QUANTIFICATION OF ENDOCRINE DISRUPTING CHEMICALS IN CENTRAL TEXAS RIVERS AND ASSOCIATIONS WITH GENOMIC VARIATION IN RED SHINERS

by

Alexis Vega Guzman, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Population and Conservation Biology
August 2022

Committee Members:

Chris C. Nice, Chair

Caitlin R. Gabor

David Rodriguez

James R. Ott

Timothy H. Bonner

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

Pursuing this Master's degree would not have been possible without the support from many people in my life and those I have met over this insane two-year journey.

First, I would like to thank my main advisor, Dr. Chris C. Nice for all the support, patience, and encouragement. Without his guidance and involvement throughout this process, none of this work would have been possible. I am truly thankful for Chris' mentorship and dedication in helping me build my confidence, think independently, and drive for success in both my academic and personal life. Chris has provided me with life-long knowledge I know I will continue to use in the next chapter of my life. I would also like to thank the rest of my committee, Drs. Caitlin R. Gabor, David Rodriguez, James R Ott, and Timothy H. Bonner.

Additionally, this endeavor would not have been possible without the generous support from my far away friends and Texas State alumni, Zach Forsburg and Kate Bell. Zach and Kate have been my life-support and sometimes virtual unpaid therapists when times were rough. I am forever grateful for the extended support you both have given me, even with your busy lives. I have learned so much from you both and value our friendships deeply.

Thanks also go to all the current and past graduate students of the Martin, Nice, and Ott lab socials, for their moral support and always great conversations. Cheers!

Last but not least, I would like to thank my family, and especially my dad for keeping me motivated and cheering me on during this process. I would also like to thank my cat Coochie for keeping me company and emotional support during all the late nights.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

Over the past decade, chemical pollution has increased in freshwater systems, including increases in endocrine disrupting compounds. One of the most prevalent of these chemical pollutants is 17α-ethinylestradiol (EE2), a synthetic estrogen used for contraceptives. Exposure to EE2 under laboratory conditions has been demonstrated to alter reproduction, development, and behavior of aquatic organisms. Most water quality assessments, however, have detection thresholds that are substantially higher than most observed concentrations of EE2. Furthermore, detection thresholds are higher than biologically relevant concentrations of EE2 as understood from laboratory studies. In this study, I used a sensitive assay to quantify EE2 concentrations in water and red shiner fish, *Cyprinella lutrensis*, collected from five central Texas rivers. I detected EE2 in all rivers at concentrations that have been shown to cause adverse effects in aquatic organisms. The observed concentrations of EE2 in water and fish suggest that aquatic organisms in central Texas commonly experience non-trivial exposure to EE2. Additionally, given that EE2 can influence reproduction and disrupt development, EE2 exposure could represent a substantial selective pressure for aquatic organisms. I used a Genotype-Environment-Association (GEA) approach to ask whether genomic variation in red shiners was associated with variation in EE2 concentrations. For this, I generated 33,902 single nucleotide polymorphisms (SNPs) for 152 red shiners from 14 localities in five rivers. GEA analysis using Redundancy Analysis included EE2 concentrations in water and fish as predictors of genomic variation as well as other environmental

predictors of water quality and patterns of land use in watersheds. Variance partitioning revealed significant proportions of genomic variation explained by my predictors and complex interactions among them. My results indicate that EE2 is not substantially associated with genotypic variation but represents a significant contaminant in central Texas rivers.

USING GENOTYPE-ENVIRONMENT-ASSOCIATION ANALYSIS TO ASSESS THE IMPACT OF ENDOCRINE DISRUPTING COMPOUNDS ON RED SHINERS, *CYPRINELLA LUTRENSIS*, FROM CENTRAL TEXAS RIVERS

I. INTRODUCTION

Increasing chemical pollution of freshwater resources has become a global conservation concern (Tang et al., 2021). Of particular interest in this investigation are endocrine disrupting compounds (EDCs), a class of chemical pollutants that include natural hormones and synthetic compounds that have the ability to disrupt endocrine systems by mimicking, inhibiting, or disrupting hormonal pathways in aquatic organisms (Dodds, 1938; EPA, 2002; Bhandari et al., 2015). EDCs have been demonstrated to alter reproduction, development and physiology of fishes and other aquatic organisms in laboratory studies (Schultz et al., 2003; Guillette Jr and Edwards, 2008). However, while EDCs potentially represent a major challenge for aquatic conservation, to date few studies have quantified the influence of EDCs in natural systems. Here, we begin to explore the influence of EDCs on aquatic organisms by quantifying how an EDC affects a widespread fish from central Texas using the framework of Genotype-Environment-Association (GEA) analysis.

We focused on a specific EDC, 17α-ethinylestradiol (EE2) found in fish and waters of five central Texas rivers. EE2 is a synthetic estrogen used in contraceptives and can cause adverse physiological, developmental and reproductive effects in aquatic organisms even at low concentrations. For example, in rainbow trout (*Oncorhynchus mykiss*), short-term exposure to EE2 results in reduced fertility (Schultz et al., 2003;

Guillette Jr and Edwards, 2008) and the production of intersex individuals (Nash et al., 2004; Depiereux et al., 2014). In a previous study (Guzman and Nice, in prep), we reported EE2 concentrations in the water of five central Texas rivers that equaled or exceeded concentrations shown to cause adverse effects in laboratory experiments (Guzman and Nice, in prep) (Table 1). We also observed EE2 in detectable concentrations in tissues of the fish, red shiner, *Cyprinella lutrensis*, sampled at several sites within each of the five rivers. Here we use a Genotyping-By-Sequencing approach to generate genotypes from across the genomes of sampled red shiners. These multi-locus genotypes we then used in a Genotype-environment association (GEA) analysis. GEA analysis is used to address how populations of organisms may be adapted, or are adapting, to their environment by identifying genomic signatures associated with environmental variation (Savolainen et al., 2013; Menamo et al., 2021; Seetharam et al., 2021; Postolache et al., 2021). GEA identifies associations between single nucleotide polymorphisms (SNPs) and environmental parameters (Grummer et al., 2019), including EE2. Any such associations are interpreted as potential evidence of the influence of selection by the environmental variables incorporated in the model.

Empirical reviews have compared the results of GEA studies implemented using several alternative statistical methods and have found redundancy analysis (RDA) to be most powerful. When comparing distance-based redundancy analysis (GEA-dbRDA) with redundancy analysis of components (cRDA), GEA-dbRDA was found to be the most powerful and successful multivariate method in reducing rates of false positives by accounting for neutral processes shaping underlying population genetic structure (Forester et al., 2018). Thus, herein we use the GEA-dbRDA framework to address two

questions: 1) are concentrations of EE2 in river water correlated with patterns of genetic variation observed in red shiners, and, if so, 2) how much of the total genetic variance in red shiners is attributable to EE2 concentrations compared to other factors including spatial variables, and sample site specific variation in land cover, water temperature (°C), pH, dissolved oxygen (mg/L), specific conductance (uS/cm), and flow (ft^3/s), and potential sources of EE2 contamination such as wastewater treatment plants (count)? The answers to these questions provide a foundation for future investigations of the potential influence of EDCs on aquatic organisms.

II. MATERIALS AND METHODS

Fish Biology

The red shiner, *Cyprinella lutrensis* is a widely distributed and abundant primary freshwater fish found in the central US (Matthews and Hill, 1977; Marsh-Matthews and Matthews, 2000). Red shiners can tolerate a variety of harsh and fluctuating environmental conditions and have the ability colonize degraded habitats (Matthews and Hill, 1977; Marsh-Matthews et al., 2011). We chose to focus on red shiners in our study because of their tolerance to a variety of conditions, reasoning that evidence of EDC influences in a freshwater species might indicate that these influences extend to more specialized, or less tolerant, aquatic species.

