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

LAUREN K. LUCAS*[†], ZACHARIAH GOMPERT[†], J. RANDY GIBSON[‡], KATHERINE L. BELL*,
C. ALEX BUERKLE[§] AND CHRIS C. NICE*

Department of Biology, Texas State University, San Marcos, TX, U.S.A.
[†]*Department of Biology, Utah State University, Logan, UT, U.S.A.*
[‡]*U.S. Fish & Wildlife Service, San Marcos Aquatic Resources Center, San Marcos, TX, U.S.A.*
[§]*Department of Botany, University of Wyoming, Laramie, WY, U.S.A.*

SUMMARY

1. We studied genetic variation in four endangered animal taxa in the largest freshwater spring complex in the southwestern U.S.A., Comal Springs (TX): *Eurycea* salamanders, *Heterelmis* riffle beetles, *Stygobromus* amphipods and *Stygoparnus* dryopid beetles. They inhabit a spring complex with nearly 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 demographic 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. An 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 tested were associated with environmental conditions and provide some evidence for potential local adaptation to slightly variable conditions across habitat patches 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 recolonise from elsewhere within the spring complex. However, genetic variants affecting survival in particular habitat patches could be lost during such droughts.

Keywords: approximate Bayesian computation, aquatic invertebrate, endangered species, genotyping-by-sequencing, neotenic salamander

Introduction

Comparative phylogeography can help explain the historical and contemporary 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, Grogan & Taylor, 2003; Marten, Braendle & Brandl, 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, Freeland & Okamura, 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, Guzik & Wilmer, 2010; Robertson, Guzik &

Correspondence: Lauren K. Lucas, Department of Biology, 5305 Old Main Hill, Utah State University, Logan, UT 84322-5305, U.S.A. E-mail: llucas@usu.edu

Murphy, 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 to 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 end of the 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, Huey & Schmidt, 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).

The limestone aquifer systems of the Edwards Plateau in Texas, U.S.A., are biotically diverse and home to a large number of locally endemic species (Longley, 1981; Brune, 2002, e.g., Figure S1 in Supporting Information). Here, we focus on endemism found in Comal Springs, the largest spring system not only in the Edwards Plateau but also in the southwestern United States. It consists of 425 spring openings that feed into Landa Lake and six spring runs (C. Norris *et al.*, unpubl. data). 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 U.S.A., 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 (Fig. 1b) are the most susceptible to the cessation of flow during droughts (C. Norris *et al.*, unpubl. data). Spring runs R1, R2 and R3 (Fig. 1b) 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 metres 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, Fig. 1b), 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 G. Schindel, unpubl. data). The flow decreases because of climate variation and general warming (Loáiciga, Maidment & Valdes, 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 withdrawals, Comal Springs is home to several endemic and federally listed species. They are found in several habitat patches within the Comal Springs complex (referred to as subpopulations herein; each coloured two-letter code in Fig. 1b–d is placed at a subpopulation). The springs have relatively stable temperatures of about 23.3 °C, nearly neutral pH and are surrounded by 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. For example, droughts may lower the water table such that previously used conduits are no longer accessible. Furthermore, toxins from chemical spills into the aquifer system could easily affect most or all subpopulations within the spring complex.

