7-1-2012

Reproductive phenology of fishes of the Rio Grande, New Mexico : a genes-to-community approach

Trevor James Krabbenhoft

Trevor J. Krabbenhoft
Candidate

Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Thomas F. Turner, Chairperson

Scott L. Collins

Charles Cunningham

Kathleen G. O'Malley

REPRODUCTIVE PHENOLOGY OF FISHES OF THE RIO GRANDE, NEW MEXICO: A GENES-TO-COMMUNITY APPROACH

by

TREVOR J. KRABBENHOFT

B.S., Zoology, North Dakota State University, 2003
M.S., Biology, University of South Carolina, 2006

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

Biology

The University of New Mexico
Albuquerque, New Mexico

July, 2012

DEDICATION

For my parents…a different kind of harvest.

ACKNOWLEDGEMENTS

I sincerely thank my major advisor, Tom Turner, for all of the help and guidance he offered on this dissertation research. I also greatly appreciate the time and insight given by my dissertation committee members: Scott Collins, Charlie Cunningham, and Kathleen O'Malley. This dissertation was greatly improved as a result of their input.

I wish to acknowledge my collaborator on Chapter 1, Steven Platania, for his support of the young-of-year fish sampling portion of this research. I also thank Rob Dudley, Howard Brandenburg, Mike Farrington, Jen Hester, Adam Barkalow and the rest of the crew at American Southwest Ichthyological Researchers for their assistance with Chapter 1 data collection. Thanks to Lex Snyder for curatorial assistance at Museum of Southwestern Biology and her support of this research. I greatly appreciate the logistical and curatorial support afforded by MSB Division of Fishes.

Next-generation sequencing and data analysis for Chapter 3 was conducted with the assistance and insight of Michelle Forys, George Rosenberg, Jennifer Kavka, Rob Miller and the Center for Evolutionary and Theoretical Immunology 454 Working Group. I am grateful for their contributions to Chapter 3.

I wish to thank members of the Turner Lab for their camaraderie and insightful discussions these past five years: Megan Osborne, Ayesha Burdett, Wade Wilson, Tom Kennedy, Tyler Pilger, Tracy Diver, Evan Carson and Jon Kawatachi. I am thankful for the friendship and mentoring of Tom Munroe, Steve and Yvonne Ross, and Mike Collyer during my graduate experience.

Finally, I wish to thank my wife and collaborator, Corey Krabbenhoft, who contributed to this research at all stages and made it a lot more fun along the way.

REPRODUCTIVE PHENOLOGY OF FISHES OF THE RIO GRANDE, NEW MEXICO: A GENES-TO-COMMUNITY APPROACH

by

TREVOR J. KRABBENHOFT

B.S., Zoology, North Dakota State University, 2003
M.S., Biology, University of South Carolina, 2006
Ph.D., Biology, University of New Mexico, 2012

ABSTRACT

Reproductive phenology is a key life history attribute of long lived organisms that can strongly affect reproductive success and, ultimately, drive community composition. Understanding the ecological causes and consequences and genetic mechanisms shaping reproductive timing is key to predicting the outcome of environmental change (e.g., climate change). The three chapters that comprised this dissertation were focused on elucidating the ecological and genetic underpinnings of reproductive timing in the fish community of an arid-land river, the Rio Grande, New Mexico.

In Chapter 1, we assessed reproductive phenology in the Rio Grande fish community with four years of young-of-year sampling data and spanning 16 years, from 1995 to 2010. Spawning data suggested that, in addition to known spatial habitat partitioning of resources, species also partition resources temporally by differential spawning periodicity. Inter-annual

variation in environmental conditions (e.g., river discharge) appears to drive community-level shifts in reproductive phenology. However, the magnitude of phenological shifts differed among species. We discuss these data in light of biotic-interactions among species and scenarios for future climate change.

In Chapter 2, we examined within- and among-species DNA sequence variation in a candidate gene, *Clock*, which is a key circadian rhythm gene that may shape reproductive phenology of fishes of the Rio Grande. Previous research has demonstrated a role for *Clock* in migratory and reproductive timing in disparate organisms, from songbirds to salmon. In this study, we tested whether patterns of allele length variation in *Clock* are consistent with: (1) among-species differences in reproductive timing, (2) phylogenetic inertia, or (3) functional-constraint in this key circadian gene. We present evidence that all three of these evolutionary processes may shape patterns of variation in *Clock* observed in Rio Grande fishes.

Finally, in Chapter 3 we broadened our search for genetic underpinnings of reproductive timing beyond *Clock*, to assess amino acid sequence variation in dozens of candidate genes among three species of Rio Grande cyprinid fishes: fathead minnow (*Pimephales promelas*), red shiner (*Cyprinella lutrensis*), and Rio Grande silvery minnow (*Hybognathus amarus*). We used next-generation 454 DNA sequencing to characterize the transcriptomes and explore sequence level variation among these three species. Using gene ontology analysis, we identified 86 candidate genes with functions potentially associated with reproductive timing and circadian rhythms. The 86 candidate genes had a total of 342 inferred amino acid substitutions among the three species, which could have functional implications and underlie some of the species-specificity in reproductive biology of these three species. Additionally, fifteen of the candidate

genes had simple sequence repeats in their inferred amino acid sequences, which might be targets of natural selection for shaping circadian rhythms and reproduction. The variation described in this study affords candidate loci for future comparative studies of reproductive timing.

