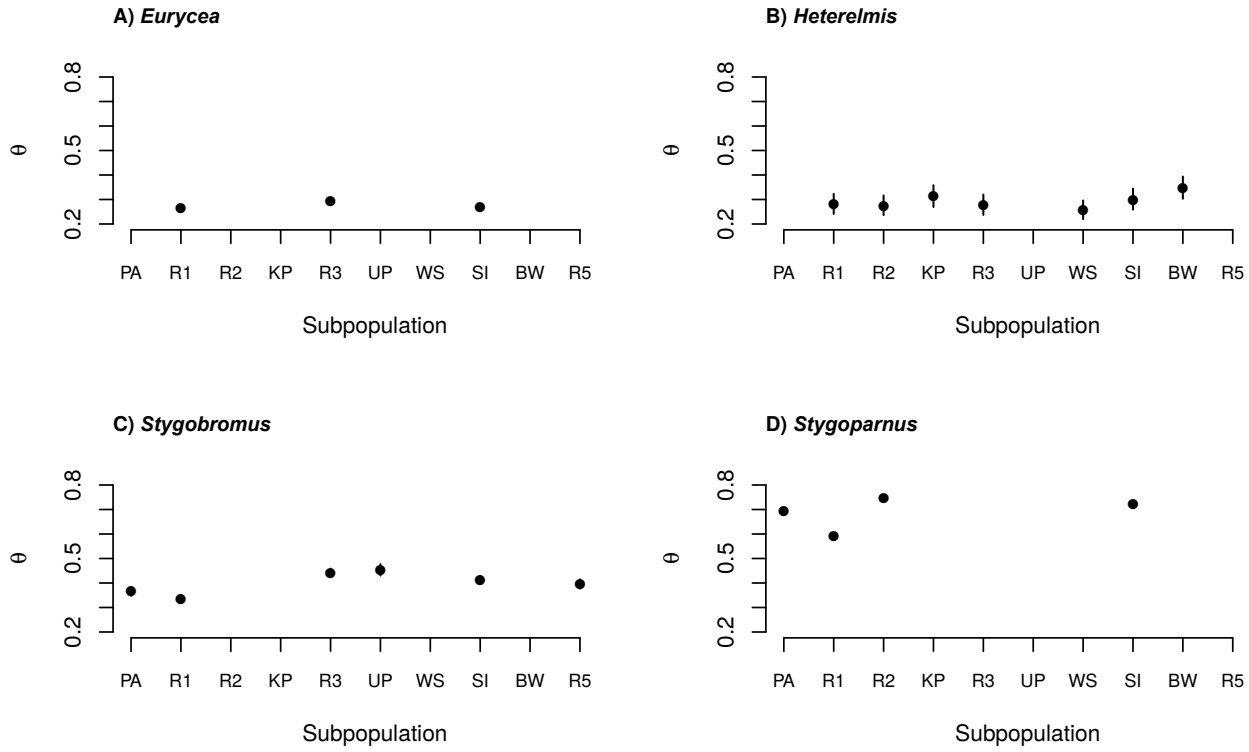

Figure 4: θ, an estimate of genetic diversity based on the allele frequency distribution, across subpopulations within each taxon. Dots are point estimates and lines are 95% credible intervals.

Information). Some of these SNPs with higher F_{ST}s were associated with environmental variables (see below for details). Genome-average pairwise F_{ST}s ranged from 0.047-0.054 between *Eurycea* subpopulations, 0.045-0.061 between *Heterelmis* subpopulations, 0.036-0.077 between *Stygobromus* subpopulations and 0.064-0.077 between *Stygoparnus* subpopulations (Table S1). Patterns of differentiation were different among the four taxa (Figure 5). *Eurycea* subpopulations were nearly equally differentiated. *Heterelmis* subpopulations displayed a correspondence between genetic diversity and geographic space. *Stygobromus* and *Stygoparnus* subpopulations were structured but not geographically. There was no association between genetic differentiation and distance in three of the taxa but the pattern is marginally significant for *Heterlemis* (Mantel test p-value for *Eurycea*: 0.4974, *Heterelmis*: 0.0965, *Stygobromus*: 0.4588, *Stygoparnus*: 0.3754).

We tested demographic models for each taxon using ABC. Subpopulation growth was the only parameter that did not correlate with at least one of the summary statistics across all datasets (Figure S2). In all datasets, summary statistics were correlated with one another to various degrees (Figure S3). In all datasets, all observed summary statistics fell within the distribution of the simulated summary statistics. Thus, we felt confident about the ability of our chosen ABC summary statistics to capture the relevant properties of the data.

The island model with equal gene flow among subpopulations (model 2) had the highest posterior probability for all four taxa: 100% for *Eurycea*, 88% for *Heterelmis*, 100% for *Stygobromus* and 59% for *Stygoparnus* (Table 2). Migration rate (m) parameters had relatively wide posterior probability distributions, with the exception of m for *Stygoparnus*; however, all posterior probability distributions were different than the uniform prior distributions (Figure 6). On average, 0.549 of *Eurycea* subpopulations were made up of new migrants each generation (95% credible interval (CI): 0.023-0.902); 0.631 of *Heterelmis* subpopulations were made up of new migrants each generation (CI: 0.205-0.877); and m was 0.343 (CI: 0.025-0.825) and 0.152 (CI: 0-0.768) for *Stygobromus* and *Stygoparnus*, respectively (Figure 6).

Table 3: ABC posterior probabilities for each hypothesized gene flow model for each taxon.

Taxon	Model 1: None	Model 2: Equal	Model 3: Unidirectional	Model 4: Bidirectional	Model 5: Groundwater
Eurycea	0.0002	0.9968	0.0001	0.0005	0.0023
Heterelmis	0.0070	0.8788	0.0235	0.0046	0.0862
Stygobromus	0.0003	0.9962	0.0000	0.0000	0.0035
Stygoparnus	0.1330	0.5940	0.1521	0.1105	0.0105

Some environmental conditions were relatively similar across subpopulations, like pH (range 7-7.2), and others were more variable, such as specific conductivity (range 406-500 μS/cm, Table 3). Based on the PCA including environmental data from all subpopulations, PC 1 explained 46.8% of the variation and represented a positive, strong relationship among temperature, specific conductivity and TDS; DO and substrate size were strongly negative associated with other variables (Table S2). We found roughly the same relationship among variables when conducting PCAs for each taxon to look for evidence of local adaptation

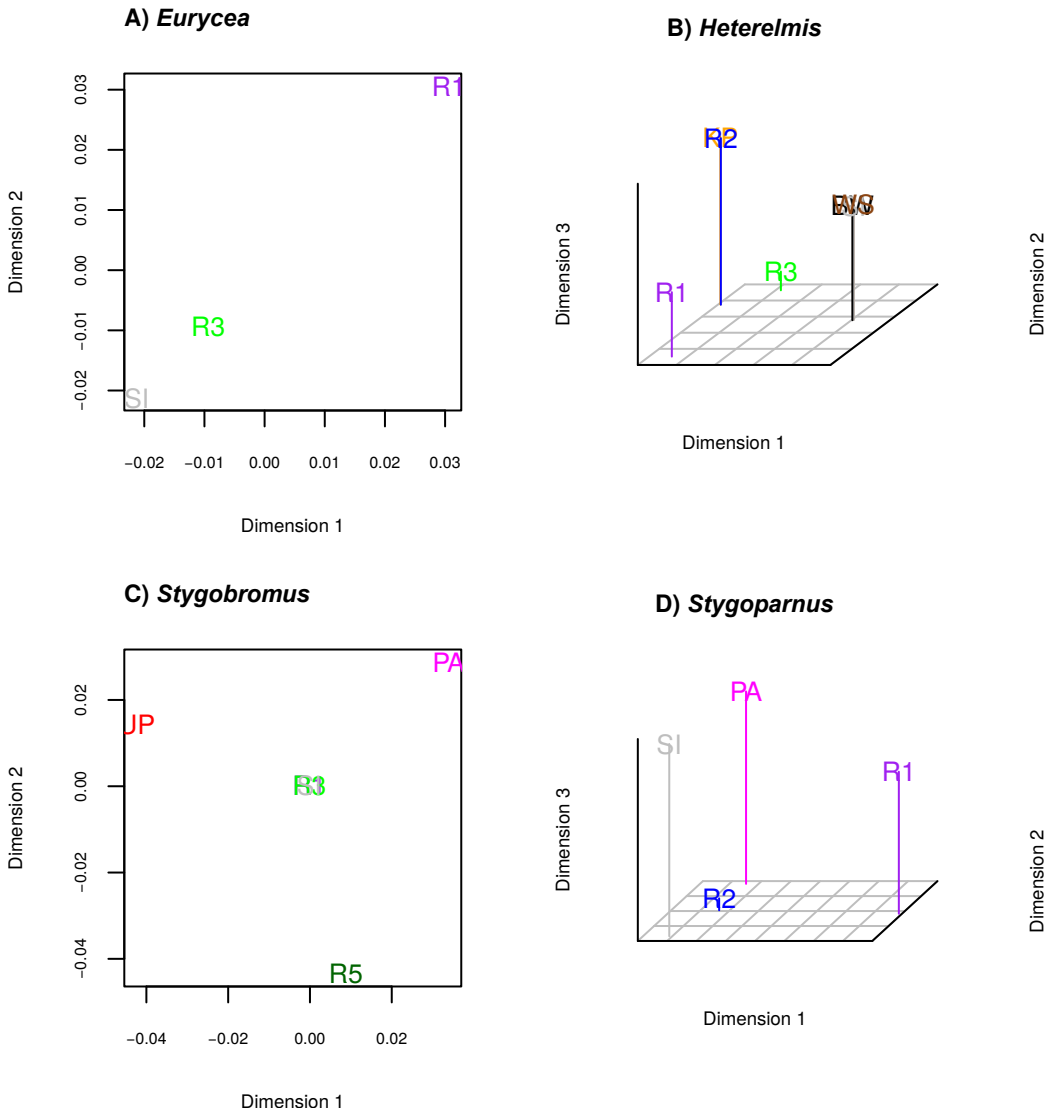

Figure 5: Non-metric multidimensional scaling (NMDS) using genome-average pairwise F_{ST}s shows subpopulation differentiation within each taxon. Note in pane B: R2 and KP overlap as do BW, SI, WS. Note in pane C: R1, R3, SI overlap.