We focus on the comparative phylogeography 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 S1).
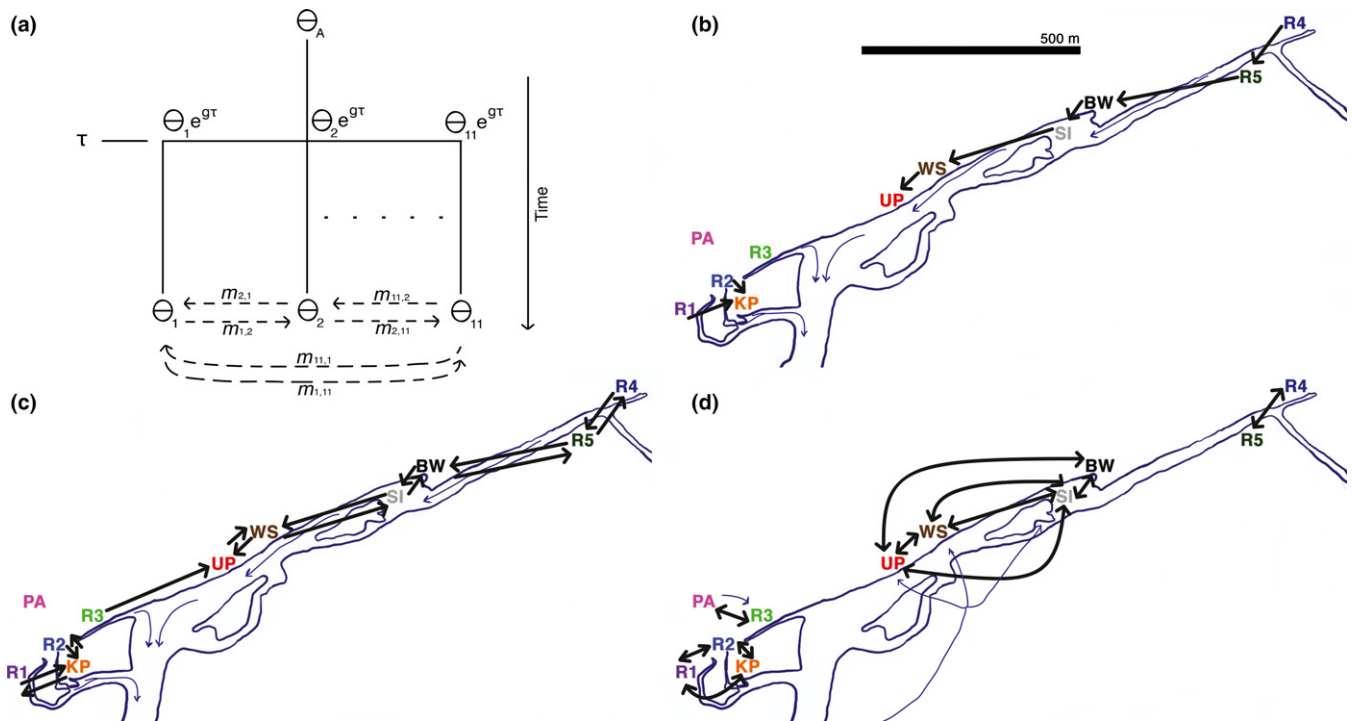
Fig. 1 Model testing using ABC. Pane (a) is the genealogy we simulated at each contig. Panes (b–d) are maps of Comal Springs showing the localities of the 11 subpopulations (coloured two-letter codes). We tested five models of gene flow for each taxon: no migration among subpopulations, an island model, 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.

From here on, we will refer to these taxa by their generic names: *Eurycea*, *Heterelmis*, *Stygobromus* and *Stygoparnus*. The latter three of these are recognised as federally endangered species and are only found in one other spring complex. 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 and are generally found under substrate at or near spring openings (e.g., Barr & Spangler, 1992). In addition, Edwards Plateau *Eurycea* species are thought to lay eggs and seek refuge within subterranean habitats when aboveground conditions are unfavourable (Chippindale *et al.*, 2000; Fries, 2002; Bendik & Gluesenkamp, 2013). Later-instar *Heterelmis* larvae drift, perhaps to locate favourable habitat for pupation (C. Norris, unpubl. data), whereas *Stygobromus* and *Stygoparnus* are stygobionts and are found in one shallow well within 110 m from Comal Springs (subpopulation PA in Fig. 1b; Gibson, Harden & Fries, 2008). These slight differences in habitat use across the four taxa might affect their ranges of dispersal and thereby 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 S1). For example, the smaller taxa may be able to access small conduits unavailable to *Eurycea*.