Data Collection

Fish (n = 1-18/site) were collected from 14 sites as a part of a previous study quantifying EE2 concentrations in both water and fish across the San Antonio, Guadalupe, San Marcos, Colorado and San Gabriel rivers (Guzman and Nice, in prep) in

central Texas (Figure 1.). Fish were collected using seines and euthanized with MS-222 (Tricaine Methanesulfonate, Western Chemical, Inc.). We obtained mass (g) and standard length (SL, mm) of each individual using a digital balance and dial calipers, respectively, and placed each fish into individual ziplock bag. At the time of collection, we obtained a water sample and obtained additional measurements at each site including longitude, latitude, and altitude (alt; m). Fish and water samples were transported back to the laboratory on ice and stored at -20C for later EE2 quantification. We also retrieved historical water quality metrics: (water temperature (temp; °C), pH, dissolved oxygen (DO; mg/L), specific conductance (SpC; uS/cm), and flow discharge (ft^3/s) for years 2018-2021. These water quality data were obtained from Texas Commission on Environmental Quality (TCEQ) surface water reporting tool. Flow discharge values are obtained from water monitoring stations located within 33 km of sampling sites, however, given the sparseness in water quality metrics, the data used is information available for river segments where sites are located.

Potential sources of EE2

Riverine systems are highly influenced by surrounding landscapes and in-stream processes (Poole, 2002) and these processes might influence EE2 concentrations in waters. A recent review revealed land development to be one of the primary factors negatively impacting biodiversity and ecosystems (Elmqvist et al., 2013). Anthropogenic factors including land development can affect water quality and the intensity of the impact can vary over spatial and temporal scales (Poole, 2002; Vrebos et al., 2017). Texas is one of the leading states in rapid land development which has led to the largest increase in impervious cover in the U.S. over the past 35 years (Xian et al., 2011;

Lombardi et al., 2020). Impervious cover (IC) associated with land development has been shown to increase run-off and associated pollutant loads into waterways (Page et al., 2015). Therefore, we calculated the percent of undeveloped (%), developed (%), agriculture (%), and water (%) within subwatersheds (HUC12) for each site. To quantify IC near rivers and sampling sites, we obtained and imported maps of the United States and World Imagery from ESRI, Texas watersheds (HUC12; from USDA), and land cover raster data from the National Land Cover Database (NLCD, 2019) (http://www.mrlc.gov) into ArcMap 10.6.1 (ESRI). All layers were projected in Albers Conical Equal Area using NAD 1983 to match the NLCD raster and clipped into the shape of Texas before calculations. We calculated the percent of impervious cover within subwatersheds (HUC12) for each site. We chose to analyze impervious cover at the HUC12 level because data is readily available across the US conterminous and at the level water quality management decisions at made (EPA, 2017). To calculate the amount of impervious cover for each site, we reclassified land cover classes into four categories: water (%), undeveloped (%), developed (%), and agriculture (%). Using the tabulate area tool, reclassified land cover classes were calculated per subwatershed polygons were sites were located. This was calculated dividing area of reclassified land cover class by subwatershed area, multiplied by 100, and included these attributes in our analyses.

In addition, EE2 contamination of surface waters can occur through discharge from wastewater treatment plants (WWTPs) that are unable to completely remove many of these compounds (Llamas, 2015; Lesser et al., 2018). To examine whether some sites are more vulnerable to contamination from these discharges, we analyzed the number of surrounding WWTPs with active permits that can discharge contaminates. To locate

WWTPs near rivers and sampling sites, we obtained and imported maps of the United States and World Imagery from ESRI, Texas watersheds (HUC12; from USDA), and WTTP location data (https://www.esri.com). We counted the number of WWTPs within subwatersheds (HUC12) for each site and were included in analysis.

Sample preparation, extractions, and EE2 quantification

Fish and water samples were stored at -20°C until thawed for EE2 extraction and analyses following US Environmental Protection Agency (EPA, 2009), (Gabor et al., 2016), and manufacturers kit protocol (Abraxis Ecologiena, Tokiwa Chemical Industries, Tokyo, Japan). Due to the chemical function and structure of EE2, ELISAs have a higher probability of detecting the biological activity of this environmental pollutant and so were used.

Fish and water samples were extracted with C18 Single Phase Extraction (SPE) columns (SepPak Vac 3cc/500mg; Waters Inc., Milford, MA, United States), using vacuum manifold pressure. Sample pre-treatment using SPEs is a common analytical procedure applied to increase the sensitivity and thus, quantification and detection of EE2 as recommended by the kit manufacturer. SPE columns were primed with 4 ml of HPLC grade methanol, followed by 4 ml of distilled water. Whole fish samples were homogenized with 4 ml of distilled water and 4 ul HPLC grade acetic acid using a 10 ml borosilicate glass cell tissue homogenizer (WSF Industries, Inc., Tonawanda, NY, USA). We pulled homogenates and water samples through primed SPE columns. Fish samples were eluted with 4 ml of HPLC grade ethyl acetate whereas; water samples were eluted with 4ml of methanol. Elutions were collected in 13 x 10 mm borosilicate test tubes.

We dried fish and water sample elutants by placing sample tubes into a 37°C

water bath under nitrogen gas (Airgas., Austin, TX USA). Dried fish and water samples were resuspended in 900 ul distilled water, 100 ul 100% methanol to a total volume of 1000 ul. EE2 standards (0, 0.05, 0.15, 0.5, 3.0 ng/L; all containing 10% methanol) and resuspended samples were then pipetted in duplicates on MoAb-Coated 96-well EE2 ELISA plates (Abraxis Ecologiena; assay range of 8.2 pg/ml - 5000 pg/ml; sensitivity (80% B/Bo) of 30 pg/ml) containing 100 ul of antigen-enzyme conjugate solution. We read ELISA sample absorbances on a spectrophotometer at 450nm (BioTex 800XS), and then quantified EE2 concentrations in surface waters (ng/L) and whole fishes (ng/g).

We generated sample standard curves for each plate and calculated means, standard deviations (SDs), and coefficients of variation (CVs). EE2 analyses using ELISA methods are highly reproducible with a CV maximum of < 10% across samples (Gomes et al., 2021). As a precautionary measure, we processed all water samples and randomly fish number of fish per site on one plate to ensure EE2 concentrations were detectable in our sites and enough variation was present. For this plate, the intra-assay CV was 0.75%. Six other plates were used for the remaining fish samples, the intra-assay CV ranged from 3.07% to 5.66%, while the inter-assay CV was 4.44%. For all seven plates (first and fish only plates), the intra-assay CV ranged from 0.75% to 5.66% and the inter-assay CV was 3.82%.

III. MOLECULAR GENETIC METHODS

DNA sequencing and data collection

DNA was extracted from fin clips taken from 152 individual fish using QIAgen's DNeasy Blood and Tissue kit (QIAgen Inc.). We used the GBS protocol described by (Meyer and Kircher, 2010; Parchman et al., 2013; Gompert et al., 2014; Mandeville et al.,

2015) to generate a genomic library for each individual fish. Briefly, DNA samples were fragmented using two restriction enzymes (EcoRI and Msel) (New England Biolabs; NEB Inc), and a unique 8- to 10-bp barcoded adapter was ligated to the restriction fragments for each individual. Using iProof high fidelity polymerase (Biorad, Inc.), we performed two rounds of PCR on the restriction-ligation products and pooled the resulting individual libraries into a single multiplexed library which was then sequenced at the University of Texas at Austin Genomic Sequencing and Analysis Facility (GSAF; Austin, Texas, USA). The library was processed via Pippen Prep quantitative electrophoresis unit (Sage Science, Beverly, MA) with size selection of fragments between 300 to 400 base pair (bp) and then sequenced with a full run of Illumina Novaseq SR100 SP. The choice of a fragment size range between 300 and 400bp was based on previous GBS library preparation insert what organism using this range which successfully produced sequence data (V. A. Sotola, pers. Comm.).