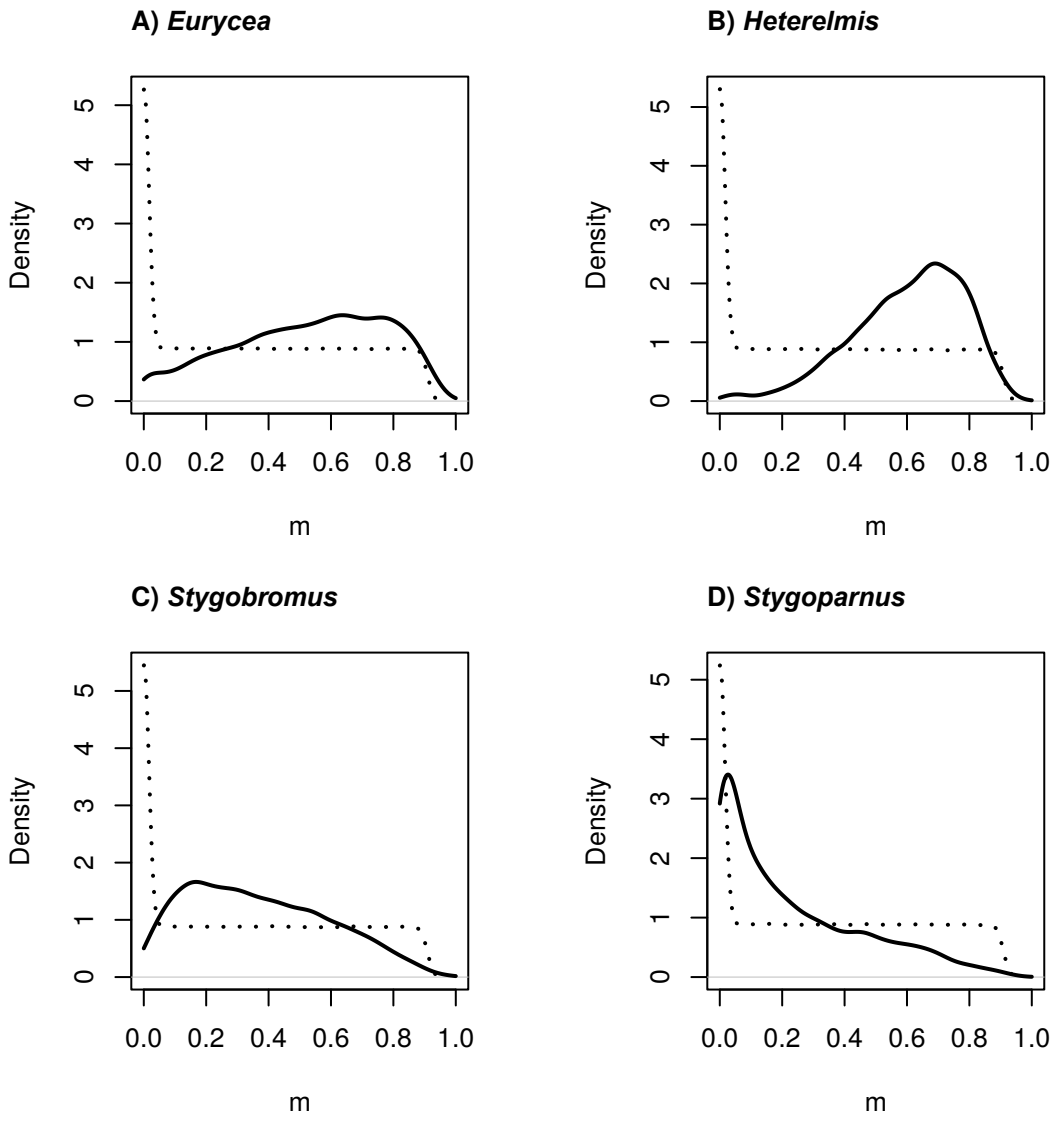

Figure 6: Model-averaged ABC prior (dotted line) and posterior (solid line) probability distributions for migration rate (m) for each taxon.

(Table S3 B). Environmental differences were somewhat structured by geography, according to PC 1. Subpopulations R1, R2, R3, WS, UP and KP had similar environmental conditions to each other, as did the group: SI, R4, R5. The environment at subpopulation BW was different from all other subpopulations (Figure S4).

Table 4: Median values for each environmental variable for each subpopulation. Primary and secondary substrate size measurements are based on the Wentworth scale.

Subpop.	Elevation (m)	Max. depth (m)	Temp. (°C)	pH	DO (mg/L)	Sp. Cond. (μS/cm)	TDS	1° sub.	2° sub.
BW	189.2	0.06	22.83	7.0	5.4	413.7	0.27	10	11
KP	189.1	0.12	23.39	7.1	4.6	408.0	0.26	5	7
PA	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
R1	189.6	0.34	23.28	7.1	5.2	408.7	0.26	9.5	9
R2	189.6	0.18	23.25	7.1	5.1	406.6	0.26	8	8.5
R3	189.3	0.09	23.22	7.1	5.2	424.3	0.27	10	8.5
R4	N/A	0.04	23.69	7.1	5.0	495.9	0.32	5.5	9.5
R5	N/A	0.03	23.67	7.1	4.7	497.5	0.32	9	4
SI	188.7	0.52	23.50	7.1	4.9	500.1	0.32	6	6
UP	189.0	0.23	23.61	7.1	5.1	412.2	0.26	8	7
WS	189.2	0.05	23.61	7.2	5.1	406.4	0.26	10	8

After performing the Mantel tests and partial Mantel tests, we asked if the number of statistically significant correlations between highly differentiated SNPs and geographic distance or environmental PC scores, respectively, within a taxon was more than we would expect with a 5% cutoff rate. None of the 19 highly differentiated SNPs in the *Eurycea* dataset were significantly associated with geographic distance. We attribute this, in part, to so few subpopulation comparisons (three only). Four of the 24 highly differentiated SNPs in the *Heterelmis* dataset were significantly associated with geographic distance, which is more than we would expect by chance. One of the 24 highly differentiated SNPs in the *Stygobromus* dataset was significantly associated with geographic distance, which is roughly the number of significant correlations expected by chance. Three of the 21 highly differentiated SNPs in the *Stygoparnus* dataset were significantly associated with geographic distance, again, which is more than expected by chance. However, none of these p-values were significant following false discovery rate (FDR) correction. See Table S3 A for a list of the significant highly differentiated SNPs and p-values.

To explore evidence of local adaptation in each dataset, we performed partial Mantel tests using pairwise F_{ST} of highly differentiated SNPs and PC scores based on the environmental variables collected from the subpopulations relevant to each dataset, while controlling for geographic distance. One SNP of the 19 in the *Eurycea* dataset was associated with the environment (specifically, PC 1 scores). This is the number of signficiant associations we would expect by chance; however, the p-value was significant following FDR correction. Three SNPs in the *Heterelemis* dataset were significantly associated with PC 1; however, none of the three p-values were significant following FDR correction. Seven of the 24 SNPs in the *Heterelmis* dataset were associated with PC 2. Again, none of these p-values were significant following FDR correction. One SNP of the 24 in the *Stygobromus* dataset was associated with PC 1, which is expected by chance, but this p-value was significant following

FDR correction. Four of the 24 SNPs were associated with PC 2, but none of these p-values were significant following FDR correction. Last, two of the 21 SNPs in the *Stygoparnus* dataset were associated with the environment (PC 1), and both p-values were significant following FDR correction. See Table S3 B for a list of the significant highly differentiated SNPs, p-values and the proportion of variance explained by the PCs mentioned above.

Discussion

Gene flow and local adaptation

The gene flow model with the most support for all four taxa was the island model in which there is equal gene flow among all subpopulations. These ABC results were consistent with the descriptive patterns of genetic variation and structure: similar levels of genetic diversity (θ) among subpopulations, low genome-average pairwise F_{ST}s, and the lack of statistically significant associations between genetic differentiation and distance. Each of the four taxa had high, but potentially different migration rates (posterior distributions for m were wide, but the posterior means were different), ranging from 15 to 55% (posterior means) of subpopulations made up of new migrants per generation. The stygobionts, *Stygobromus* and *Stygoparnus*, had relatively lower migration rate estimates, perhaps because their habitats are inherently more isolated. Though, these levels of gene flow are enough to prevent subpopulation isolation within Comal Springs for each taxon (i.e., more than one migrant per generation, Wright, 1931; Slatkin, 1985). All four taxa did not seem to be constrained by the direction of water flow or our conception of their dispersal abilities.

Given the data in this paper reflect a long evolutionary history, alternative explanations for all taxa fitting the island model and high m estimates include the fact that Comal Springs previously was a continuous spring-fed marsh, perhaps making gene flow easier (Lande, 1999). Comal Springs was a continuous spring-fed marsh up until the spring water was impounded in 1847 and channelized in 1936, becoming a heavily-used city park. Furthermore, whereas the public is advised to stay out of the springs that support the endangered taxa, the taxa may rarely experience human-mediated gene flow (e.g., throwing rocks, children using aquarium nets).

We found at least one SNP in each dataset that was associated with aspects of the Comal Springs environment, after asking if the number of significant correlations between pairwise F_{ST} of highly differentiated SNPs and environment was more than we would expect due to chance (more than 5% of the time) or after FDR correction. These associations were consistent with the hypothesis that these SNPs reflect local adaptation. It is interesting that we found any associations given the relatively similar environmental conditions as well as the high migration rate estimates (Holt & Gomulkiewicz, 1997). However, it is important to note that these SNPs may not be directly under selection. That is, if a SNP contributes to local adaptation, meaning it is under different selection in different environments, we would find a correlation between allele frequency and the environment. But not every SNP whose allele frequency is correlated with the environment is directly under selection; the correlation could be caused by drift or by a SNP linked to the genetic region under selection instead (e.g.,

Haldane, 1948; Slatkin, 1973; Coop *et al.*, 2010). For the SNPs that did not show a relationship between genetic differentiation and environmental correlates, either local adaptation is not the explanation for the differentiation observed at these SNPs, or we may have not yet identified the relevant environmental variables (i.e., aspects of the subsurface environment). We may have found more associations between highly differentiated SNPs and environment in general with more genetic markers and specifically if we had more subpopulations represented in the *Eurycea* and *Stygoparnus* datasets. Thus, we do not necessarily have strong evidence of local adaptation, but we should take seriously the potential for local adaptation, and further investigation is warranted. A logical next step would be to perform reciprocal transplant experiments or performance assays (Kawecki & Ebert, 2004).

Conservation management

Moritz (1999) defined a management unit (MU) as demographically independent where growth rate depends on local birth and death rates rather than on immigration. Whereas an evolutionary significant unit (ESU) shows long-term independent evolution or strong adaptive differentiation. Maintenance of MUs are important for the long-term persistence of an ESU. Previous phylogeographic studies examined some of the four taxa of this study at various geographic scales, albeit with genetic markers with comparatively lower resolution. Lucas *et al.* (2009) found IBD across populations of *Eurycea* to suggest each *Eurycea* population per spring complex is an ESU. The work of Ethridge *et al.* (2013) and Ethridge *et al.* (2013) suggested there was one MU in the Comal Springs *Stygobromus* population and more than one MU for *Heterelmis*. However, as is the case with the latter studies, allele frequency differentiation (F_{ST}) should not be used by itself to identify MUs because the same F_{ST} can result in different migration rates for different population sizes (Allendorf & Luikart, 2009). Here we use both F_{ST} and patterns of gene flow and now know there is not much genetic structure within Comal Springs, considerable gene flow among subpopulations and evidence that is consistent with the hypothesis that there is some local adaptation to subpopulations in Comal Springs. Thus, we suggest considering the entire Comal Springs complex as both the MU and the ESU for all four spring-endemic taxa.

In 2012, the USFWS approved a habitat conservation plan for managing the Edwards Aquifer to preserve the federally-listed species at Comal Springs as well as the other major spring complex in Texas, San Marcos Springs (for details, see: http://www.eahcp.org/). The plan includes recommendations for how much water will be available in these spring systems in periods of drought. The plan adds a fifth stage to the existing critical period management, which describes well withdrawal reduction measures to be taken if the aquifer level drops below 190.5 mamsl. This water level is just slightly above the level at which Comal Springs would dry, particularly the spring runs. If part of the spring system temporarily dries, any localized extinctions may be naturally recolonized from elsewhere, based on our results from model testing with ABC. However, based on our tests for local adaptation, genetic diversity at SNPs potentially important for surviving in particular subpopulations of Comal Springs could be lost in such situations.

While the habitat conservation plan assures that Comal Springs will sustain suitable habitat no matter the threats to the aquifer any given year, there is still the potential loss

of water quantity and the possibility of catastrophic water quality issues. Because of this possibility there are captive breeding programs for three of the taxa in this study and others. Space at the captive breeding facility is limited. As such, conservation managers have to make tough decisions about how wild-caught individuals should be structured in captivity. Based on our evidence that adaptive variation could be partitioned among subpopulations, if possible, we suggest individuals from subpopulations could be kept in separate tanks to maintain those alleles important for local adaptation. Alternatively, the NMDS plots (Figure 5) could be used to decide which subpopulations should be grouped together in captivity if space is limited. For example, *Heterelmis* from R2 and KP could be pooled in captivity, as could BW, SI and WS, while R1 and R3 could be kept separate. However, if space does not allow for these designs, all individuals collected from the wild could be pooled in a tank. After all, wild gene flow estimates are high and pooling individuals like this would increase random mating and thereby total genetic variation (Hartl *et al.*, 1997). We hope these results will help maintain the genetic variation of the taxa that rely on the stable conditions at Comal Springs.