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, Lucas *et al.* (2009) 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. Gonzalez and colleagues (unpubl. data) found no pattern of IBD among sampled populations of the riffle beetles *Heterelmis comalensis* and *H. glabra*, perhaps suggesting different dispersal capabilities of *Heterelmis* and *Eurycea*. Here, we take a comparative approach with four endemic taxa and examine gene flow on a smaller scale. 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 1 km² area in the Great Artesian Basin of central Australia (Robertson *et al.*, 2014). Notably, our study takes advantage of relatively recent advances in DNA technology, genotyping-by-sequencing, to sample thousands of markers across the genome to provide a more accurate representation of demographic history (i.e., different regions of DNA with different functional constraints might evolve at different rates; Patterson, 1999; Gompert *et al.*, 2006; Elshire *et al.*, 2011) and to identify genetic regions that underlie local adaptation which requires many markers (Knowles & Maddison, 2002).

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 approximate Bayesian computation (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 11 habitat patches 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 11 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; T. Gonzalez *et al.*, unpubl. data and Ethridge, Gibson & Nice, 2013, respectively). We used the DNeasy 96 Blood and Tissue Kit (QIAGEN Sciences, Germantown, MD, U.S.A.) to extract DNA from entire *Stygoparnus* individuals.

We followed the protocol described by Gompert *et al.* (2012) and Parchman *et al.* (2012) for preparing and analysing reduced genomic complexity libraries for each individual (genotyping-by-sequencing); 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 labelled the fragments of each individual with barcodes (individual identification sequences) to allow for multiplexing hundreds of individuals in one sequencing lane (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, U.S.A.) and assessed library quality and concentration with a NanoDrop spectrophotometer (NanoDrop products, Wilmington, DE, U.S.A.) and a Bioanalyzer (Agilent, Inc., Santa Clara, CA, U.S.A.). 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. 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 min, 60 °C for 2 min and 72 °C for 10 min) to ensure that the PCR product would be dominated by double-stranded frag-

	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			

Table 1 Number of individuals sampled from each subpopulation of four sympatric animal taxa in Comal Springs. PA = Panther Canyon Well, R1 = Spring Run 1, R2 = Spring Run 2, R3 = Spring Run 3, R4 = Spring Run 4, R5 = Spring Run 5, KP = Kiddie Pool, UP = Upwelling, WS = West Shore, SI = Spring Island, BW = Backwater Springs.

ments. 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. The filtered sequences were 86–92 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, U.S.A.) to perform a *de novo* assembly. We found 6204 contigs in the *Eurycea* dataset, 494 contigs in the *Heterelmis* dataset, 4980 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, U.S.A.). Refer to the Supporting Information for assembly details. Each sequence within a contig is referred to as a haplotype. 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 SNP, we used custom Perl scripts in conjunction with samtools and bcftools (Li *et al.*, 2009). We have high confidence in the SNPs identified using strict criteria (see Supporting Information), but we have likely failed to identify those SNPs with rare alleles. We found 7035 SNPs in the *Eurycea* dataset, 545 SNPs in the *Heterelmis* dataset, 5432 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). Some contigs had multiple SNPs, though the average number of SNPs per contig is close to 1 (the mean across taxa is 1.04). Our downstream ABC-based analyses model haplotypes explicitly, including linkage disequilibrium among SNPs within a contig, and thus our migration rate estimates and model comparisons take linkage among SNPs in the same contig into account. Our estimates of F_{ST} do not explicitly account for linkage disequilibrium. However, since having few tightly linked SNPs would only affect uncertainty of

parameter estimates not the point estimates, our other downstream analyses (e.g., NMDS) should be unaffected by linkage among SNPs.

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; see Supporting Information). 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 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 visualise subpopulation differentiation using genome-average pairwise F_{ST}s. NMDS does not force bifurcating relationships among subpopulations but collapses the data from multiple subpopulations into a few dimensions. 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. The number of dimensions chosen minimises the stress (the disagreement between the new, collapsed configuration and the predicted values from the regression). 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 randomisations.