We used BOWTIE version 1.1.2 to identify and exclude PhiX control sequences (Langmead and Salzberg, 2012). We then used a combination of custom scripts, BCFTOOLS and SAMTOOLS (Li et al., 2009) to process the sequence reads. We removed unique adapter and barcode sequences from the raw sequence reads, corrected up to one mutation within barcodes, and attached individual IDs to reads using a custom Perl script. There is no reference genome for *Cyprinella lutrensis*, therefore, we used a de novo assembly strategy in combination with a clustering strategy from the dDOCENT variant calling pipeline (Puritz et al., 2014). The dDOCENT pipeline identifies unique sequence reads for each individual and we retained sequences with at least four copies per individual and that were shared by at least four individuals. The assembly of these

reads were carried out using CD-hit (Li and Godzik, 2006; Fu et al., 2012) with an 80% similarity threshold. Consensus reads from the resulting contigs then served as scaffolds for a reference-based assembly of all reads using aln and samse algorithms from BWA 0.7.5a-r405 (Li and Durbin, 2009). Next, we used SAMTOOLS ver. 0.1.19 and BCFTOOLS ver. 0.1.19 to call Single Nucleotide Polymorphisms – SNPs) and to calculate Bayesian posterior probabilities that a site is variable (Li et al., 2009). We required 80% of individuals to have at least one read at a specific locus for that locus to be included in the dataset. If a contig had more than one SNP, one was randomly selected and used for subsequent analyses, to maximize independence among markers. For every individual, the resulting genotype likelihoods from BCFTOOLS was used to calculate allele frequencies and loci with minor allele frequency (MAF) less than 0.05 were excluded. Lastly, we used the VCFFILTER ver. 0.1.1.9 to filter individuals containing a maximum number < 20 of missing data, a base quality score < 30, and a mapping quality score < 30.

IV. COMPUTATIONAL METHODS

Genetic structure

We examined the genetic structure of red shiner populations distributed within and among the five stream systems in central Texas using Entropy (Gompert et al., 2014; Mandeville et al., 2015). Entropy is a hierarchical model that estimates an individual's ancestry to a user-determined number of clusters or populations using a Bayesian framework. The Entropy approach is analogous to Structure (Pritchard et al., 2000; Falush et al., 2003) but differs in that Entropy accounts for uncertainty in genotype estimates resulting from sequence depth, alignment variation, and genotyping errors

(Gompert et al., 2014). Entropy estimates posterior genotype probability distributions, population allele frequencies and admixture proportions. We ran models with numbers of clusters, k, ranging from k = 2-16. Two Markov Chain Monte Carlo (MCMC) simulations were run for each k with 100,000 iterations, sampling every 10th iteration and, excluding the first 5,000 iterations of each run. We used Gelman-Rubin diagnostic statistics (Gelman and Rubin, 1992) and effective sample sizes to check the MCMC chains for convergence and stabilization (using the CODA package in R) (Plummer et al., 2006). Posterior probability parameter estimates from model runs were averaged across chains and k's for genotype estimates and illustrated as means, medians, and 95% credible intervals. The estimates of mean posterior genotype probabilities were used to summarize the distribution of genetic variation among individuals through principal component analysis (PCA) using the *prcomp* function (VEGAN package in R). Patterns of population structure across sampled populations were visualized by plotting admixture proportions in barplots.

Genotype-environment association (GEA)

One caveat with GEA is the need to account for spatial autocorrelation of allele frequencies and environmental data. Ecological and population genetic studies have traditionally used Mantel tests (Mantel, 1967) to test for spatial autocorrelation. However, the Mantel test has been shown to be problematic because the test assumes a regular sampling design that is almost never achieved in real studies (Legendre et al., 2015). Comparative simulations performed by Legendre and Fortin (2010) and Legendre et al. (2015) examined the comparative power of the Mantel test and distance-based Moran's Eigenvector Mapping (dbMEM) functions, formally known as principal coordinates of

neighbor matrices (Borcard and Legendre, 2002; Borcard et al., 2004, 2018; Dray et al., 2006). Simulations demonstrated that the Mantel test assumptions of linearity and homoscedasticity are likely violated for non-regular (i.e. real) sampling designs (Legendre and Fortin, 2010; Legendre et al., 2015). In comparison, dbMEM analysis was found to be successful in its ability to detect signals of spatial structure in univariate and multivariate geographical datasets across varying spatial scales and with non-regular designs (Legendre et al., 2015), while controlling for spatial correlations in species-environment relationships (Dray et al., 2006; Legendre et al., 2015). While many studies on riverine systems have used dbMEMs to control for space (Cilleros et al., 2017; Zhukov et al., 2019; Ramos et al., 2021), these methods might not fully account for spatial autocorrelation in systems which possess inherent directionality and stream order such as flows in riverine systems. A modification of the MEM approach is Asymmetric eigenvector maps (AEMs), an eigen-based spatial filtering method, which considers geographical space asymmetrically and therefore may be better suited to account for directional processes found in rivers (Blanchet et al., 2008, 2011). AEMs may also be more effective at modeling spatial structure at different scales (Blanchet et al., 2008) and has been shown to outperform MEM on an empirical dataset under directional processes (Blanchet et al., 2011). Given the geographical asymmetry of riverine systems (flow direction is upstream-downstream), we controlled for spatial autocorrelation on allele frequencies associated with the geographic distribution of populations by using AEMs (Blanchet et al., 2008, 2011) to measure the proportion of genotypic variance explained by spatial autocorrelation among sampling localities.

Following Blanchet et al. (2008, 2011), we first constructed a connection

diagram, similar to a minimum spanning tree, to link sites to one another in an upstream-downstream (unidirectional) flow pattern within their corresponding river network. Using this connection diagram, we constructed a site-by-edge matrix, representing a sequence of edges connecting each site to the river network origin influencing it (i.e., if site A is connected to the river network origin, site A = 1, if not, site A = 0; Figure 2C). We built the site-by-edge matrix using the *aem.build.binary* function (ADESPATIAL package) on the site coordinates and edge possibilities identified from the connection diagram (site 0 in Figure 2B). We then used the site-by-edge matrix to generate weights based on the differences in distance between sites within each river using inverse squared distances. For this, we first created a matrix of Euclidean distances from the site geographical coordinates (as performed in MEMs) and then removed edges directly linked to the river network origin in the site-by-edge matrix. Next, we created a vector to determine the length of each edge (i.e., distance from site saa to sab) and calculated weights by dividing by the maximum length of the network, this resulting value was squared and then subtracted from 1. AEM eigenfunctions were constructed using the *aem* function (ADESPATIAL package) on the site-by-edges matrix with weighted vectors for each site. We also ran AEM without weights but there were no differences (results not shown).