Acknowledgements

This research was funded by Texas Parks and Wildlife (Section 6) to LKL and CCN. We thank Chad Norris and colleagues for detailed environmental data and the USFWS Aquatic Resources Center staff, especially Joe Fries and Val Cantu, for their help collecting specimens. To run the ABC simulations, we used computer clusters at the University of Wyoming, Texas State University, Utah State University as well as the Extreme Science and Engineering Discovery Environment (XSEDE, previously Teragrid, through allocation award MCB110082 to CAB), which is supported by National Science Foundation grant number ACI-1053575. This project was improved by input from Bob Hall, Jim Ott, Kenneth Ostrand, Tom Devitt, and Andy Gluesenkamp.

Part 1: Supporting Information

Table S1. Genome-average pairwise F_{ST}s for each taxon.
Table S2. Relationships among environmental variables recorded at Comal Springs.
Table S3. Mantel and partial Mantel results.
Figure S1. Distribution of F_{ST}s across SNPs for each pair of subpopulations per taxon.
Figure S2. Patterns of correlations of ABC parameters with ABC summary statistics for each taxon.
Figure S3. Patterns of correlations between all pairs of ABC summary statistics for each taxon.
Figure S4. Environmental similarities among subpopulations based on PCA.

Part1: Supporting Information

Table S1: Genome-average pairwise F_{ST}s for each taxon. Averages are above the diagonal and 95% credible intervals are below the diagonal.

A) *Eurycea*

	R1	R3	SI
R1	0	0.051	0.054
R3	0.049-0.052	0	0.047
SI	0.052-0.055	0.046-0.048	0

B) *Heterelmis*

	BW	KP	R1	R2	R3	SI	WS
BW	0	0.056	0.059	0.056	0.057	0.045	0.049
KP	0.051-0.061	0	0.058	0.054	0.057	0.061	0.060
R1	0.054-0.065	0.053-0.064	0	0.055	0.060	0.061	0.061
R2	0.051-0.061	0.048-0.059	0.049-0.061	0	0.056	0.056	0.055
R3	0.051-0.062	0.051-0.063	0.054-0.066	0.050-0.062	0	0.060	0.059
SI	0.041-0.050	0.055-0.067	0.055-0.067	0.051-0.062	0.053-0.065	0	0.051
WS	0.044-0.054	0.054-0.066	0.055-0.067	0.050-0.061	0.053-0.0657	0.046-0.056	0

C) *Stygobromus*

	PA	R1	R3	R5	SI	UP
PA	0	0.066	0.056	0.071	0.059	0.077
R1	0.064-0.068	0	0.044	0.062	0.047	0.069
R3	0.054-0.058	0.043-0.046	0	0.051	0.036	0.058
R5	0.069-0.074	0.060-0.064	0.050-0.053	0	0.054	0.073
SI	0.057-0.061	0.045-0.048	0.035-0.037	0.053-0.056	0	0.061
UP	0.074-0.079	0.066-0.071	0.056-0.060	0.071-0.076	0.059-0.063	0

D) *Stygoparnus*

	PA	R1	R2	SI
PA	0	0.076	0.064	0.070
R1	0.0759-0.077	0	0.074	0.078
R2	0.064-0.065	0.073-0.074	0	0.064
SI	0.070-0.071	0.077-0.078	0.064-0.065	0

Table S2: PC 1 and PC 2 loadings explaining the relationship among environmental variables recorded at Comal Springs.

Environmental variable	PC 1 (46.8%)	PC 2 (19.8%)
Max. depth (m)	0.08740491	0.1947870
Temperature (°C)	0.42551118	-0.3304372
pH	0.18653963	-0.6655191
DO (mg/L)	-0.43158072	0.1274411
Specific conductivity (μS/cm)	0.40942043	0.3873596
TDS	0.39021054	0.4168354
Primary substrate	-0.32804406	-0.1808025
Secondary substrate	-0.40340973	0.1929342

Table S3: Mantel (A) and partial Mantel (B) results. A) Highly differentiated SNPs associated with geographic distance and p-values. B) Results of the partial Mantel tests in which there are relationships between pairwise F_{ST} of highly differentiated SNPs and environmental PC scores, controlling for geographic distance. The significant SNPs, p-values, proportion of variance explained by the principal component (PC) and PC loadings for each environmental variable are included.

A)

Taxon	Locus	p-value
Heterelmis	74	0.0032
Heterelmis	273	0.04731
Heterelmis	414	0.0065
Heterelmis	538	0.0214
Stygobromus	130	0.0271
Stygoparnus	42669	0.0439
Stygoparnus	89870	0.0408
Stygoparnus	144043	0.0453

B)

Taxon	Locus	p-value	PC (% var.)	Elevation	Max. depth	T	pH	DO	Sp. Cond.	TDS	1 ° substrate	1 ° substrate
Eurycea	2833	0.0001	1 (92.5%)	-0.343	0.297	0.361		-0.368	0.363	0.363	-0.365	-0.363
Heterelmis	39	0.0120	1 (53.2%)	0.332	-0.318	-0.313	-0.144	0.327	-0.369	-0.349	0.350	0.428
Heterelmis	252	0.0240	1 (53.2%)	0.332	-0.318	-0.313	-0.144	0.327	-0.369	-0.349	0.350	0.428
Heterelmis	471	0.0163	1 (53.2%)	0.332	-0.318	-0.313	-0.144	0.327	-0.369	-0.349	0.350	0.428
Heterelmis	39	0.0391	2 (25.4%)	0.229	-0.240	0.425	0.533	-0.306	-0.354	-0.397	-0.058	-0.214
Heterelmis	57	0.0071	2 (25.4%)	0.229	-0.240	0.425	0.533	-0.306	-0.354	-0.397	-0.058	-0.214
Heterelmis	224	0.0539	2 (25.4%)	0.229	-0.240	0.425	0.533	-0.306	-0.354	-0.397	-0.058	-0.214
Heterelmis	247	0.0135	2 (25.4%)	0.229	-0.240	0.425	0.533	-0.306	-0.354	-0.397	-0.058	-0.214
Heterelmis	273	0.0045	2 (25.4%)	0.229	-0.240	0.425	0.533	-0.306	-0.354	-0.397	-0.058	-0.214
Heterelmis	472	0.0257	2 (25.4%)	0.229	-0.240	0.425	0.533	-0.306	-0.354	-0.397	-0.058	-0.214
Heterelmis	538	0.0032	2 (25.4%)	0.229	-0.240	0.425	0.533	-0.306	-0.354	-0.397	-0.058	-0.214
Stygobromus	991	0.0298	1 (65.9%)		0.028	0.370		-0.450	0.437	0.434	-0.287	-0.445
Stygobromus	64	0.0459	2 (23.9%)		-0.762	0.125		-0.174	-0.044	-0.030	0.572	-0.210
Stygobromus	862	0.0190	2 (23.9%)		-0.762	0.125		-0.174	-0.044	-0.030	0.572	-0.210
Stygobromus	975	0.0173	2 (23.9%)		-0.762	0.125		-0.174	-0.044	-0.030	0.572	-0.210
Stygobromus	3176	0.0157	2 (23.9%)		-0.762	0.125		-0.174	-0.044	-0.030	0.572	-0.210
Stygoparnus	13330	0.0001	1 (91.7%)	-0.368	0.312	0.364		-0.331	0.368	0.368	-0.343	-0.368
Stygoparnus	97174	0.0010	1 (91.7%)	-0.368	0.312	0.364		-0.331	0.368	0.368	-0.343	-0.368

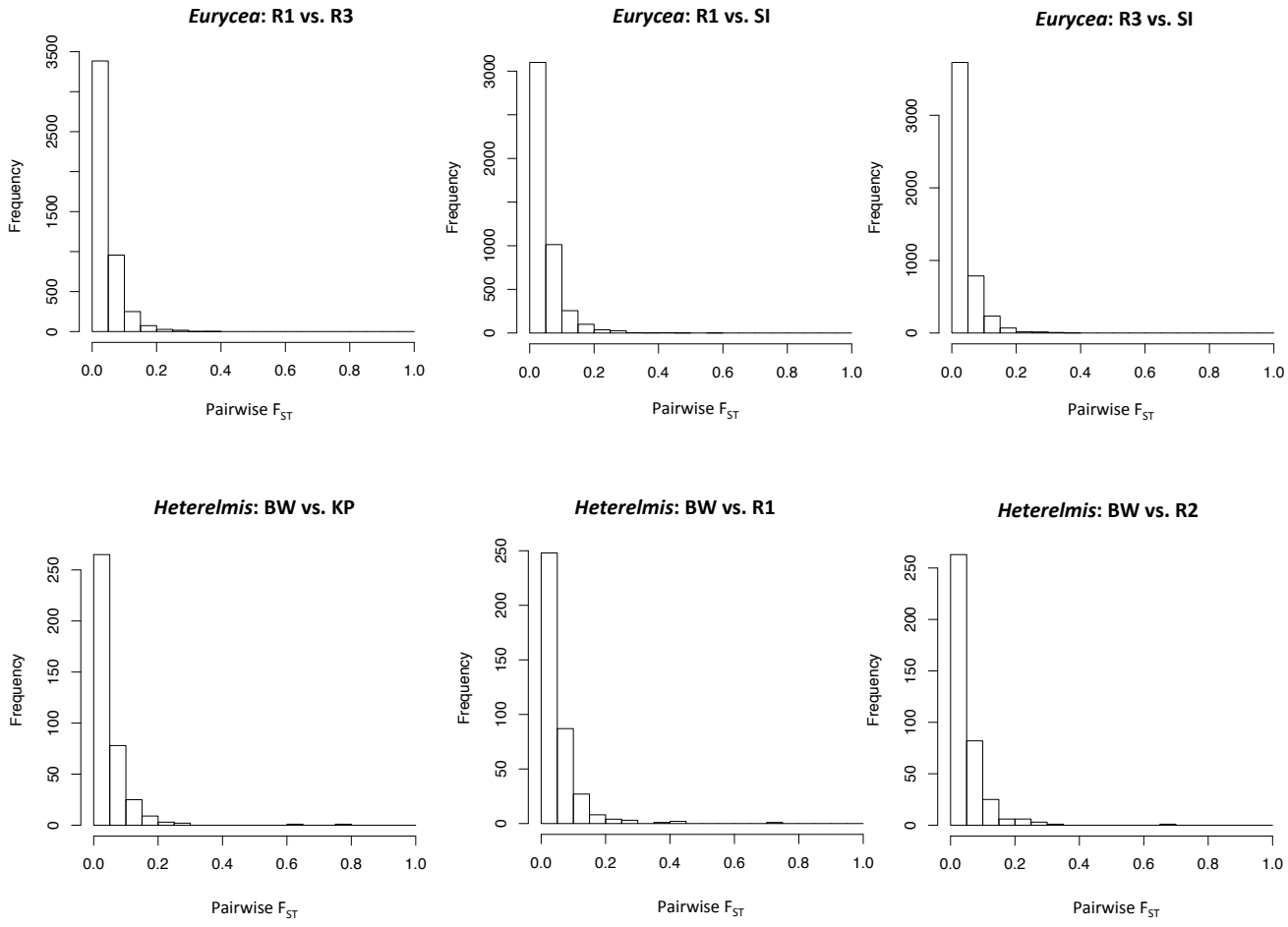

Figure S1: Distribution of F_{ST}s across SNPs for each pair of subpopulations per taxon.