We used approximate Bayesian computation (ABC, Beaumont, Zhang & Balding, 2002; Nielsen & Beaumont, 2009; Bertorelle, Benazzo & Mona, 2010; Csilléry *et al.*, 2010) to test which models (hypotheses) of gene flow best explained patterns of genetic variation in the data. We tested five competing models: (i) no gene flow among subpopulations; (ii) an island model with equal or constant gene flow among all subpopulations; (iii) a stepping stone model with unidirectional gene flow along surface stream flow (Fig. 1b); (iv) a stepping stone model with bidirectional gene flow with and against surface stream flow (Fig. 1c) and (v) gene flow among subpopulations fed by the same groundwater sources (Fig. 1d). 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 a consistently higher temperature by 0.5 °C. We cannot calculate the likelihood of each hypothesised gene flow model (i.e. the probability of obtaining the data given the model and specific parameter values), so we approximate the model likelihood using ABC (Nielsen & Beaumont, 2009). First, we specified the data for ABC. We identified our 86–92 bp contigs 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 that there is no recombination within a contig but free recombination between contigs. 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 contig datasets for ABC; instead, we used a subset of our full datasets to include one contig per individual for higher coverage contigs. Specifically, in the *Eurycea* dataset, we identified 475 variable contigs with data for 10 or more individuals in each sampled subpopulation. We used the 174 variable contigs with data for five or more *Heterelmis* individuals in each sampled subpopulation, and the 496 variable contigs with data for five or more *Stygobromus* individuals in each sampled subpopulation. There were 61 180 variable contigs with data for five individuals per subpopulation in the *Stygoparnus* dataset, which was too many to run ABC practically, so we randomly sampled 500 of the 61 180 contigs using R and a custom Perl script.

Second, we used a custom Perl script and the software ms (Hudson, 2002) to simulate demographic histories (Fig. 1a) and calculate the corresponding summary statistics. We simulated the genealogy at each contig, 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 (Fig. 1a). 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; see Supporting Information for details): (i) the five gene flow models; (ii) effective population size, N_e; (iii) migration rate (*m*); (iv) time since divergence, τ; (v) mutation rate per fragment, μ; and (vi) growth rate, *g*. We also estimated mean Θ, ancestral Θ, subpopulation Θ, and the number of migrants per generation per subpopulation ($4N_e^{\text{anc}}m$). 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 that the observed summary statistics were similar to a large number of simulated summary statistics. We simulated data for 11 subpopulations for each taxon to include all potential subpopulations in Comal Springs (Fig. 1b–d). We used the same sample sizes of the observed datasets.

We then calculated the mean, variance, and skew of five haplotype-based summary statistics that describe genetic diversity within each subpopulation: expected heterozygosity (2pq), the average number of nucleotide differences between pairs of haplotypes in the sample (π, Tajima, 1983), the number of segregating or polymorphic sites within a genetic locus (i.e., contig) (*S*, Watterson, 1975), the number of private haplotypes (i.e., unique haplotypes in one subpopulation and no other) and the proportion of contigs in which the rarer haplotype 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 alleles in a growing subpopulation and an excess of moderate-frequency alleles 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.

A key to successful application of these ABC methods is how well the summary statistics capture the relevant properties of the data (Nielsen & Beaumont, 2009). After running approximately 20% of the total number of simulations for each dataset, we ran diagnostic tests to ensure: (i) parameters were correlated with summary statistics and (ii) 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 generalised 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