To limit multicollinearity among environmental predictors we calculated variance inflation factors (VIF) (see Table 2 for predictors) before the GEA-RDA analyses. Following best practices we required a VIF threshold of < 10. Individual fish mass, fish [EE2], water [EE2], water (%), developed (%), altitude, pH, DO, specific conductance, flow, and WWTPs were scaled and were incorporated as predictors in GEA using RDA. For GEA-RDA with AEMs, we first collapsed the genotype probability matrix using the

vegdist function (in the VEGAN; (Oksanen et al., 2013)) into a pairwise Euclidean genetic distance matrix among individuals to reduce computational time. To determine whether any AEM axes were significantly associated with among-individual distances, a distance-based (dbRDA) RDA using the *dbrda* function (in the VEGAN package; (Legendre and Anderson, 1999)) was performed. When this model was found to be significant, a permutational ANOVA was then used to identify significant AEM axes. We then implemented forward step-wise model selection with 9999 permutations using the *ordiR2step* function. The *ordiR2step* function computes similar statistics as AIC model selection but is used for ordinations. To understand the explanatory power of the geographical variables, environmental variables, and [EE2] on genomic variation in red shiners (GEA-RDA), we performed variance partitioning using the *varpart* function with 9999 replicates in the VEGAN package (Oksanen et al., 2013)). Variance partitioning was performed for combinations of predictors, specifically: all spatial predictors, all environmental site variables (excluding EE2), and water [EE2] and fish [EE2]. Venn diagrams were used to visualize the partitioning of variance for each of these analyses.

V. RESULTS

Sequencing generated an average of 1.8 million sequence reads per individual. Following filtering, retained 33,902 SNPs which form the basis for inspection of population genetic structure and GEA-RDA for 152 individuals in our study.

Population genetic structure

Four groups were identified among all individuals based on PCA of individual genotypes and admixture proportions calculated from Entropy. Results from PCA using

individual genotype probabilities of red shiners revealed four major clusters: 1) Colorado, 2) San Gabriel, 3) Guadalupe and San Marcos, 4) San Antonio (Figure 3). Principal component (PC) axes I and II together explained 42.2% of total genetic variation and divided individuals into groups by river. The first PC axis divided individuals from eastern rivers (Colorado and San Gabriel) from individuals from western rivers (San Antonio, San Marcos and Guadalupe) and explained 39.7% of the genotypic variance, while the second PC axis explained 2.5% of the variation and further divided individuals from eastern rivers into distinct groups by river basin.

Admixture proportions were calculated in Entropy for k = 2 – 16. We did not find support for biologically meaningful clusters and model convergence was poor at and after k = 5, therefore, we present results for only k = 2–4 models. For k = 2, individuals were separated based on river basin (Figure 4A). The model at k = 2 also shows individuals in the Guadalupe and San Marcos rivers share ancestry with individuals from the adjacent San Antonio, Colorado, and San Gabriel rivers. At k=3, individuals from the Colorado, and San Gabriel rivers form a distinct group, but individuals from the Guadalupe and San Marcos rivers maintain a pattern of admixture (Figure 4B). At k = 4, San Gabriel river individuals are assigned to the fourth genetic cluster with Colorado river individuals sharing ancestry with the San Gabriel group (Figure 4C).

GEA-RDA

We identified three AEM spatial variables. Permutational ANOVA revealed that all three AEM variables (AEM axes 2, 5, and 10) were significantly associated with genomic variation in red shiners (Table 2). These AEM's were then included in the full RDA model. A forward step-wise model selection procedure was then implemented for

the dbRDA analysis selecting among all 10 environmental predictors and the three significant AEM variables. The most parsimonious model contained eight environmental variables and two AEM as predictors. This selected model explained 42.35% of the total genomic variation (Table 4; adjusted R2 = 0.4235). Permutational ANOVA (9999 replicates) performed on the model selected from forward step-wise model selection indicated that water temperature, DO, specific conductance, flow, water (%), developed (%), EE2 concentration in water, WWTP, and AEM 2 and 5 were significant predictors of among-individual genetic distances. Notably, EE2 concentration in water was identified as a significant predictor, but not EE2 concentrations in fish.

Variance partitioning was used to examine how much genotypic variance was attributable to each environmental site variables, water [EE2], and AEM spatial variables in red shiners. The full model explained 42.35% of the total genotypic variation. Of this, spatial predictors (AEM 2 and 5) explained only 1.2%, [EE2] in water explained 0.1%, and all other site variables (water temperature, DO, specific conductance, flow, water (%), developed (%), and WWTP) explained 31.3% of the total genomic variation. A total of 59.2% of the total genomic variation was unexplained (Figure 5B). Further, 3.5% of the genotypic variation could not be disentangled from site variables and EE2 concentrations in water. There was also 4.1% of genomic variation shared between (confounded with) spatial and site variables and EE2 concentrations. Given the 3.5% shared genotypic variance between site variables and EE2 concentrations in water, we also inspected the proportion of variance attributed to between individual site variables including EE2 concentrations in water as a site variable and EE2 concentrations in fish, while including spatial predictors.

While we noted that EE2 concentrations in fish were not significant predictors in explaining the among-individual genetic variation, we examined the associations between the eight significant predictors in the permutational ANOVA with EE2 concentrations in fish (Figure 5A). Variance partitioning with EE2 concentrations in fish did not explain any genomic variation and no variance was shared between EE2 concentrations in fish and space (AEM 2 and 5). Despite the lack of variance explained, there is similar and an increase in genomic variance attributed to space and site variables. Space explained similar genotypic variance at 1.2%, with an increase of shared variance of 4.8% for space and site variables. While site variables (water temperature, DO, specific conductance, flow, water (%), developed (%), WWTP, and EE2 concentrations in water) showed an increase in genomic variance from 31.3% to 34%.

VI. DISCUSSION

Quantifying the extent of genetic variation associated with chemical pollution is an important step toward understanding how organisms are responding or might respond to future chemical pollutant exposure. In a previous study (Guzman and Nice, in prep), we found concentrations of EE2 in red shiner whole body tissue and water samples that were equal to, or greater than, concentrations that have been shown to cause adverse effects in aquatic organisms. The current study used GEA-RDA analysis to investigate the potential influence of EE2 concentrations in natural populations. The GEA-RDA analysis explored possible associations between fish genotypic variation and predictor variables including EE2 concentrations (reported by (Guzman and Nice, in prep), other site environmental variables, and spatial variables across 14 red shiner populations

distributed among five central Texas rivers. GEA-RDA analysis using the final dbRDA model revealed that EE2 concentrations in water only accounted for only 0.1% of the total genotypic variation present in red shiners. The lack of a selective response to EE2 concentrations might suggest that the red shiner populations sampled lack heritable genomic variation to respond to selection induced by variable EE2 concentrations. Alternatively, our results could indicate that EE2 might not represent a selective force for red shiners. Since the *Cyprinella* genera is widespread and abundant, our results could imply that other species in this genera do not possess heritable genomic variation for response to EE2 or in the case that EE2 does not represent a selective force, other species of *Cyprinella* are potentially able to persistent in waterways with EE2 exposures. However, considering we found that out of the total 42.35% explained variance we were able to explain, a large portion of it, 31.3%, was attributable to eight site variables. This suggests that red shiners have experienced and responded to selective pressures from water temperature, DO, specific conductance, flow, water (%), developed (%), and WWTP. However, of these, all except water (%) were significant predictors. Because water temperature, DO, specific conductance, and flow are water quality variables that vary across different temporal and spatial scales, it is predictable that environmentally tolerant red shiners, possess the genomic variation to respond to different and changes in environmental conditions.

Overall, our results suggest that red shiners might lack the heritable genetic variation to respond to EE2, but not for other variables that drive among site variation. It is also worth noting that although EE2 concentrations did not show patterns that indicated it is a selective agent, there were proportions of genomic variance could not be

disentangled from other site variables and EE2 concentrations. Due to general fluctuations in water quality and limited knowledge on how EE2 and other environmental variables interact, patterns that might indicate selection might not be detected until significant exposure has occurred. Given that EE2 exposure has important consequences for aquatic organisms and given the potentially complex dynamics of EE2 in natural waterways, we feel more study of the consequences of EE2 exposure on genomic variation is warranted.