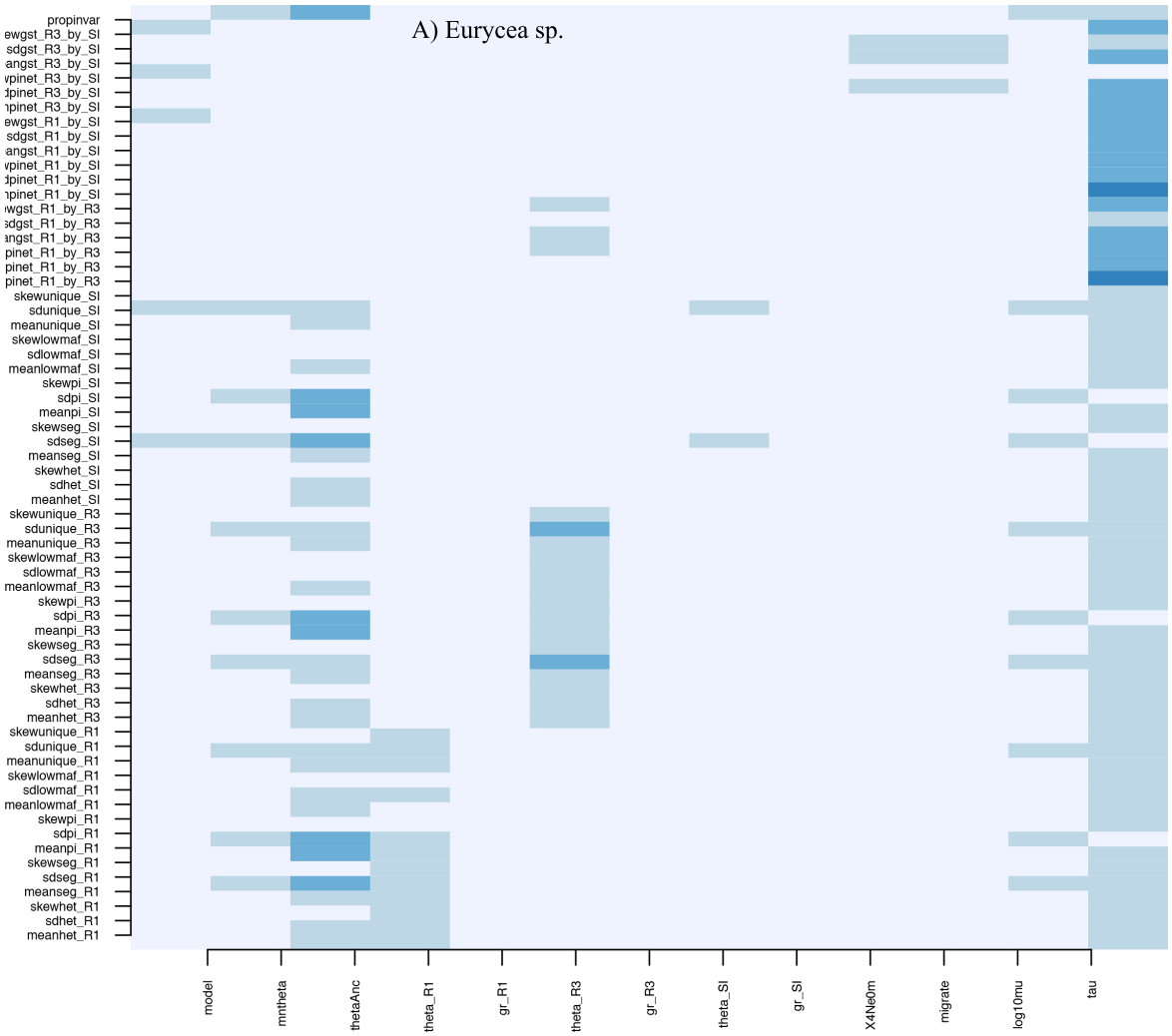

Figure S2: ABC diagnostics: Patterns of correlations of parameters (x-axis) with summary statistics (y-axis) for each taxon. The shades of blue represent five classes of correlation coefficients from white to dark blue: 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, 0.8-1.

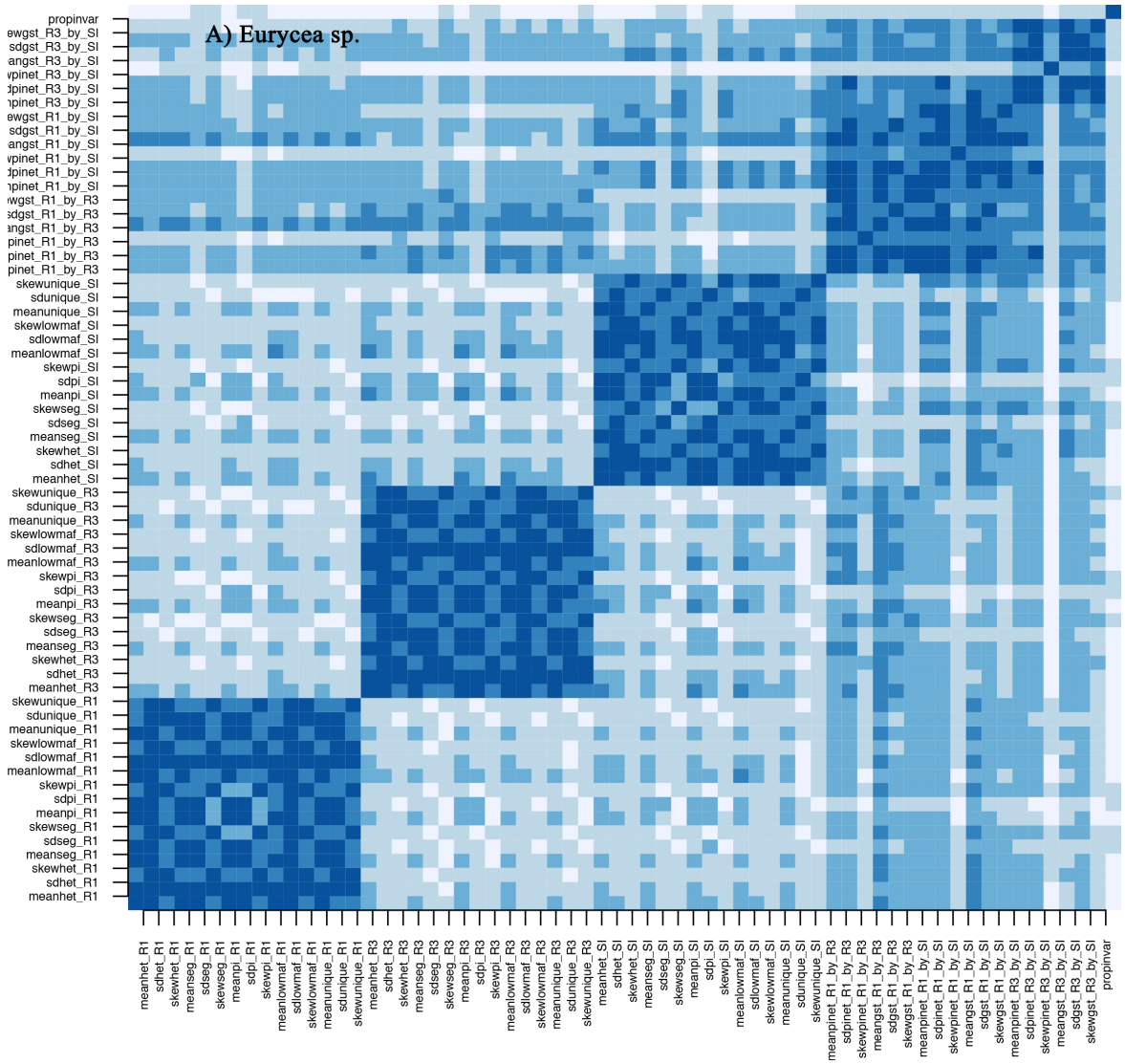

Figure S3: ABC diagnostics: Patterns of correlations between all pairs of summary statistics for each taxon. The shades of blue represent five classes of correlation coefficients from white to dark blue: 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, 0.8-1.

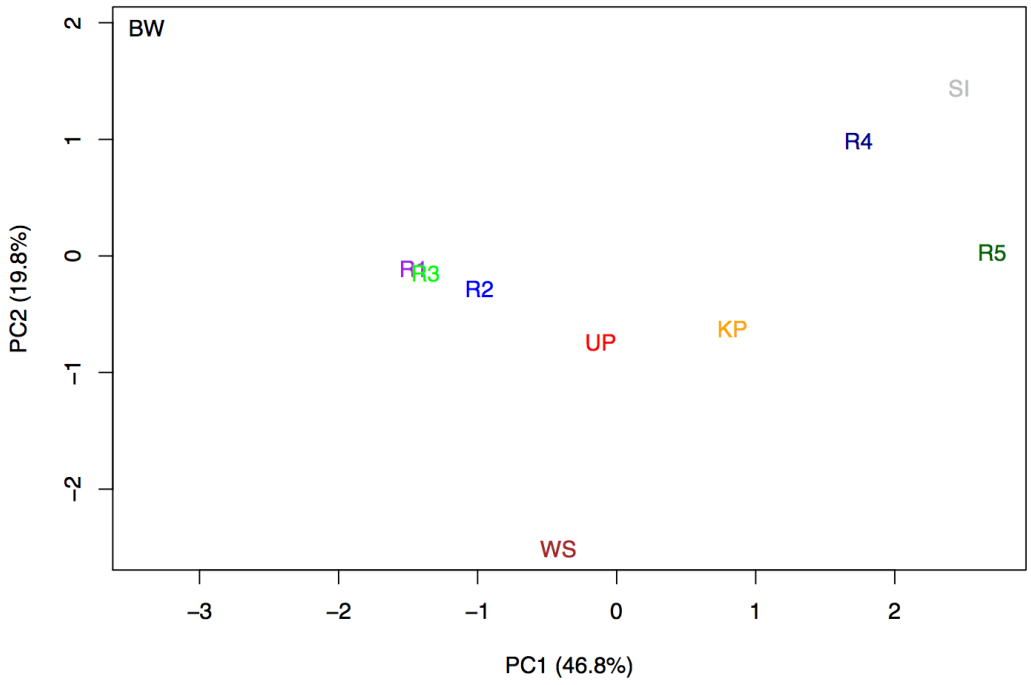

Figure S4: Environmental similarities among subpopulations based on PCA using the environmental values in Table 3.

Part 2: Comparative population genomics across the Texas Hill Country: *Heterelmis* Riffle Beetles and *Stygobromus* Amphipods

Chris C. Nice[1] and Lauren K. Lucas[1,2]

[1]Department of Biology, Texas State University, San Marcos, TX 78666, USA
[2]Department of Biology, Utah State University, Logan, UT 84322, USA

Introduction

Conservation genetics methods have been applied in a variety of contexts to provide critical information to inform management decisions (Allendorf & Luikart, 2009). Applications include the assessment of natural variation with and among populations or species of conservation concern, quantification of the potential for inbreeding or outbreeding depression in captive propagation programs, and the analysis of genealogical information for pedigree analysis. In addition, molecular genetics tools can be used to delineate lineages and identify the units of conservation (Allendorf & Luikart, 2009; Moritz, 1994; Forister *et al.*, 2008), which might be especially important in cases where the systematics of a lineage are not well-studied, or where cryptic lineages might occur due to morphological conservation or convergence (Niemiller *et al.*, 2008; Ethridge *et al.*, 2013). Here we examine the patterns of variation within two lineages of spring-endemic invertebrates from the Edwards Plateau of central Texas, *Heterelmis* riffle beetles (Family: Elmidae), and *Stygobromus* amphipods (Family Crangonyctidae). Both of these groups include species of conservation concern (*Heterelmis comalensis* found in Comal Springs and San Marcos Springs, and *Stygobromus pecki* found in Comal Springs and Hueco Springs) and congeneric species occurring in springs throughout the Edwards Plateau (Gonzales, 2008; Ethridge *et al.*, 2013). Our goal is to quantify allele frequency differences among populations and species within these lineages and to compare patterns of population genetic differentiation between these lineages. We use next-generation sequencing technology to generate multi-locus data for both lineages.