There is a long history of the study of geographic patterns of genetic variation, and some of the earliest tests of selection on genetic markers were based on identifying loci that showed extreme allele frequency differences among populations (Lewontin & Krakauer, 1973; Coop et al., 2010). Specifically, we (i) chose highly differentiated SNPs based on F_{ST}; (ii) examined whether these SNPs exhibit patterns of IBD to control for IBD when testing for local adaptation; and (iii) tested for correlations between genetic and environmental differences for these SNPs, which would support the hypothesis of local adaptation. Due to our limited sampling of *Eurycea*

subpopulations, the *Eurycea* dataset is not included in this analysis. We identified SNPs with 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 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 nine environmental variables recorded from spring openings in most of the subpopulations in our datasets available to us (C. Norris et al., unpub. data; Table 3). We took the median of each variable for each subpopulation. We realise 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 visualise 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 (i.e., the subpopulations for which we had both genetic and environmental data). 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 (Fig. 2). *Eurycea* subpopulation Θs ranged from 0.26 to 0.29. Subpopulation Θs ranged from 0.26 to 0.35 and 0.33 to 0.45 for *Heterelmis* and *Stygobromus*, respectively. *Stygoparnus* subpopulation Θs had the widest range,
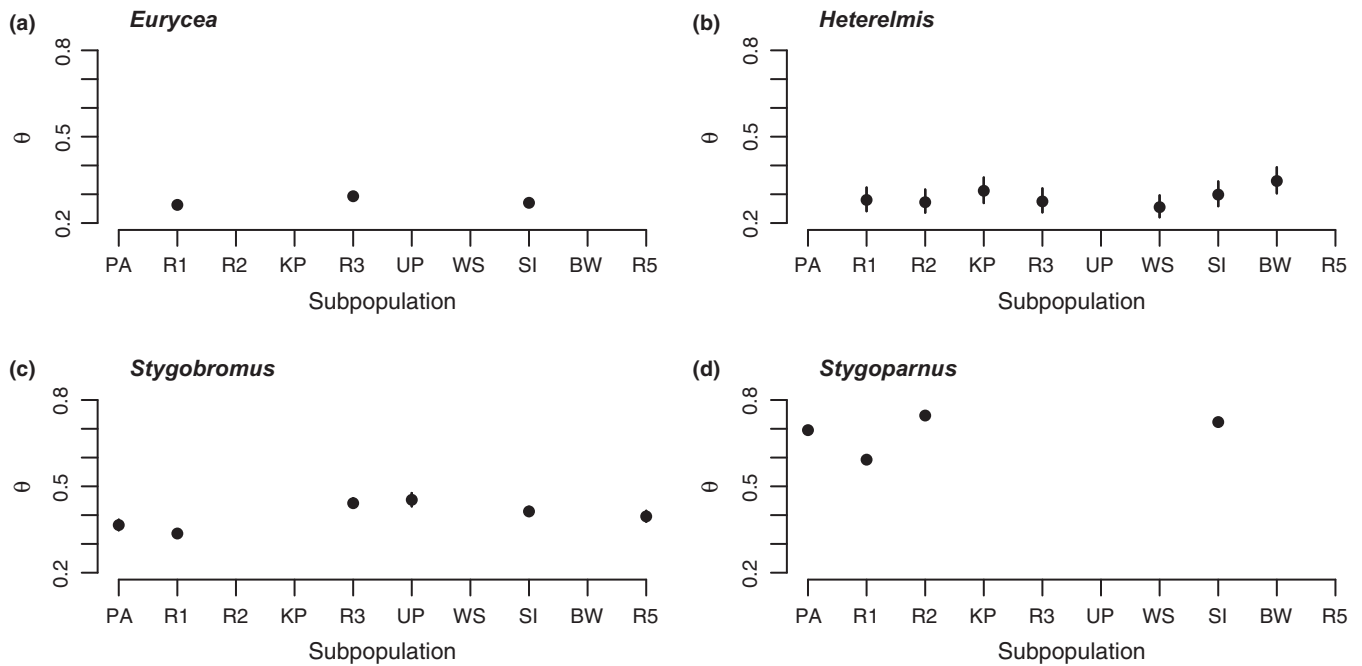
Fig. 2 Θ, an estimate of genetic diversity based on the allele frequency distribution, across subpopulations within each taxon. Dots are point estimates and lines are 95% credible intervals.