VII. CONCLUSION

This is the first population-level analysis of wild populations to explore whether chemical pollutant EE2 is a potential driver of selection. While our GEA-RDA analysis did not suggest EE2 to be a potential driver of selection in red shiners, this approach did allow us to effectively examine genomic patterns attributed to complex environmental variables. Furthermore, the observed patterns of among site variation in EE2 concentrations in this study indicate further research extending to other aquatic organisms, including those in the genus *Cyprinella*, that inhabit EE2 polluted waters is warranted. Studies on other aquatic organisms experiencing exposure to EE2 is crucial to understanding whether the lack of heritable for a response is common and thus represents a serious conservation challenge. Regardless, if other aquatic organisms are able to respond, the concentrations found across central Texas rivers itself justify emphasis on management efforts to protect freshwater systems.

Table 1. Collection site details. River, site code and coordinates of sites for water and fish collections are provided. Water and fish EE2 concentrations found at each site are reported.

River	Site	Longitude	Latitude	n	EE2 (ng/L)	n	Mean EE2 (ng/g)	SD (±)
San Antonio	SAA	29.3817	-98.4945	1	0.1358	16	0.054	0.096
San Antonio	SAB	28.9776	98.0100	1	0.2179	10	0.012	0.009
San Antonio	SAC	28.6626	-97.3887	1	0.14	11	0.014	0.029
Guadalupe	GDA	29.5358	-97.8810	1	0.2256	15	0.026	0.044
Guadalupe	GDB	29.4845	-97.4477	1	0.2081	16	0.009	0.013
Guadalupe	GDC	28.8239	-97.0305	1	0.1674	-	-	-
San Marcos	SMA	29.8571	-97.8970	1	0.1187	18	0.008	0.01
San Marcos	SMB	29.7525	-97.7811	1	0.1453	14	0.01	0.012
San Marcos	SMC	29.6686	-97.6996	1	0.1145	2	0.003	0.001
Colorado	COA	30.2092	-97.4997	1	0.1405	3	0.023	0.025
Colorado	COB	30.1120	-97.3251	1	0.1534	11	0.004	0.004
Colorado	COC	29.7052	-96.5354	1	0.1551	16	0.008	0.02
San Gabriel	SGA	30.6943	-97.2788	1	0.1923	1	0.025	-
San Gabriel	SGB	30.7020	-97.8775	1	0.1289	15	0.007	0.011
San Gabriel	SGC	30.6227	-97.8627	1	0.1253	4	0.015	0.02

Abbreviations: Site = Code to distinguish each site, EE2 ng/L = Observed EE2 concentration in water sample, Mean EE2 ng/g = Mean EE2 concentration found in fish per respective site, SD = Standard deviation of mean EE2 in fish per respective site, and n = sample size or number of samples per site.

Table 2. Sampling locality water quality measurements. Measurement values for temperature, DO, pH, specific conductance, and flow are historical data medians over the years 2018-2021.

Site	Alt	Seg ID	Seg size	Temp	DO	pH	SpC	Flow	Water	Undev	Dev	Ag	WWTP
SAA	173	1911	88.06	23.3	7.7	8	746	59.9	0.26	15.37	79.08	5.27	0
SAB	81.99	1901	155.34	24.4	7.9	8.2	1105	328	0.67	35.95	4.98	58.37	0
SAC	48	1901	155.34	24.4	7.9	8.2	1105	328	0.75	51.7	5.68	41.85	1
GDA	127.01	1804	101.27	20	8.1	8.2	713	358	0.76	55.89	6.68	36.65	1
GDB	75.98	1803	169.76	23.1	8.2	8.05	532	675	1.44	34.74	15.09	48.7	1
GDC	13.01	1803	169.76	23.1	8.2	8.05	532	675	3.51	20.72	27.57	48.18	5
SMA	160.99	1808	77.94	21.7	8.5	8	565	195	0.75	28	14.91	56.31	4
SMB	127.98	1808	77.94	21.7	8.5	8	565	195	0.53	32.27	6.2	60.98	1
SMC	106	1808	77.94	21.7	8.5	8	565	195	0.32	49.61	5.83	44.23	0
COA	112.77	1428	42.37	22.9	8.5	7.9	555	516	3.59	41.55	7.02	47.82	1
COB	112.47	1434	76.89	19.8	7.9	7.9	489	887	2.32	39.16	24.51	33.99	3
COC	53.34	1402	152.48	22.6	9.75	8.1	534	820	1.848	46.32	6.39	45.4	1
SGA	142.95	1214	32.79	19.6	7.7	7.9	439	4.45	0.15	16.36	4.1	79.37	0
SGB	377.95	1251	41.85	20	8.65	8	541	1.4	3.27	85.27	11.16	0.28	1
SGC	294.13	1250	41.34	20.5	10.1	8.1	569	5	0.27	75.94	23.55	0.22	1

Abbreviations: Alt = Altitude (m), Seg ID = Segment identification number for specific site, Seg size = size of river segment (miles), Temp. = Water temperature (°C), DO = Dissolved oxygen (mg/L), pH = pH (standard units), SpC = Specific conductance (uS/cm), Flow = flow discharge (ft^3/s) or volume (ft) of water moving down a stream or river per unit of time (second), Water = percentage of water within watershed of site, Undev. = percentage of undeveloped land cover within watershed, Dev. = percentage of developed land cover within watershed of site, Ag. = percentage of agricultural land cover within watershed of site, WWTP = number of active WTTPs within watershed of site.

Table 3. Spatial (AEM) variables identified by dbRDA analysis conducted on genetic distances. Significant AEMs were used in the final dbRDA.

	Df	Variance	F	Pr(>F)
AEM2	1	211	3.287	0.010 **
AEM5	1	146.1	2.275	0.040 *
AEM10	1	1072.3	16.708	0.001 ***

Significance levels indicated by asterisks: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 4. Significant environmental and spatial variables identified in the forward step-wise model selection on collapsed genetic distances. Table output shows significant environmental and spatial predictor variables in dbRDA analysis using permutational ANOVA. Water quality and landscape variable abbreviations can be found on Table 2.

	Df	Variance	F	Pr(>F)
Water	1	61.3	1.4304	0.1583
DO	1	118.5	2.763	0.0136 *
WWTP	1	451.1	3.507	0.0001 ***
Flow	1	159.3	3.716	0.0031 **
SpC	1	176.8	4.123	0.0001 ***
AEM2	1	121.3	2.828	0.0115 *
Dev	1	184.3	4.299	0.0007 ***
AEM5	1	141.7	3.304	0.0084 **
Temp	1	99.1	2.312	0.0357 *
EE2(ng/L)	1	92.6	2.16	0.0421 *

Significance levels indicated by asterisks: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.