TABLE OF CONTENTS

Dissertation Introduction

Extensive research has revealed that seasonal variation in spawning time is ubiquitous in freshwater fish communities (Matthews 1998). In these communities, both the onset and duration of spawning differ, particularly among reproductive guilds. For example, in temperate arid-land rivers in the southwestern United States and Australia, species can differ strongly in reproductive phenology, with two endpoints being: (1) high flow, spring spawners with short spawning seasons, and (2) low flow, summer spawners with protracted spawning seasons (Turner et al. 2010). Such differences are thought to arise from fishes timing key life history events to coincide with abiotic cues such as appropriate photoperiod, temperature, and discharge (Turner et al. 2010). Thus, inter-annual variation in environmental conditions has strong effects on adult fish community structure (Pease et al. 2006; Turner et al. 2010). High discharge years likely favor fishes with life histories tied to spawning at high flows, while low discharge years should favor low-flow spawners. In most freshwater fish species, the bulk of mortality occurs in young-of-year (YOY) fishes (Sogard 1997), and minor changes in survival rates of YOY fishes can have profound effects on abundances of adults of particular species. Consequently, there should be strong selective pressures aimed at matching the timing of presence of larvae with necessary food resources, while minimizing competition among species in a classic trade-off scenario (Pease et al. 2006; Turner et al 2010). In recruitment driven systems, such as arid-land rivers, a key component to bridging the gap between organism-environment interactions and community assembly is through elucidation of underlying (genetic) mechanisms used by fishes to determine timing of reproduction. In this dissertation,

we integrated comparative ecological and genomics approaches toward advancing our understanding of the mechanistic underpinnings of fish reproductive phenology in an arid-land river, the Rio Grande, New Mexico. The goal of this research is to connect fish sensory perception to genetic pathways for reproduction and community assembly. The Rio Grande is an ideal system in which to study comparative reproductive biology for four reasons: (1) the fish community is relatively simple with just twelve species comprising over 95% of the total number of individuals, facilitating confident morphological identification of fish larvae to species; (2) fishes in the Rio Grande utilize a wide breadth of life history strategies (Sublette et al. 1990), a requisite for comparative studies and adding to the generalizability of this study; (3) high levels of inter-annual variation in abiotic features can lend insight into how such variation affects fish reproduction and recruitment; and (4) study species are closely related to zebrafish, a model organism with a genome sequence available and for which many of these pathways have been elucidated.

This dissertation is comprised of three chapters. In the first chapter, we assessed reproductive phenology in the Rio Grande fish community with four years of young-of-year sampling data and spanning 16 years, from 1995 to 2010. The purpose of this study was twofold: (1) to provide high temporal resolution data on when Rio Grande fishes spawn, which is used as baseline data for mechanistic (genetic) studies of reproductive timing in Chapter 2, and (2) to understand how inter-annual environmental variation shapes reproductive timing in an arid-land river fish community. Spawning data suggested that in addition to known spatial habitat partitioning of resources, species also partition resources temporally by differential reproductive phenology. While rank order of spawning was generally consistent across years,

the absolute timing of spawning of each species differed across years. Date of first appearance of larvae was earlier for all species in 2008-2010 compared to 1995. Thus, these data revealed community-wide shifts in reproductive phenology among years. Shifts in spawning periodicity were not due to temperature differences, but were consistent with among-year shifts in magnitude and timing of spring flood pulse from snowmelt runoff. These results suggest that local temperature alone is insufficient to determine the effects of climate change and additional environmental variables, such as changes in river discharge, can also play an important role. Such knowledge of how and why organisms timing reproduction will help determine likely responses of fishes to future environmental disturbance (e.g., climate change). Additionally, when coupled with adaptive management practices, an understanding of the mechanisms that drive variation in spatial and temporal niche partitioning may assist restoration efforts of native fish communities in altered and regulated rivers.

In Chapter 2, we took a mechanistic approach aimed at elucidating the genetic underpinnings of among-species differences in reproductive timing revealed in chapter 1. Specifically, in this chapter we assessed DNA sequence-level variation in a key circadian rhythm gene, *Clock*, in Rio Grande minnows in relation to differences in reproductive timing among these species. *Clock* is a transcription factor and central component of the circadian machinery, and allele length variation in a poly-glutamate domain in *Clock* is correlated with latitude and timing of migration in salmon and songbirds (O'Malley and Banks 2008; Johnsen et al. 2007). We characterized the gene *Clock* in cyprinid fishes in the Rio Grande and tested three non-mutually exclusive hypotheses: (1) *Clock* is under functional constraint; (2) variation in *Clock* is due to phylogenetic non-independence of the study species (i.e., phylogenetic inertia), and (3)

Clock allele length variation corresponds to differences among species in reproductive timing. Data presented in this chapter suggests that all three hypotheses may play a role in shaping amino acid sequence variation in *Clock*.