from 0.59 to 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 S2). 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 to 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 (Fig. 3). *Eurycea* subpopulations were nearly equally differentiated. *Heterelmis* subpopulations displayed a correspondence between genetic diversity and geographic space. Conversely, *Stygobromus* and *Stygoparnus* subpopulations were somewhat differentiated, but the structure did not correspond to geographic distance. 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 S3). In all datasets, summary statistics were correlated with one another to various degrees (Figure S4). 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 (Fig. 4). 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 (95% CI: 0.205–0.877); and *m* was 0.343 (95% CI: 0.025–0.825) and 0.152 (95% CI: 0–
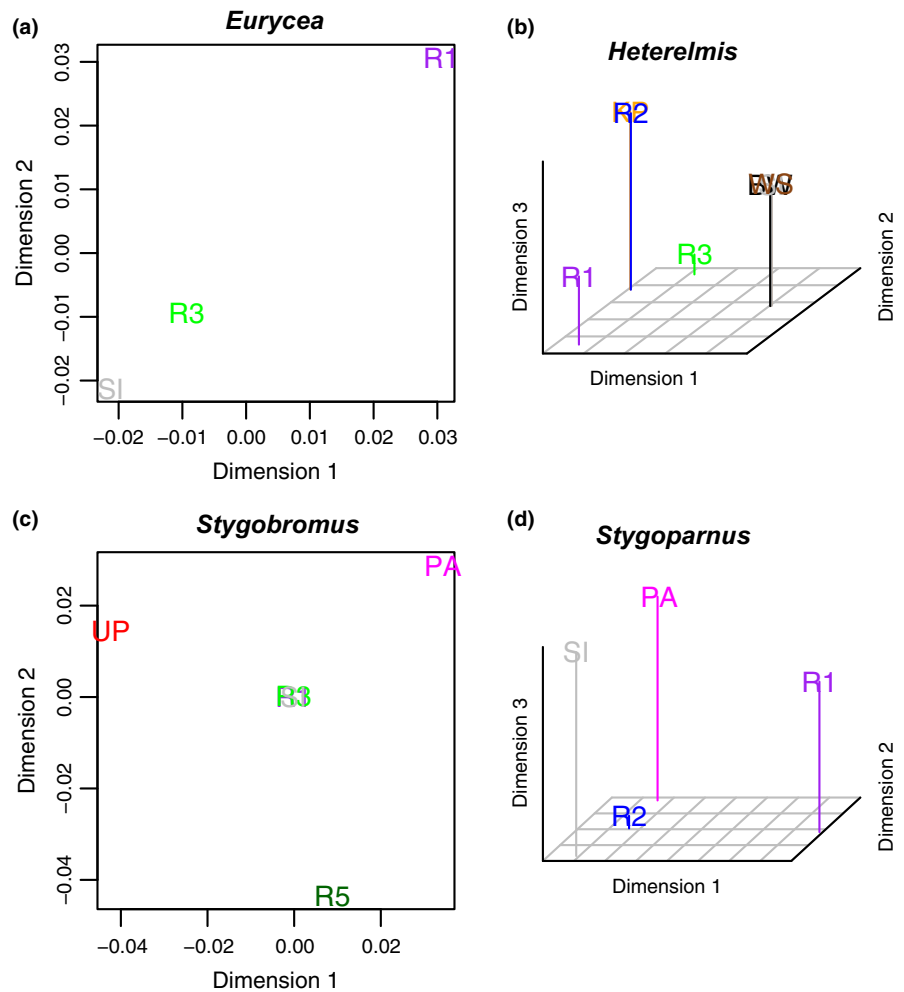
Fig. 3 Non-metric multidimensional scaling (NMDS) using genome-average pairwise F_{ST}s shows subpopulation differentiation within each taxon. Note in pane (b): R2 and KP overlap as do BW, SI, WS. Note in pane (c): R1, R3, SI overlap.