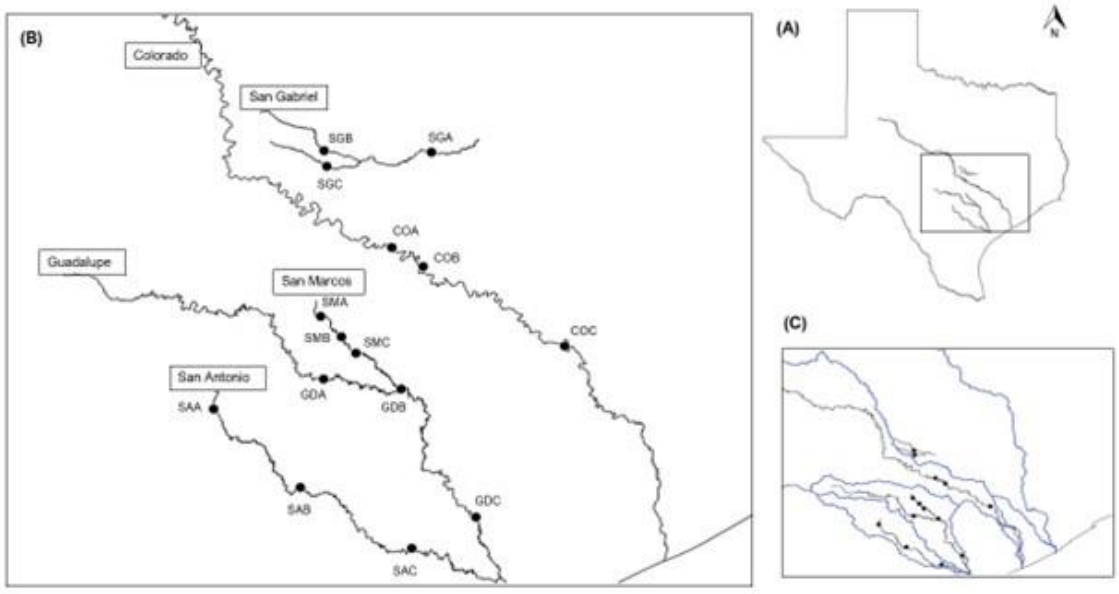

Figure 1. Water and red shiner sampling sites (n = 15). (A) River study region in Texas, (B) Name of river with site of red shiner, EE2, and water quality collections (Refer to Table 1 for coordinates). (C) River basin outlined in blue. The Guadalupe and San Marcos river are in the same river basin, Guadalupe river basin.

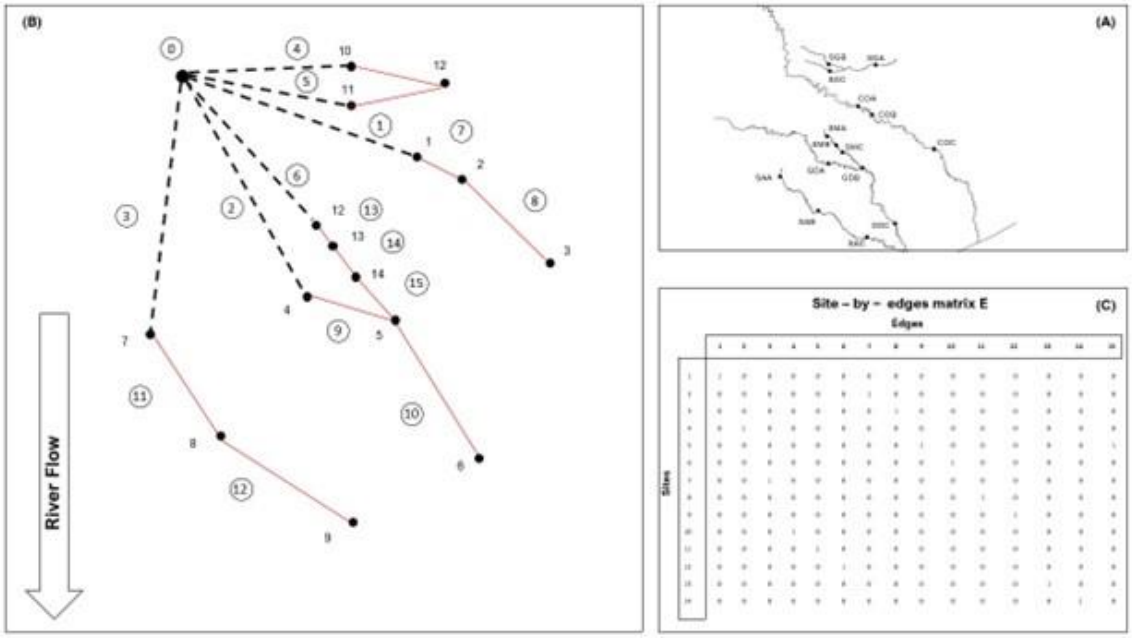

Figure 2. Schematic representations of AEM analysis of the 14 study sites. (A) Map of sampling sites for reference, (B) connection diagram showing connection of sites. Sites are numbered (see (A) for reference) and edges are numbered with enclosed circles. (C) sites-by-edges matrix E generated from connection diagram.

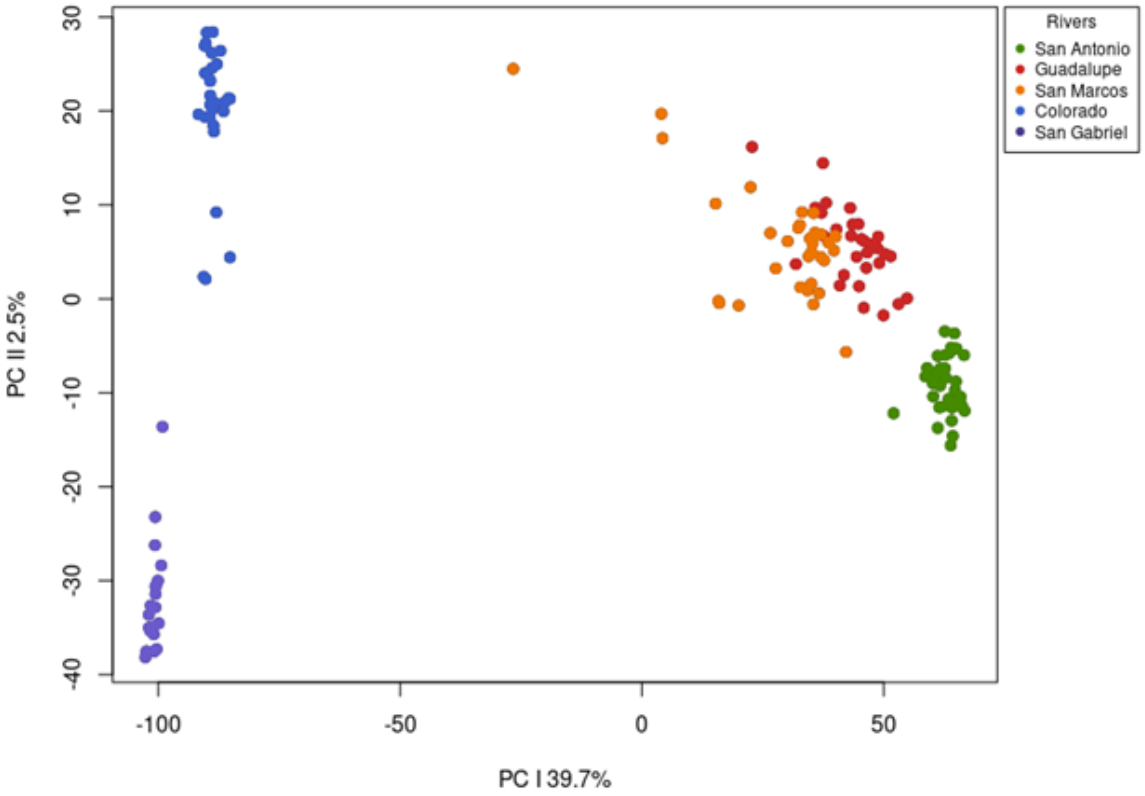

Figure 3. Principle component analysis of genetic differentiation of all individuals from each collection site. Each individual is represented by a dot. Individuals (represented as dots) are colored by river: San Antonio = green, Guadalupe = red, San Marcos = orange, Colorado = blue, and San Gabriel = purple (See figure legend).