Methods

Individuals of *Heterelmis* and *Stygobromus* were collected from spring sites in the Edwards Plateau (Table 1). Samples were preserved in ethanol until genomic DNA was extracted using standard methods (Gonzales, 2008; Ethridge *et al.*, 2013). As described in Part 1 of this report, reduced representation genomic libraries were produced following the methods of Gompert *et al.* (2012) and Parchman *et al.* (2012). Libraries were produced for each individual specimen and then pooled for sequencing on the Illumina platform. Briefly, genomic DNA is digested with two restriction enzymes, EcoR1 and Mse1. Adapters that include the Illumina sequencing primer site and an 8-10bp barcode sequence (multiplex identifier sequence) were ligated to the sticky ends of these restriction fragments. The polymerase

chain reaction (PCR) was used to amplify these fragments which were then pooled across individuals within *Heterelmis* and *Stygobromus*. Amplified fragments between 250-350bp were excised from an agarose gel and extracted using the QiaQuick Gel extraction kit (QIAGEN Sciences, Germantown, MD, USA). The resulting pooled, genomic library was sequenced using the Illumina HiSeq sequencer at The National Center for Genome Research (NCGR, Santa Fe, NM).

Sequence reads were processed and filtered following the methods of Mandeville *et al.* (2015). We used a custom script to identify sequence reads to individual using the 8-10bp barcodes and to remove the barcode and the EcoR1 restriction site sequences from each read. In the absence of a reference genome for either taxon, we used a *de novo* assembly using the DNAStar SeqMan assembler to create scaffolds for a reference-based assembly for each taxon using the BWA (Burrows Wheeler Aligner) software (Li & Durbin, 2009). Consensus sequences from the *de novo* assembly were assembled to each other to screen out any potentially paralogous loci and the remaining consensus sequences form the scaffolds for reference-based assembly of all sequence reads.

For the *Heterelmis* sequence reads, we used SAMtools and BCFtools (Li *et al.*, 2009) to identify variable nucleotide sites, requiring at least 10% of individuals to have data at a site before it can be called variable. We removed variant sites with more than two alleles to avoid retaining any potentially paralogous loci. We used custom R scripts to filter loci, keeping only one, randomly chosen variant site per contig, and keeping variable markers whose median coverage was greater than 2x. Because we obtained more sequence reads from the *Stygobromus* library, we were more conservative in filtering. (It is possible that *Heterelmis* has a smaller genome than *Stygobromus*, or there are differences in nucleotide composition or restriction site frequencies between the two genera that might explain the differences in number of reads.) For *Stygobromus*, variants were called requiring at least 50% of individuals to have data at a site, and we filtered loci to retain those with median coverage greater than 4x. For both data sets, we converted the genotype likelihoods from SAMtools and BCFtools into composite genotypes.

We estimated genetic diversity for both data sets by using SAMtools and BCFtools to estimate expected heterozygosity, π, and the scaled effective population size, Waterson's θ. The genetic diversity measure $\theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate, in this case, the genome-wide mutation rate. Thus, θ is the mutation rate-scaled effective population size. We used the expectation-maximization (EM) algorithm, employing 20 iterations for each population to achieve convergence of estimates (Li, 2011).

For analyses of population genetic variation, we combined some sampling localities with small sample sizes to create larger sample sizes for some taxa. For example, two localities of *S. dejectus*, Cascade Caverns (n=13) and Sleath Cave (n=5), were combined into one sample of *S. dejectus* (n=18). Table 1 provides sampling details. Population allele frequencies and posterior probabilities of individual genotypes at all filtered loci were calculated using the hierarchical Bayesian model described by Gompert *et al.* (2013). Markov Chain Monte Carlo were used to calculate posterior probabilities and credible intervals with two chains, each with 6000 steps and a burnin of 1000 steps. Chain mixing and convergence were assessed in R using the coda package. The resulting allele frequency estimates and genotypes were used to calculate Nei's G_{ST} (Nei, 1973), an analog of F_{ST} (hereafter called F_{ST}), between

all pairs of localities within each taxon. Non-metric Multidimensional scaling (NMDS) was used to illustrate the patterns of differentiation among sampling localities and taxa for each genus. NMDS is more appropriate than hierarchical analyses when divergence is relatively recent or in situations where reticulate or clinal patterns might occur (Lessa, 1990).

As another measure of genetic differentiation, we used the point estimates of allele frequencies from the Bayesian model to calculate Nei's D_a (Nei *et al.*, 1983; Takezaki & Nei, 1996), a genetic distance metric based on allele frequency (dis)similarity. Pairwise distances between sampling localities were then used to construct unrooted Neighbor-Joining dendrograms using the R package APE. These genetic distances were also employed to examine patterns of isolation-by-distance. Pairwise geographic distances were calculated for all localities using the Great Circle distance calculation with the GEOSPHERE package in R. The VEGAN package was used to compare the matrices of geographic and genetic distances with a Mantel test in R.

Table 1: Sample size information for population genetic analyses. Samples with * indicate pooling of individuals from more than one sampling locality (see text).

Heterelmis			***Stygobromus***		
Nominal Species	Locality:	n	Nominal Species	Locality:	n
H. comalensis			*S. dejectus**		18
	Comal Spr.s*	80	*S. flagellatus*	San Marcos Spr.s	20
	San Marcos Spr.s	28	*S. longipes**		7
H. glabra			*S. pecki*	Comal Spr.s*	78
	Caroline Spr.s	10		Hueco Spr.s	12
	Dolan Spr.s	10	*S. sp.*	Fessenden Spr.s	11
	Fern Bank	25			
	Fessenden Spr.s	11			
H. sp.					
	Indian Spr.s	22			
H. vulnerata					
	Guadalupe R. Gonzales Co.	10			
	Guadalupe R. Kendall Co.	10			
	Plum Creek	11			
	San Marcos R.	10			

Results

Heterelmis

The *Heterelmis* library produced 9.5 x 10^6 usable short sequences (84-86bp in length). Estimates of genome-wide expected heterozygosity, π, and Waterson's θ were relatively similar across all sampling localities (Fig. 1). The genetic diversity estimates from population samples of *H. comalensis* are not substantially different from estimates from other *Heterelmis* species or localities.

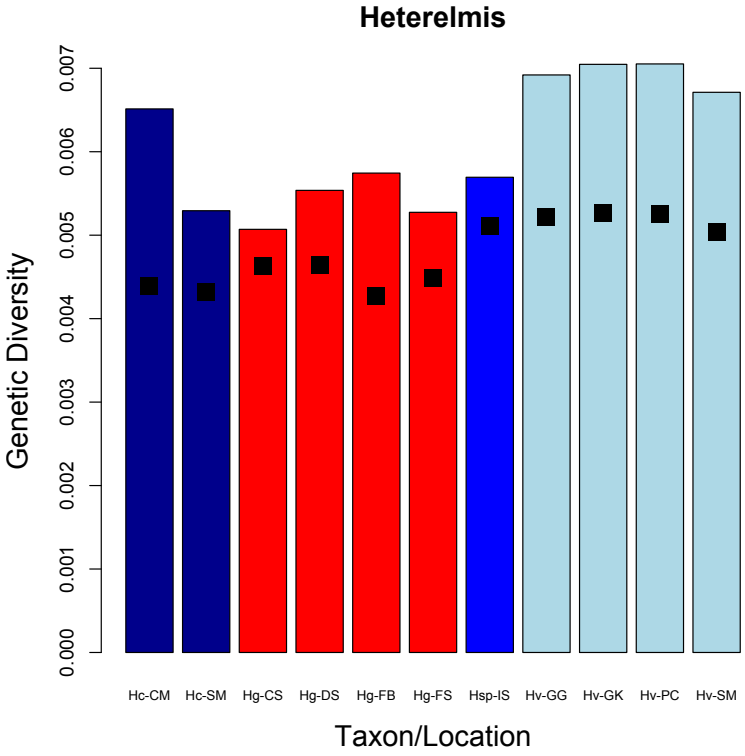

Figure 1: Genetic diversity across *Heterelmis* populations/localities. Genome-wide expected heterozygosity, π, for each locality depicted with bars, and Waterson's θ are indicated with black squares. Colors: dark blue = *H. comalensis*, red = *H. glabra*, blue = *H. sp.*, light blue = *H. vulnerata*. Locality abbreviations: Hc-CS = *H. comalensis* Comal Springs, Hc-SM = *H. comalensis* San Marcos Springs, Hg-SS = *H. glabra* Fessenden Springs, Hg-DS = *H. glabra* Dolan Springs, Hg-CS = *H. glabra* Caroline Springs, Hsp-IS = *H. sp.* Indian Springs, Hv-GK = *H. vulnerata* Guadalupe R. Kendall Co., Hv-GG = *H. vulnerata* Guadalupe R. Gonzales Co., Hv-SM = *H. vulnerata* San Marocs R., Hv-PC = *H. vulnerata* Plum Creek.

After assembly, variant calling and filtering, analyses proceeded on a data set of 116 loci (SNPs) with median coverage of 2x or higher. Estimates of differentiation between sampling localities based on F_{ST} (Table 2, Fig 2) and Nei's genetic distance (Table 3, Fig 3) showed similar patterns. Population samples from localities of *H. vulnerata* were distinctly identifiable from other populations and the first dimension of the NMDS of pairwise F_{ST} separates *H. vulnerata* from all other populations. Populations of *H. glabra* and *H. comalensis* were less differentiated from each other, although the two *H. comalensis* population samples from Comal Springs and San Marcos Springs had the smallest pairwise measures of differentiation. The *H. glabra* sample from Fessenden Springs is more distantly related to the other *H. glabra* populations and is separated from them on dimension 3. Finally, the population sample from Indian Springs (near Lake Amistad), whose nominal taxonomy is ambiguous, is also distantly related to the other *H. glabra* and *H. comalensis* populations and separated along dimension 2 (Fig.s 2, 3).

In the Neighbor-joining dendrogram based on pairwise values of Nei's D_a, samples of

Table 2: Pairwise estimates of F_{ST} (below the diaganol) and bootstrapped confidence intervals from 1000 bootstrap replicates (above the diaganol) for *Heterelmis* population samples.