Table 2 ABC posterior probabilities for each hypothesised 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

0.768) for *Stygobromus* and *Stygoparnus*, respectively (Fig. 4).

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^{-1}, 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 total dissolved solides (TDS); dissolved oxygen (DO) and substrate size were strongly negatively

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 (Table S3b). 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 S5).

After performing the Mantel tests and partial Mantel tests, we asked if the number of statistically significant
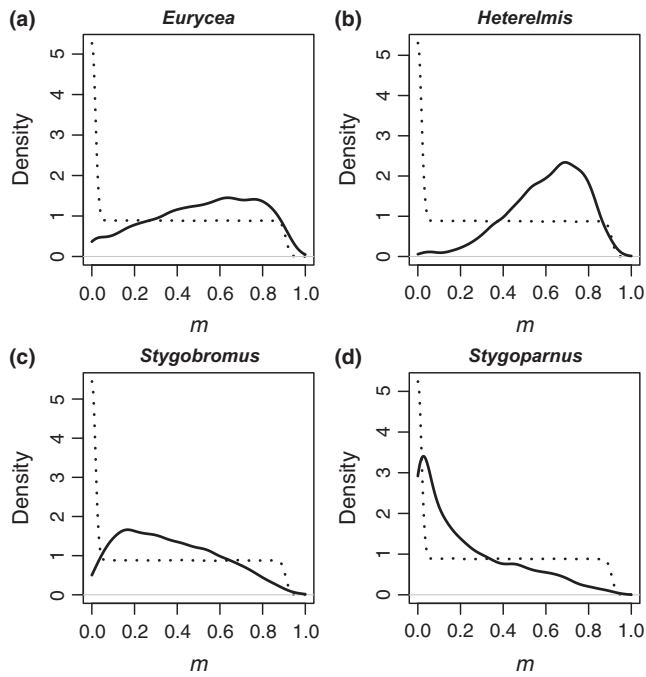
Fig. 4 Model-averaged ABC prior (dotted line) and posterior (solid line) probability distributions for migration rate (*m*) for each taxon.

correlations between highly differentiated SNPs and geographic distance or environmental PC scores, respectively, within a taxon was more than we would expect by change (5% cutoff rate). 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 S3a 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. 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 S3b 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 an 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)

Table 3 Median values for each environmental variable for each subpopulation. Primary and secondary substrate size measurements are based on the Wentworth scale (Wentworth, 1922).

Subpop.	Elevation (m)	Max. depth (m)	Temp. (°C)	pH	DO (mg L^{-1})	Sp. Cond. (µS cm^{-1})	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

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 63% (posterior means) of subpopulations made up of new migrants per generation. Generation times are unknown for the invertebrates in this study, but the generation time for these *Eurycea* is approximately a year, such that 55% of *Eurycea* subpopulations are made up of new migrants per year. For comparison, Robertson *et al.* (2014) studied genetic structure and gene flow among populations of the desert spring amphipod on a similar spatial scale as ours and found relatively lower pairwise F_{ST}s, 0.02, and higher m, 95%. In our study, the stygobionts, *Stygobromus* and *Stygoparnus*, had relatively lower migration rate estimates, perhaps because their habitats are inherently more isolated. However, these levels of gene flow are enough to prevent complete 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.

All phylogeographic studies use genetic data, which reflect evolutionary histories, with an emphasis on recent evolutionary processes or demographic events on a time scale of $2N_e$ generations (Gillespie, 2010). As such, 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 channelised 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 might occasionally experience human-mediated gene flow (e.g., throwing rocks, children using aquarium nets).