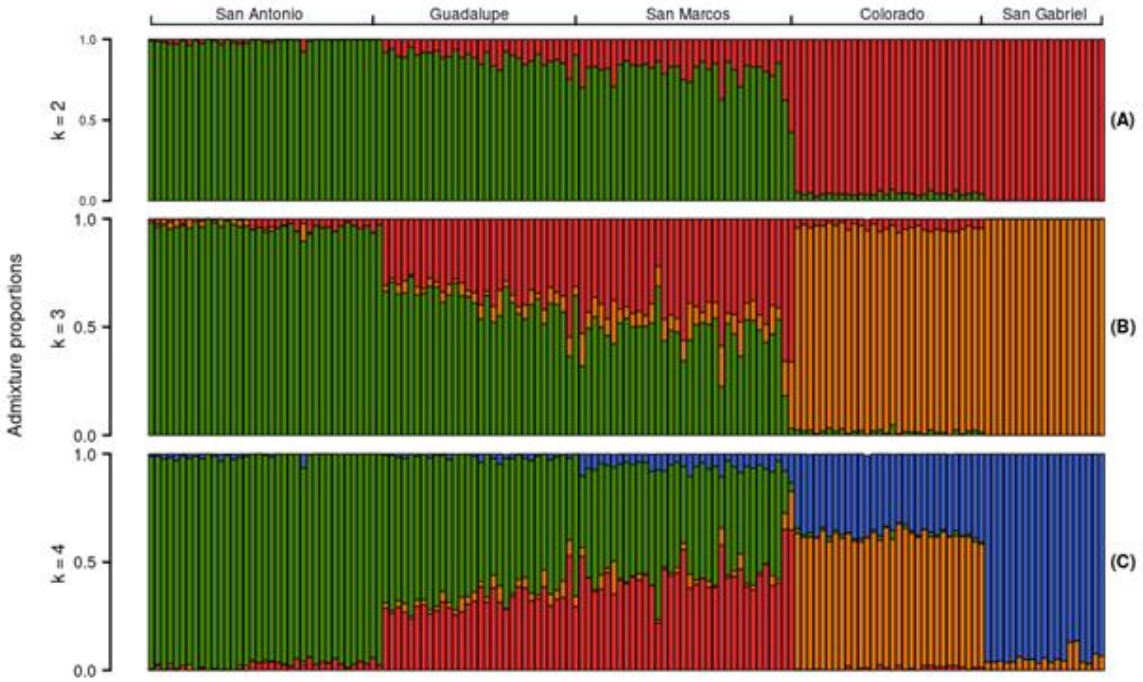

Figure 4. Admixture proportions estimated with Entropy for k = 2 (A), 3 (B), and 4 (C). Each bar represents an individual (red shiner) and colors represents the individuals genotype probability assignment to a group. Tick marks above (A) outline groups individuals by river.

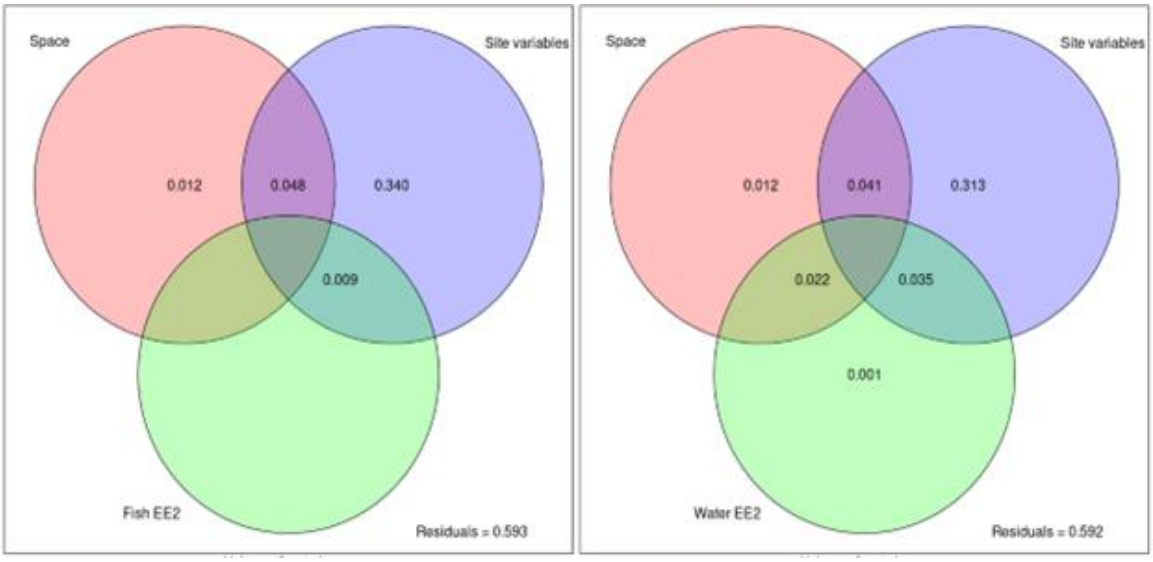

Figure 5. Venn diagrams of the variance partitioning based on the final dbRDA analysis. (A) represents the variance explained by spatial (AEMs) variables in red, site variables (see below plus EE2 in water) in blue, and EE2 in fish in green. (B) shows the variance explained by spatial (AEMs) variables, site variables, and EE2 in water. Explanatory components are: space = AEM2, AEM 5, and AEM 10, site variables = Water, DO, flow, specific conductance, developed, water temperature, and Water EE2 = EE2 in water found at sites where fish were collected (See Table 2 for abbreviations).

REFERENCES

Bhandari, R. K., Deem, S. L., Holliday, D. K., Jandegian, C. M., Kassotis, C. D., Nagel, S. C., Tillitt, D. E., Vom Saal, F. S., and Rosenfeld, C. S. (2015). Effects of the environmental estrogenic contaminants bisphenol A and 17α-ethinyl estradiol on sexual development and adult behaviors in aquatic wildlife species. General and comparative endocrinology, 214:195–219.

Blanchet, F. G., Legendre, P., and Borcard, D. (2008). Modelling directional spatial processes in ecological data. Ecological Modelling, 215(4):325–336.

Blanchet, F. G., Legendre, P., Maranger, R., Monti, D., and Pepin, P. (2011). Modelling the effect of directional spatial ecological processes at different scales. Oecologia, 166(2):357–368.

Borcard, D., Gillet, F., and Legendre, P. (2018). Spatial analysis of ecological data. In Numerical ecology with R, pages 299–367. Springer.

Borcard, D. and Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological modelling, 153(1-2):51–68.

Borcard, D., Legendre, P., Avois-Jacquet, C., and Tuomisto, H. (2004). Dissecting the spatial structure of ecological data at multiple scales. Ecology, 85(7):1826–1832.

Cilleros, K., Allard, L., Vigouroux, R., and Brosse, S. (2017). Disentangling spatial and environmental determinants of fish species richness and assemblage structure in neotropical rainforest streams. Freshwater Biology, 62(10):1707–1720.

Depiereux, S., Liagre, M., Danis, L., De Meulder, B., Depiereux, E., Segner, H., and Kestemont, P. (2014). Intersex occurrence in rainbow trout (*Oncorhynchus mykiss*) male fry chronically exposed to ethynylestradiol. PLoS One, 9(7):e98531.

Dodds, E. (1938). The significance of synthetic oestrogenic agents. Acta Medica Scandinavica, 95(S90):141–145.

Dray, S., Legendre, P., and Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (pcnm). Ecological modelling, 196(3-4):483–493.

Elmqvist, T., Fragkias, M., Goodness, J., Güneralp, B., Marcotullio, P. J., McDonald, R. I., Parnell, S., Schewenius, M., Sendstad, M., Seto, K. C., et al. (2013). Urbanization, biodiversity and ecosystem services: challenges and opportunities: a global assessment. Springer Nature.

EPA, U. (2002). Environmental protection agency.(2002). Registration eligibility science chapter for chlorpyrifos: Fate and Environmental Assessment.