Nominal Species: / Locality:	*H. comalensis* Comal Spr.s	*H. comalensis* San Marcos Spr.s	*H. glabra* Caroline Spr.s	*H. glabra* Dolan Spr.s	*H. glabra* Fern Bank	*H. glabra* Fessenden Spr.s
Comal Spr.s	0.000	0.008-0.017	0.019-0.033	0.024-0.05	0.012-0.028	0.024-0.068
San Marcos Spr.s	0.012	0.000	0.018-0.029	0.025-0.044	0.011-0.027	0.02-0.062
Caroline Spr.s	0.026	0.024	0.000	0.026-0.045	0.022-0.039	0.023-0.081
Dolan Spr.s	0.036	0.034	0.035	0.000	0.024-0.049	0.028-0.077
Fern Bank	0.020	0.018	0.03	0.036	0.000	0.027-0.088
Fessenden Spr.s	0.044	0.038	0.047	0.05	0.052	0.000
Indian Spr.s	0.093	0.085	0.081	0.093	0.097	0.101
Guadalupe R. Gonzales Co.	0.210	0.211	0.204	0.213	0.217	0.214
Guadalupe R. Kendall Co.	0.200	0.201	0.195	0.205	0.205	0.204
Plum Creek	0.211	0.210	0.206	0.213	0.218	0.213
San Marcos R.	0.195	0.195	0.191	0.196	0.202	0.201

Nominal Species: / Locality:	*H. sp.* Indian Spr.s	*H. vulnerata* Guadalupe R. Gonzales Co.	*H. vulnerata* Guadalupe R. Kendall Co.	*H. vulnerata* Plum Creek	*H. vulnerata* San Marcos R.
Comal Spr.s	0.043-0.15	0.17-0.256	0.161-0.246	0.171-0.254	0.16-0.234
San Marcos Spr.s	0.039-0.143	0.171-0.258	0.161-0.242	0.172-0.254	0.16-0.235
Caroline Spr.s	0.037-0.136	0.161-0.249	0.157-0.235	0.167-0.248	0.156-0.225
Dolan Spr.s	0.044-0.148	0.169-0.259	0.166-0.247	0.175-0.253	0.16-0.236
Fern Bank	0.044-0.166	0.176-0.263	0.168-0.245	0.176-0.265	0.165-0.243
Fessenden Spr.s	0.051-0.167	0.174-0.261	0.166-0.243	0.174-0.256	0.164-0.241
Indian Spr.s	0.000	0.183-0.28	0.181-0.267	0.183-0.274	0.179-0.262
Guadalupe R. Gonzales Co.	0.230	0.000	0.008-0.028	0.007-0.018	0.013-0.029
Guadalupe R. Kendall Co.	0.223	0.017	0.000	0.007-0.028	0.012-0.025
Plum Creek	0.228	0.012	0.016	0.000	0.009-0.029
San Marcos R.	0.219	0.021	0.018	0.018	0.000

H. vulnerata are distantly related to other *Heterelmis*. *H. comalensis* and *H. glabra* are more similar in terms of allele frequencies (i.e. smaller genetic distances) with the Indian Springs sample a bit more isolated (Fig.s 2, 3). The differentiation between *H. comalensis* and the *H. glabra* population at Fessenden Springs is greater in the analysis based on F_{ST} compared to the dendrogram based on Nei's D_a (Tables 2, 3, Fig.s 2, 3), reflecting the differences in these metrics of differentiation. It should also be noted that this Neighbor-Joining dendrogram does not represent a cladistic analysis and cannot be equated with a phylogenetic tree. Rather, it represents the relative patterns of allele frequency similarity among populations. The correlation between geographic and genetic distance (using Nei's D_a) was not significant for *Heterelmis* (Mantel statistic r: 0.07424, p=0.217). Removal of all pairwise comparisons involving the distantly related *H. vulnerata* also failed to reveal a significant pattern of isolation-by-distance (Mantel statistic r: 0.1201, p=0.154). There were no obvious patterns of differentiation that paralleled landscape features such as aquifers or rivers.

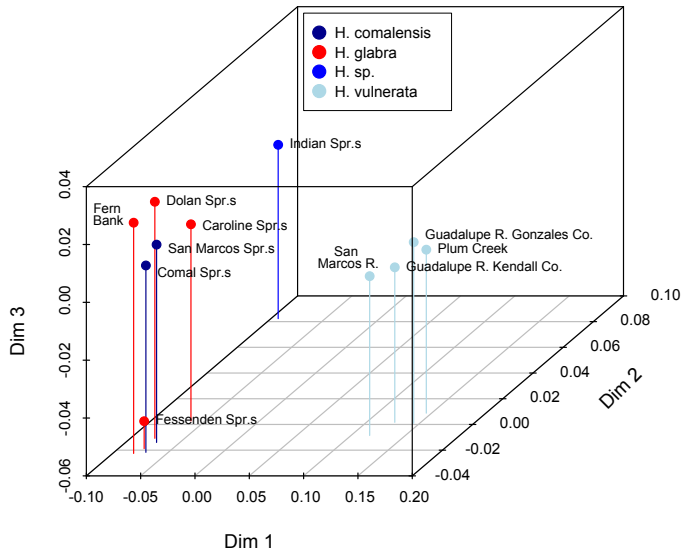

Figure 2: Non-metric Multidimensional Scaling (NMDS) ordination of pairwise, genome average F_{ST} values among *Heterelmis* populations/localities.

Table 3: Pairwise estimates of Nei's D_a for *Heterelmis* population samples. Nominal species and localities are in the same order as in Table 2.

Nominal Species: / Locality:	Hc-CS	Hc-SM	Hg-CS	Hg-DS	Hg-FB	Hg-FS	Hs-IS	Hv-GG	Hv-GK	Hg-PC	HV-SM
Comal Spr.s	0										
San Marcos Spr.s	0.006	0									
Caroline Spr.s	0.012	0.011	0								
Dolan Spr.s	0.016	0.014	0.014	0							
Fern Bank	0.010	0.009	0.014	0.016	0						
Fessenden Spr.s	0.015	0.013	0.016	0.017	0.018	0					
Indian Spr.s	0.031	0.028	0.025	0.03	0.032	0.03	0				
Guadalupe R. G	0.102	0.100	0.094	0.102	0.105	0.101	0.112	0			
Guadalupe R. K	0.098	0.096	0.092	0.100	0.100	0.097	0.110	0.012	0		
Plum Creek	0.104	0.101	0.098	0.103	0.107	0.102	0.111	0.009	0.011	0	
San Marcos R.	0.096	0.093	0.089	0.094	0.099	0.097	0.109	0.014	0.013	0.013	0

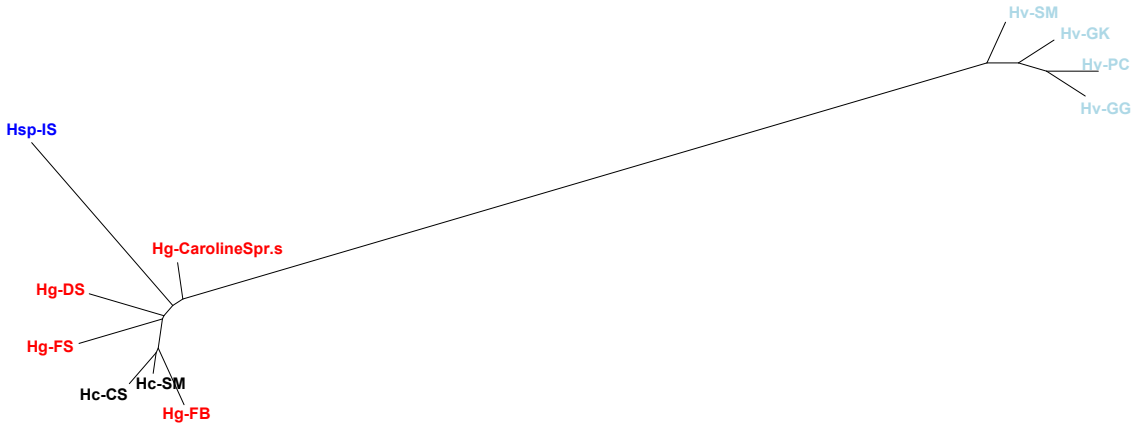

Figure 3: Unrooted Neighbor-joining dendrogram based on pairwise values of Nei's D_a among *Heterelmis* populations/localities. Locality label colors match colors in Fig. 2. Locality abbreviations: Hc-CS = *H. comalensis* Comal Springs, Hc-SM = *H. comalensis* San Marcos Springs, Hg-FS = *H. glabra* Fessenden Springs, Hg-DS = *H. glabra* Dolan Springs, Hg-CarolineSpr.s = *H. glabra* Caroline Springs, Hsp-IS = *H. sp.* Indian Springs, Hv-GK = *H. vulnerata* Guadalupe R. Kendall Co., Hv-GG = *H. vulnerata* Guadalupe R. Gonzales Co., Hv-SM = *H. vulnerata* San Marocs R., Hv-PC = *H. vulnerata* Plum Creek.

Stygobromus

The *Stygobromus* library produced 24 x 10^6 usable short sequences (84-86bp in length). Estimates of genome-wide expected heterozygosity, π, and Waterson's θ were relatively similar across all sampling localities (Fig. 4). As was the case for *Heterelmis*, the samples of the endangered taxon, *S. pecki*, are not substantially different from estimates from other *Stygobromus* species or localities. In fact, the sample from Comal Springs had the highest genome-wide heterozygosity, however, this is also the largest sample and is pooled across many subpopulations (Table 1), which might inflate these diversity estimates.

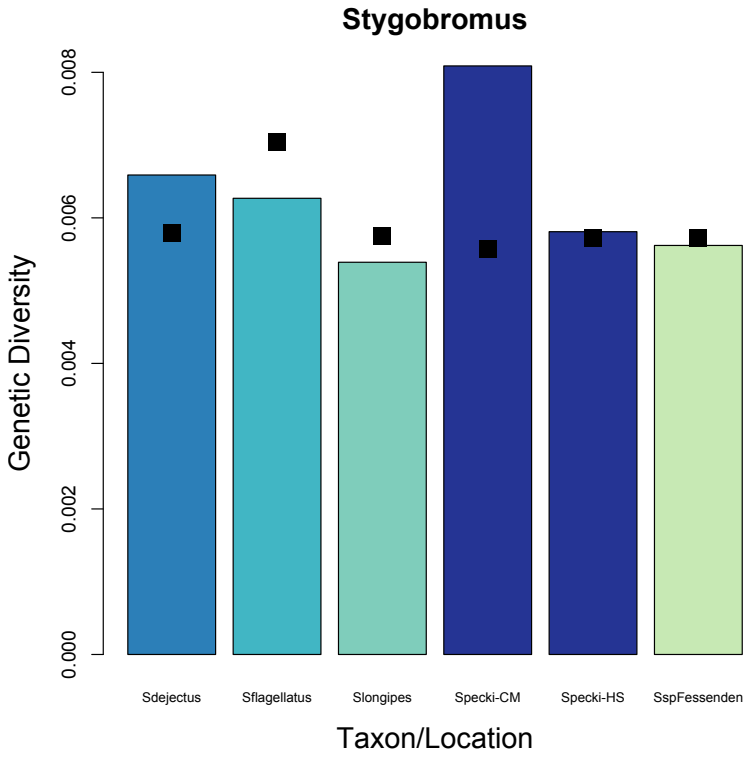

Figure 4: Genetic diversity across *Heterelmis* populations/localities. Genome-wide expected heterozygosity, π, for each locality depicted with bars, and Waterson's θ are indicated with black squares. The nominal species are indicated with different colors. Taxon/locality abbreviations: Sdejectus = *S. dejectus*, Sflagellatus = *S. flagellatus* San Marcos Springs, Slongipes = *S. longipes*, Specki-CM= *S. pecki* Comal Springs, Specki-HS = *S. pecki* Hueco Springs, SspFessenden = *S. sp.* Fessenden Springs (refer to Table 1).

Analyses were conducted on 129 loci with at least a median of 4x coverage. Measures of population differentiation clearly distinguished species and populations. In the NMDS ordination of populations by F_{ST}, the samples of *S. pecki* from Comal Springs and Hueco Springs are the least differentiated (Fig. 5). Dimension 1 distinguished *S. flagellatus* from all other species. Dimension 2 separates *S. dejectus* from other taxa, and dimension 3 separates *S. longipes*. The population sample from Fessenden Springs, whose taxonomic status is ambiguous, appears to be most similar to *S. pecki* (Fig. 5, Table 4).

Table 4: Pairwise estimates of F_{ST} (below the diaganol) and bootstrapped confidence intervals from 1000 bootstrap replicates (above the diaganol) for *Stygobromus* population samples.