We found at least one SNP in each dataset examined 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). Genetic differentiation reflects a quantitative balance between divergent selection and gene flow (Gillespie, 2010), and with higher gene flow, stronger selection is required for populations to diverge. With that said, even with modest selection, there would likely be small but non-zero F_{ST} for selected loci (i.e., no fixed differences, but some low level of genetic differentiation and thus local adaptation would be possible). We suspect this is the case here, but the details depend on the strength of selection, which we do not know. It is important to note that these SNPs we identified 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 even more genetic markers and more subpopulations represented in the 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 is important for the long-term persistence of an ESU. As 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 or divergence times (Allendorf & Luikart, 2009), here we use both F_{ST} and patterns of gene flow. There is little genetic structure within Comal Springs and considerable

gene flow among subpopulations. Despite the high levels of gene flow, there is also evidence that is consistent with the hypothesis that there is some local adaptation to specific habitat patches within 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 localised extinctions may be naturally recolonised 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 these threats, captive propagation programmes have been established for the three endangered invertebrate taxa in this study. 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 (Fig. 3) 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 &

Clark, 1997). We hope these results will inform management strategies to maintain genetic variation in these spring-endemic taxa at Comal Springs.

References

Allendorf F.W. & Luikart G. (2009) *Conservation and the Genetics of Populations.* John Wiley & Sons, Hoboken.

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.

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

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

Bilton D.T., Freeland J.R. & Okamura B. (2001) Dispersal in freshwater invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.

Brune G.M. (2002) *Springs of Texas*, vol. **2**. Texas A&M University Press, College Station.

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

Chippindale P.T., Price A.H., Wiens J.J. & Hillis D.M. (2000) Phylogenetic relationships and systematic revision of cen-

tral Texas hemidactyliine plethodontid salamanders. *Herpetological Monographs*, **14**, 1–80.

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

Csilléry K., Blum M.G., Gaggiotti O.E. & François O. (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S. *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Ethridge J.Z., Gibson J.R. & Nice C.C. (2013) Cryptic diversity within and amongst spring-associated *Stygobromus* amphipods (Amphipoda: Crangonyctidae). *Zoological Journal of the Linnean Society*, **167**, 227–242.

Fries J.N. (2002) Upwelling flow velocity preferences of captive adult San Marcos salamanders. *North American Journal of Aquaculture*, **64**, 113–116.

Gibson J.R., Harden S.J. & Fries J.N. (2008) Survey and distribution of invertebrates from selected springs of the Edwards Aquifer in Comal and Hays counties, Texas. *The Southwestern Naturalist*, **53**, 74–84.

Gillespie J.H. (2010) *Population Genetics: A Concise Guide*. The Johns Hopkins University Press, Baltimore.

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 two 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.

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

Hartl D.L. & 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.

Holt R.D. & Gomulkiewicz R. (1997) How does immigration influence local adaptation? A reexamination of a familiar paradigm. *The American Naturalist*, **149**, 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.

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 University Press, Princeton.

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

Lewontin R. & Krakauer J. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N. *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.K., Gompert Z., Ott J.R. & Nice C.C. (2009) Geographic and genetic isolation in spring-associated *Eurycea* salamanders endemic to the Edwards Plateau region of Texas. *Conservation Genetics*, **10**, 1309–1319.

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**, t5448.

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 of the United States of America*, **70**, 3321–3323.

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

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., Gompert Z., Mudge J., Schilkey F.D., Benkman C.W. & Buerkle C. (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

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

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.

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*, **16**, 393–430.

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

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

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

Wakeley J. (2009) *Coalescent Theory: An Introduction*, Roberts & Company Publishers Greenwood Village, Colorado.

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

Wentworth C.K. (1922) A scale of grade and class terms for clastic sediments. *The Journal of Geology*, **30**, 377–392.

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

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

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Methods. Additional details about molecular and statistical methods.

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. Photographs and relative sizes of the four spring-endemic taxa in this study.

Figure S2. Distribution of F_{ST}s across SNPs for each pair of subpopulations per taxon.

Figure S3. Patterns of correlations of ABC parameters with ABC summary statistics for each taxon.

Figure S4. Patterns of correlations between all pairs of ABC summary statistics for each taxon.

Figure S5. Environmental similarities among subpopulations based on PCA.

(Manuscript accepted 1 March 2016)