EPA, U. E. P. A. (2009). USEPA (US Environmental Protection Agency). 2009. National lakes assessment 2007: A Collaborative Survey. epa/841/r-09/001.

EPA, U. E. P. A. (2017). USEPA (US Environmental Protection Agency). 2017. National lakes assessment 2012: Technical Report. epa841-r-16-114.

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164(4):1567–1587.

Forester, B. R., Lasky, J. R., Wagner, H. H., and Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate Genotype–Environment Associations. Molecular Ecology, 27(9):2215–2233.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). Cd-hit: accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23):3150–3152.

Gabor, C. R., Zabierek, K. C., Kim, D. S., da Barbiano, L. A., Mondelli, M. J., Bendik, N. F., and Davis, D. R. (2016). A non-invasive water-borne assay of stress hormones in aquatic salamanders. Copeia, 104(1):172–181.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7(4):457–472.

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., and Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Molecular Ecology, 23(18):4555–4573.

Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., and Taylor, E. B. (2019). Aquatic landscape genomics and environmental effects on genetic variation. Trends in Ecology & Evolution, 34(7):641–654.

Guillette Jr, L. J. and Edwards, T. M. (2008). Environmental influences on fertility: can we learn lessons from studies of wildlife? Fertility and Sterility, 89(2):e21–e24.

Guzman, A. and Nice, C. C. (2022). Endocrine disrupting chemicals detected in central Texas surface water and fish at biologically effective concentrations [Unpublished manuscript - in preparation].

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with BOWTIE 2. Nature Methods, 9(4):357–359.

Legendre, P. and Anderson, M. J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecological Monographs, 69(1):1–24.

Legendre, P. and Fortin, M.-J. (2010). Comparison of the mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources, 10(5):831–844.

Legendre, P., Fortin, M.-J., and Borcard, D. (2015). Should the mantel test be used in spatial analysis? Methods in Ecology and Evolution, 6(11):1239–1247.

Lesser, L. E., Mora, A., Moreau, C., Mahlknecht, J., Hernández-Antonio, A., Ramírez, A. I., and Barrios-Piña, H. (2018). Survey of 218 organic contaminants in groundwater derived from the world's largest untreated wastewater irrigation system: Mezquital Valley, Mexico. Chemosphere, 198:510–521.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics, 25(14):1754–1760.

Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13):1658–1659.

Llamas, M. A. (2015). Degradación de Disruptores Endocrinos de Estructura Fenólica en Procesos de Tratamiento de Aguas Residuales. PhD thesis, Universitat de València.

Lombardi, J. V., Perotto-Baldivieso, H. L., and Tewes, M. E. (2020). Land cover trends in south Texas (1987–2050): potential implications for wild felids. Remote Sensing, 12(4):659.

Mandeville, E. G., Parchman, T. L., McDonald, D. B., and Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of catostomus species. Molecular Ecology, 24(8):1856–1872.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Research, 27(2 Part 1):209–220.

Marsh-Matthews, E. and Matthews, W. J. (2000). Spatial variation in relative abundance of a widespread, numerically dominant fish species and its effect on fish assemblage structure. Oecologia, 125(2):283–292.

Marsh-Matthews, E., Matthews, W. J., and Franssen, N. R. (2011). Can a highly invasive species re-invade its native community? the paradox of the red shiner. Biological Invasions, 13(12):2911–2924.

Matthews, W. J. and Hill, L. G. (1977). Tolerance of the red shiner, *Notropis lutrensis* (*Cyprinidae*) to environmental parameters. The Southwestern Naturalist, pages 89–98.

Menamo, T., Kassahun, B., Borrell, A., Jordan, D., Tao, Y., Hunt, C., and Mace, E. (2021). Genetic diversity of ethiopian sorghum reveals signatures of climatic adaptation. Theoretical and Applied Genetics, 134(2):731–742.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols, 2010(6):pdb–prot5448.

Nash, J. P., Kime, D. E., Van der Ven, L. T., Wester, P. W., Brion, F., Maack, G., Stahlschmidt-Allner, P., and Tyler, C. R. (2004). Long-term exposure to environmental concentrations of the pharmaceutical ethynylestradiol causes reproductive failure in fish. Environmental Health Perspectives, 112(17):1725–1733.

NCLD, N.-U. (2019). National Land Cover Dataset (nlcd) 2019 land cover (CONUS).

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., Wagner, H., et al. (2013). Package '*vegan*'. Community Ecology Package, Version, 2(9):1–295.

Page, J. L., Winston, R. J., Mayes, D. B., Perrin, C., and Hunt III, W. F. (2015). Retrofitting with innovative stormwater control measures: Hydrologic mitigation of impervious cover in the municipal right-of-way. Journal of Hydrology, 527:923–932.

Parchman, T., Gompert, Z., Braun, M. J., Brumfield, R., McDonald, D., Uy, J., Zhang, G., Jarvis, E., Schlinger, B., and Buerkle, C. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology, 22(12):3304–3317.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. R News, 6(1):7–11.

Poole, G. C. (2002). Fluvial landscape ecology: addressing uniqueness within the river discontinuum. Freshwater Biology, 47(4):641–660.

Postolache, D., Oddou-Muratorio, S., Vajana, E., Bagnoli, F., Guichoux, E., Hampe, A., Le Provost, G., Lesur, I., Popescu, F., Scotti, I., et al. (2021). Genetic signatures of divergent selection in European beech (*Fagus sylvatica* l.) are associated with the variation in temperature and precipitation across its distribution range. Molecular Ecology, 30(20):5029–5047.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2):945–959.

Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. (2014). dDOCENT: a radseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ, 2:e431.

Ramos, E. A., de Morais-Junior, C. S., Rodrigues-Filho, C. A., Sánchez-Botero, J. I., Melo Júnior, M., and Novaes, J. L. (2021). Influence of spatial and environmental factors on the structure of a zooplankton metacommunity in an intermittent river. Aquatic Ecology, pages 1–11.

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics, 14(11):807–820.

Schultz, I. R., Skillman, A., Nicolas, J.-M., Cyr, D. G., and Nagler, J. J. (2003). Short-term exposure to 17α-ethynylestradiol decreases the fertility of sexually maturing male rainbow trout (*Oncorhynchus mykiss*). Environmental Toxicology and Chemistry: An International Journal, 22(6):1272–1280.

Seetharam, K., Kuchanur, P. H., Koirala, K., Tripathi, M. P., Patil, A., Sudarsanam, V., Das, R. R., Chaurasia, R., Pandey, K., Vemuri, H., et al. (2021). Genomic regions associated with heat stress tolerance in tropical maize (zea mays l.). Scientific Reports, 11(1):1–11.

Tang, Z., Liu, Z.-h., Wang, H., Dang, Z., and Liu, Y. (2021). A review of 17α-ethynylestradiol (ee2) in surface water across 32 countries: Sources, concentrations, and potential estrogenic effects. Journal of Environmental Management, 292:112804.

Vrebos, D., Beauchard, O., and Meire, P. (2017). The impact of land use and spatial mediated processes on the water quality in a river system. Science of the Total Environment, 601:365–373.

Xian, G. Z., Homer, C. G., Dewitz, J., Fry, J., Hossain, N., and Wickham, J. (2011). Change of impervious surface area between 2001 and 2006 in the conterminous United States. Photogrammetric Engineering and Remote Sensing, 77(8):758–762.

Zhukov, O. V., Bondarev, D. L., Yermak, Y. I., and Fedushko, M. P. (2019). Effects of temperature patterns on the spawining phenology and niche overlap of fish assemblages in the water bodies of the Dnipro river basin. Ecologica Montenegrina, 22:177–203.