Nominal Species: /Locality:	*S. dejectus*	*S. flagellatus* San Marcos Spr.s	*S. longipes*	*S. pecki* Comal Spr.s	*S. pecki* Hueco Spr.s	*S. sp.* Fessenden Spr.s
S. dejectus	0	0.138-0.190	0.066-0.160	0.051-0.162	0.037-0.123	0.067-0.149
S. flagellatus	0.162	0	0.141-0.202	0.135-0.194	0.134-0.189	0.115-0.171
S. longipes	0.110	0.171	0	0.056-0.136	0.047-0.114	0.066-0.147
S. pecki-CS	0.101	0.163	0.093	0	0.024-0.085	0.040-0.114
S. pecki-HS	0.077	0.161	0.079	0.052	0	0.044-0.093
S. sp.	0.106	0.142	0.104	0.075	0.066	0

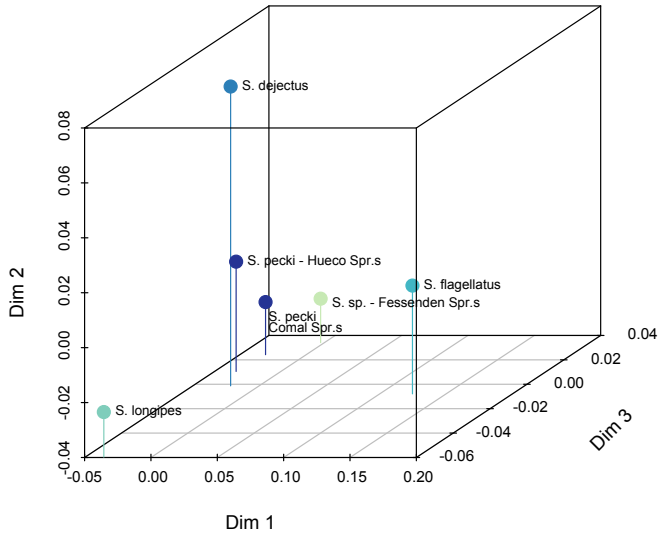

Figure 5: Non-metric Multidimensional Scaling (NMDS) ordination of pairwise, genome average F_{ST} values among *Stygobromus* populations/localities.

The dendrogram based on pairwise estimates of Nei's D_a (Fig. 6) illustrates patterns of differentiation that are similar to those observed for pairwise F_{ST} (Fig. 5). The most differentiation appears between *S. flagellatus* and all other samples. The two *S. pecki* samples are the two least differentiated populations. The other two nominal species, *S. dejectus* and *S. longipes*, along with the Fessenden Springs sample are more similar to *S. pecki* compared to *S. flagellatus*. However, the Fessenden Springs sample does not appear to be as closely related to *S. pecki* compared to the ordination based on pairwise F_{ST} (Fig. 5). This difference is a function of both the genetic distance metrics (F_{ST} vs. Nei's D_a), and the algorithms used to illustrate the patterns. The unrooted Neighbor-joining algorithm forces bifurcating relationships which might not reflect the actual history of the taxa and populations. The correlation between geographic and genetic distance (using Nei's D_a) was not significant for *Stygobromus* (Mantel statistic r:-0.05013, p=0.33889). As with *Heterelmis*, there were no obvious patterns of differentiation that paralleled landscape features such as aquifers or

838 rivers.

Table 5: Pairwise estimates of Nei's D_a for *Heterelmis* population samples. Nominal species and localities are in the same order as in Table 2.

Nominal Species: /Locality:	*S. dejectus*	*S. flagellatus* San Marcos Spr.s	*S. longipes*	*S. pecki* Comal Spr.s	*S. pecki* Hueco Spr.s	*S. sp.* Fessenden Spr.s
S. dejectus	0					
S. flagellatus	0.091	0				
S. longipes	0.043	0.089	0			
S. pecki-CS	0.045	0.093	0.036	0		
S. pecki-HS	0.033	0.088	0.030	0.024	0	
S. sp.	0.046	0.080	0.040	0.030	0.027	0

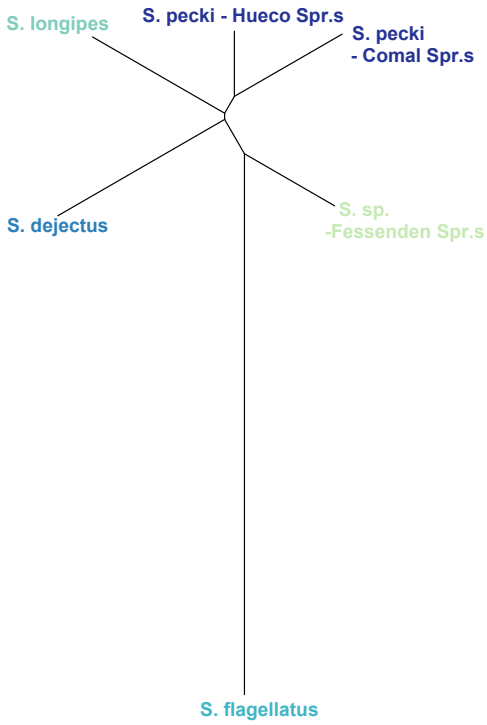

Figure 6: Unrooted Neighbor-joining dendrogram based on pairwise values of Nei's D_a among *Stygobromus* populations/localities. Locality label colors match colors in Fig.s 4 and 5.

839 Discussion

840 In the second part of this project, we used GBS data to examine the relationships between
841 species endemic to Comal Springs and congeneric species and populations across the Edwards
842 Plateau. Specifically, we focused on populations of *H. comalensis* and *S. pecki* and congeners

from central Texas with the goal of surveying the geographic distribution of genetic variation as well as illustrating the genetic distances between nominal taxa. This was undertaken with the goal of placing *H. comalensis* and *S. pecki* into a comparative and biogeographical context. A full systematic treatment of *Heterelmis* and *Stygobromus* is not possible with the current sampling.

Levels of genetic diversity within *H. comalensis* and *S. pecki* are not substantially different from diversity observed in congeners (Fig.s 1, 4). Estimates of pairwise population differentiation (F_{ST}) and genetic distance (Nei's D_a) were used to illustrate relationships based on allele frequency similarities. Within *Heterelmis*, *H. comalensis* appears to be closely related to *H. glabra*, but quite distinct from *H. vulnerata*. The population sample from Indian Springs, whose taxonomic status is ambiguous due to lack of clearly identifying morphological characters, is more similar to the *H. comalensis* - *H. glabra* cluster of populations than to *H. vulnerata*, but appears distinct, possibly representing an independent lineage (Fig.s 2, 3). These patterns, which are based on estimates of population differentiation and allele frequency similarity using many nuclear markers, comport with the patterns detected in a phylogenetic analysis of mitochondrial DNA (mtDNA) sequence variation (Gonzales, 2008) which also showed close relationships between *H. comalensis* and *H. glabra*, with *H. vulnerata* quite distantly related. However, unlike the patterns detected using mtDNA, the Fern Bank sample of *H. glabra* appears to be much less distinct from the *H. comalensis*.

For the amphipods, *S. pecki*, from Comal Springs and Hueco Springs, have the least amount of differentiation among *Stygobromus* populations sampled, but samples of *S. longipes*, *S. dejectus* and an unknown or undescribed taxon from Fessenden Springs are all somewhat closely related to *S. pecki*. *Stygobromus flagellatus* sampled from San Marcos Springs is distantly related to the other taxa (Fig.s 5, 6). Ethridge *et al.* (2013) reported similar patterns in a survey of mtDNA sequence variation.

These population-level data form the basis of an improved understanding of the patterns of geographic genetic variation for these spring-endemic taxa in the Edwards Plateau. However, they do not fully resolve many of the taxonomic issues that continue to persist for both *Heterelmis* and *Stygobromus* from central Texas. For example, Ethridge *et al.* (2013) uncovered complex patterns of relatedness among samples nominally considered as *S. flagellatus*. In the current data, only one sampling locality for *S. flagellatus* was included, which leaves open the question of whether there might be more than one cryptic lineage within this nominal species as suggested by Ethridge *et al.* (2013). Future investigations with a broader geographic and taxonomic scope, and including in-depth analyses of morphological variation, with be required to fully resolve the systematics of both of these groups from the Edwards Plateau.

The overall picture of the geographic distribution of of genetic variation in both taxa is interesting from the perspective that there does not appear to be a strong relationship between geographic distance among populations and their patterns of differentiation. Mantel tests failed to detect any significant correlation between geographic distance and genetic distance. Nor was there any obvious evidence that genetic variation is organized by river system or along aquifer boundaries, which comports with patterns observed in some other Edwards Plateau, spring-associated organisms (e.g. Lucas *et al.*, 2009), but differs from what is sometimes observed in fish where variation can be structured by river drainage

(e.g. Richardson & Gold, 1995). The absence of evidence of obvious phylogeographical structure might reflect a complex history of colonization of a complicated karst landscape by the members of each genus, during which simple patterns of isolation-by-distance have been erased or were never established. Certainly, the isolation of springs on the Edwards Plateau could have contributed to the complex biogeographical history of these organisms. Alternatively, the sampling in the present study might be insufficient, both geographically and taxonomically, to detect what might be subtle biogeographical patterns. We hope to rectify this potential sampling problem in the near future.

Beyond testing evolutionary and biogeographical hypotheses, molecular genetics data have been touted as a solution to conservation problems where population enumeration is difficult or impossible. In this context, indirect estimates are best used in a relative and non-quantitative comparative context because the translation of genetic diversity estimates into estimates of actual population sizes requires numerous and potentially dubious assumptions. Molecular data can be, and have been, used for indirect estimates of population parameters such as the effective population size, N_e, which are often of interest in conservation and management situations. Despite the attraction for such uses of molecular data, indirect estimates based on measures of genetic diversity require a large number of assumptions and often rely on estimates of mutation rates. One standard procedure using a single population genetic sample is to decompose estimates of θ to obtain an estimate of effective population size, N_e (Roman & Palumbi, 2003; Allendorf & Luikart, 2009; Hare *et al.*, 2011). This works in the context of the neutral theory of molecular evolution (Kimura, 1983). Assuming that mutation and genetic drift are the only evolutionary forces acting on population genetic variation and that populations are at an equilibrium between mutation and drift, then $\theta = 4N_e\mu$, where μ is the mutation rate. We have obtained genome-wide estimates of θ for *Heterelmis* and *Stygobromus* population samples (Fig.s 1, 4). However, we have no understanding of mutation rate variation in these taxa and there are no estimates from any closely related taxa. Further complicating this situation, meaningful comparisons across lineages, for example, between *H. comalensis* and congeneric populations, requires that mutation rates are not different among lineages. Estimation of evolutionarily relevant mutation rates is difficult, and empirical evidence indicates that there is substantial variation among lineages (Baer *et al.*, 2007; Haag-Liautard *et al.*, 2007). Another complication is that the effective population size, N_e, estimated in this way is likely not equivalent to the census population for a variety of good reasons (Hartl *et al.*, 1997; Allendorf & Luikart, 2009; Waples, 1991; Hare *et al.*, 2011). The most important of these reasons for the present case is that indirect estimates of N_e based on the decomposition of θ are long-term average population sizes which are not relevant for estimation of contemporary census population size (Waples, 1991; Hare *et al.*, 2011). Given these problems, we suggest that relative comparisons of genetic diversity is most appropriate for the data reported here. In the future, two-sample approaches, involving population samples at two time points, or more explicitly model-based methods could be used for a more precise estimation of population size. Our estimates of θ and expected heterozygosity, π, for both lineages suggest that the endangered taxa are not genetically depauperate with respect to congeneric populations, and, assuming equal mutation rates, these populations' average (long-term) sizes have not been not dramatically different from their congeners (Fig.s 1, 4).

In conclusion, our analysis of population genetic variation within and among lineages of *Heterelmis* and *Stygobromus* indicate that the populations of endangered species have levels of genetic diversity that are comparable to populations of congeneric species from central Texas. The patterns of differentiation among lineages indicate that both *H. comalensis* and *S. pecki* have closely related species in the near vicinity: *H. comalensis* is closely related to the nominal *H. glabra*, and *S. pecki* is distinct from, but presumably recently diverged from *S. longipes*, *S. dejectus*, and an unnamed lineage from Fessenden Springs. Neither the riffle beetles, nor the amphipods show patterns of isolation-by-distance given the geographic extent of sampling in the current study. In addition and somewhat unexpectedly, the geographic organization of genetic variation does not appear to follow river drainages or aquifer boundaries in either lineage. Expansion of the taxonomic and geographic breadth of investigations of these lineages in future studies will be required for a comprehensive understanding of the evolutionary histories of these organisms in the Edwards Plateau area and beyond.

Acknowledgements

This research (Part 2) was funded by Texas Parks and Wildlife (Section 6) to LKL and CCN. We thank especially Randy Gibson, Josh Etheridge, Tina Gonzales, Joe Fries, Pete Diaz, Ethan Chapell, Jean Krejca, Andy Gluesenkamp, and the USFWS Aquatic Resources Center staff for their help collecting specimens. This project benefitted greatly from discussions with Zach Gompert, Alex Buerkle, Kate Bell, Randy Gibson and Andy Gluesenkamp. For allowing access to *Stygobromus* sampling localities, the authors thank Cascade Caverns (S. Kyle), Cave-Without-A- Name (T. Summers and M. Burrell), Magic Springs (W. McAllister), CM Cave (J. Eisenhauer), Comal County Water Recreation District #1, the Nature Conservancy (S. McWilliams and J. Karges), the Balcones Can- yonlands Preserve (M. Sanders), the Devils River State Natural Area (J. Ranzau and D. Riskind), the Edwards Aquifer Authority (G. Schindel), the Camp Bullis Military Training Reservation (L. Cooksey), Texas State University (G. Longley and V. Castillo), Southwestern University (B. Pierce), and the cities of Austin, Georgetown, New Braunfels, and San Marcos.

References

Allendorf, F. W. & Luikart, G. (2009). *Conservation and the genetics of populations*. John Wiley & Sons.

Baer, C. F., Miyamoto, M. M. & Denver, D. R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, 8, 619–631.

Barr, C. & Spangler, P. (1992). A new genus and species of stygobiontic dryopid beetle, stygoparnus comalensis (coleoptera: Dryopidae), from comal springs, texas. *Proceedings of the Biological Society of Washington*, 105, 40–54.

Beaumont, M. A., Zhang, W. & Balding, D. J. (2002). Approximate bayesian computation in population genetics. *Genetics*, 162, 2025–2035.

968 Bendik, N. & Gluesenkamp, A. (2013). Body length shrinkage in an endangered amphibian
969 is associated with drought. *Journal of Zoology*, 290, 35–41.

970 Bertorelle, G., Benazzo, A. & Mona, S. (2010). Abc as a flexible framework to estimate
971 demography over space and time: some cons, many pros. *Molecular ecology*, 19, 2609–
972 2625.

973 Bilton, D. T., Freeland, J. R. & Okamura, B. (2001). Dispersal in freshwater invertebrates.
974 *Annual review of ecology and systematics*, 159–181.

975 Brune, G. M. (2002). *Springs of Texas*, vol. 1. Texas A&M University Press.

976 Buerkle, C. A. & Gompert, Z. (2013). Population genomics based on low coverage sequencing:
977 how low should we go? *Molecular Ecology*, 22, 3028–3035.

978 Chippindale, P. T., Price, A. H., Wiens, J. J. & Hillis, D. M. (2000). Phylogenetic relation-
979 ships and systematic revision of central texas hemidactyliine plethodontid salamanders.
980 *Herpetological monographs*, 1–80.

981 Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J. K. (2010). Using environmental
982 correlations to identify loci underlying local adaptation. *Genetics*, 185, 1411–1423.

983 Csilléry, K., Blum, M. G., Gaggiotti, O. E. & François, O. (2010). Approximate bayesian
984 computation (abc) in practice. *Trends in ecology & evolution*, 25, 410–418.

985 Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. &
986 Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (gbs) approach for high
987 diversity species. *PLoS ONE*, 6, e19379.

988 Ethridge, J. Z., Gibson, J. R. & Nice, C. C. (2013). Cryptic diversity within and amongst
989 spring-associated stygobromus amphipods (amphipoda: Crangonyctidae). *Zoological Jour-*
990 *nal of the Linnean Society*, 167, 227–242.

991 Forister, M. L., Nice, C. C., Fordyce, J. A., Gompert, Z. & Shapiro, A. M. (2008). Consid-
992 ering evolutionary processes in the use of single-locus genetic data for conservation, with
993 examples from the Lepidoptera. *Journal of Insect Conservation*, 12, 37–51.

994 Fries, J. N. (2002). Upwelling flow velocity preferences of captive adult san marcos salaman-
995 ders. *North American journal of aquaculture*, 64, 113–116.

996 Gibson JR, Harden SJ, F. J. (2008). Survey and distribution of invertebrates from se-
997 lected springs of the edwards aquifer in comal and hays counties, texas. *The Southwestern*
998 *Naturalist*, 53, 74–84.

999 Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Buerkle, C. A. & Forister, M. L.
1000 (2013). Geographically multifarious phenotypic divergence during speciation. *ECOLOGY*
1001 *AND EVOLUTION*, 3, 595–613.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L. & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between tow butterfly species. *Evolution*, 66, 2167–2181.

Gompert, Z., Nice, C. C., Fordyce, J. A., Forister, M. L. & Shapiro, A. M. (2006). Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology*, 15, 1759–1768.

Gonzales, T. K. (2008). *Conservation genetics of the Comal Springs riffle beetle (*Heterelmis comalensis*) populations in central Texas, with examination of molecular and morphological variation in Heterelmis sp. throughout Texas*. Master's thesis.

Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D. L., Charlesworth, B. & Keightley, P. D. (2007). Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature*, 445, 82–85.

Haldane, J. (1948). The theory of a cline. *Journal of genetics*, 48, 277–284.

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., Ruegg, K. & Palstra, F. (2011). Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology*, 25, 438–449.

Hartl, D. L., Clark, A. G. & Clark, A. G. (1997). *Principles of population genetics*, vol. 116. Sinauer associates Sunderland.

Hermoso, V., Linke, S., Prenda, J. & Possingham, H. (2011). Addressing longitudinal connectivity in the systematic conservation planning of fresh waters. *Freshwater Biology*, 56, 57–70.

Hey, J. & Nielsen, R. (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, 167, 747–760.

Holt, R. D. & Gomulkiewicz, R. (1997). How does immigration influence local adaptation? a reexamination of a familiar paradigm. *American Naturalist*, 563–572.

Hudson, R. R. (2002). Generating samples under a wright–fisher neutral model of genetic variation. *Bioinformatics*, 18, 337–338.

Hughes, J. M., Huey, J. A. & Schmidt, D. J. (2013). Is realised connectivity among populations of aquatic fauna predictable from potential connectivity? *Freshwater Biology*, 58, 951–966.

Humphreys, W. (2009). Hydrogeology and groundwater ecology: Does each inform the other? *Hydrogeology Journal*, 17, 5–21.

Kawecki, T. J. & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology letters*, 7, 1225–1241.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Knowles, L. L. & Maddison, W. P. (2002). Statistical phylogeography. *Molecular Ecology*, 11, 2623–2635.

Lande, R. (1999). *Extinction risks from anthropogenic, ecological, and genetic factors*. Princeton, New Jersey, Princeton University Press.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17, 183–189.

Lessa, E. P. (1990). Multidimensional-analysis of geographic genetic-structure. *Systematic Zoology*, 39, 242–252.

Li, H. (2011). A statistical framework for snp calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27, 2987–2993.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. *et al.* (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25, 2078–2079.

Loáiciga, H., Maidment, D. & Valdes, J. (2000). Climate-change impacts in a regional karst aquifer, texas, usa. *Journal of Hydrology*, 227, 173–194.

Longley, G. (1981). The edwards aquifer: Earth's most diverse groundwater ecosystem? *International Journal of Speleology*, 11, 12.

Lucas, L., Gompert, Z., Ott, J. & Nice, C. (2009). Geographic and genetic isolation in spring-associated *Eurycea* salamanders endemic to the Edward's Plateau region of Texas. *Conservation Genetics*, DOI: 10.1007/s10592–008–9710–2.

Mandeville, E. G., Parchman, T. L., McDonald, D. B. & Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of catostomus species. *Molecular ecology*, 24, 1856–1872.

Marten, A., Braendle, M. & Brandl, R. (2006). Habitat type predicts genetic population differentiation in freshwater invertebrates. *Molecular Ecology*, 15, 2643–2651.

Meyer, M. & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010, pdb–prot5448.

Moritz, C. (1994). Defining evolutionarily significant units for conservation. *Trends in Ecology & Evolution*, 9, 373–375.

Moritz, C. (1999). Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas*, 130, 217–228.

Murphy, N. P., Guzik, M. T. & Wilmer, J. W. (2010). The influence of landscape on population structure of four invertebrates in groundwater springs. *Freshwater Biology*, 55, 2499–2509.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70, 3321–3323.

Nei, M., Tajima, F. & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, 19, 153–170.

Nielsen, R. & Beaumont, M. A. (2009). Statistical inferences in phylogeography. *Molecular ecology*, 18, 1034–1047.

Niemiller, M. L., Fitzpatrick, B. M. & Miller, B. T. (2008). Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae : *Gyrinophilus*) inferred from gene genealogies. *Molecular Ecology*, 17, 2258–2275.

Page, T. J. & Hughes, J. M. (2014). Contrasting insights provided by single and multispecies data in a regional comparative phylogeographic study. *Biological Journal of the Linnean Society*, 111, 554–569.

Parchman, T. L., Z. Gompert, a. J. M., Shilkey, F. D., Benkman, C. W. & Burkle, C. A. (2012). Genome-wide association of genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, 21, 2991–3005.

Patterson, C. (1999). *Evolution*. Comstock Publishing Associates.

Richardson, L. R. & Gold, J. R. (1995). Evolution of the cyprinella lutrensis species-complex. ii. systematics and biogeography of the edwards plateau shiner, cyprinella lepida. *Copeia*, 28–37.

Robertson, H. L., Guzik, M. T. & Murphy, N. P. (2014). Persistence in the desert: ephemeral waterways and small-scale gene flow in the desert spring amphipod, *Wangiannachiltonia guzikae*. *Freshwater Biology*, 59, 653–665.

Roman, J. & Palumbi, S. R. (2003). Whales before whaling in the north atlantic. *science*, 301, 508–510.

Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics*, 75, 733–756.

Slatkin, M. (1985). Gene flow in natural populations. *Annual review of ecology and systematics*, 393–430.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236, 787–792.

1106 Storfer, A. (1999). Gene flow and endangered species translocations: a topic revisited.
1107 *Biological Conservation*, 87, 173–180.

1108 Tajima, F. (1983). Evolutionary relationship of dna sequences in finite populations. *Genetics*,
1109 105, 437–460.

1110 Takezaki, N. & Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees
1111 from microsatellite dna. *Genetics*, 144, 389–399.

1112 Wakeley, J. (2009). *Coalescent theory: an introduction*, vol. 1. Roberts & Company Pub-
1113 lishers Greenwood Village, Colorado.

1114 Waples, R. S. (1991). Genetic methods for estimating the effective size of cetacean popula-
1115 tions. *Report of the International Whaling Commission (special issue)*, 13, 279–300.

1116 Watterson, G. (1975). On the number of segregating sites in genetical models without
1117 recombination. *Theoretical population biology*, 7, 256–276.

1118 Wentworth, C. K. (1922). A scale of grade and class terms for clastic sediments. *The Journal*
1119 *of Geology*, 377–392.

1120 Whitaker, R. J., Grogan, D. W. & Taylor, J. W. (2003). Geographic barriers isolate endemic
1121 populations of hyperthermophilic archaea. *Science*, 301, 976–978.

1122 Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16, 97.