While Chapter 2 focuses on whether reproductive phenology is shaped by allelic variation in a single gene, *Clock*, in Chapter 3 we adopted a multi-gene approach aimed at characterizing among-species variation in dozens of candidate genes putatively involved in reproduction and circadian rhythms. We employed next-generation DNA transcriptome sequencing and data mining to characterize protein-coding DNA sequence variation among three co-occurring species of Rio Grande cyprinid fishes with differences in life history strategies, reproductive biology, and circadian rhythms. We relied heavily on functional annotation of the zebrafish (*Danio rerio*) genome to characterize gene identities and ontologies. Amino acid variation was compiled and compared across genes and species as part of this exploratory study as a necessary first step toward understanding how natural selection has shaped the genomes of these three related, but ecologically quite different, species.

Together, the three chapters of this dissertation employ a genes-to-community approach toward understanding reproductive timing in the Rio Grande fish community.

Literature Cited

Johnsen A, Fidler AE, Kuhn S, Carter KL, Hoffmann A, Barr IR, Biard C, Charmantier A, Eens M, Korsten P, Siitari H, Tomiuk J, Kempenaers B (2007) Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies. Molecular Ecology 16(22):4867-4880.

Matthews WJ (1998) *Patterns in freshwater fish ecology*. Chapman and Hall Publishers.

O'Malley KG, Banks MA (2008) A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) *Clock* gene: evidence for selection on PolyQ length variants. Proceedings of the Royal Society of London, B 275:2813-2821.

Pease AA, Davis JJ, Edwards MS, Turner TF (2006) Habitat and resource use by larval and juvenile fishes in an arid-land river (Rio Grande, New Mexico). *Freshwater Biology* 51:475-486.

Sogard, S.M. (1997) Size-selective mortality in the juvenile stage of teleost fishes: a review. *Bulletin of Marine Science* 60(3):1129-1157.

Sublette JE, Hatch MD, Sublette M (1990) *The fishes of New Mexico*. University of New Mexico Press, Albuquerque.

Turner TF, Krabbenhoft TJ, Burdett AS (2010) Reproductive phenology and fish community structure in an arid-land river system. *In: Community Ecology of Stream Fishes* (Gido K, Jackson D, Eds.). *American Fisheries Society Symposium* 73:427-446.

Chapter 1: REDUCED FLOWS IN AN ARID-LAND RIVER DECREASE TEMPORAL NICHE PARTITIONING IN THE LARVAL FISH COMMUNITY.

Trevor J. Krabbenhoft, Steven P. Platania and Thomas F. Turner

Abstract

Knowledge of how inter-annual environmental variation affects timing of key life history events is essential for predicting the effects of climate change. In this study, we assessed reproductive phenology in an arid-land river fish community with four years of young-of-year sampling data and spanning 16 years, from 1995 to 2010. Spawning data suggested that in addition to known spatial habitat partitioning of resources, species also partition resources temporally by differential spawning periodicity. Additionally, while rank order of spawning was generally consistent across years, the absolute timing of spawning varied. Julian date of first appearance of larvae for each of the eight most abundant species was 4.3 – 28.1 days earlier for years 2008-2010 compared to 1995. Moreover, number of days spawning time advanced in 2008-2010 versus 1995 was greater for later spawning fishes, suggesting a reduction in temporal niche partitioning in reproductive phenology, and possible truncation of the community-wide spawning season. Given large resource overlap among larval fishes in this community, these findings may have important implications for the magnitude of inter-specific competition. Shifts in spawning periodicity were not due to among-year temperature differences, but were consistent with magnitude and duration of spring flood pulse from snowmelt runoff. These results suggest that local temperature alone is insufficient to determine the effects of climate change and additional environmental variables, such as changes in river discharge, can also play an important role. Additionally, these data suggest there is plasticity in reproductive timing, but beg the question of whether resource needs are met when reproductive phenology is altered. Knowledge of how and why organisms timing reproduction will help determine likely responses of fishes to future environmental disturbance

(e.g., climate change). Additionally, when coupled with adaptive management practices, an understanding of the mechanisms that drive variation in spatial and temporal niche partitioning may assist restoration efforts of native fish communities in altered and regulated rivers.

Introduction

As global climate change has garnered support over the past two decades (Intergovernmental Panel on Climate Change 2007), biologists have become increasingly interested in how such change drives shifts in timing of key life history events (Bradshaw & Holzapfel 2008; Parmesan 2006; Parmesan & Yohe 2003; Walther *et al.* 2002). For example, researchers have demonstrated climate-change driven phenological shifts in flowering time in angiosperms (Bradley *et al.* 1999; Fujisawa & Kobayashi 2010; Gordo & Sanz 2005), bird migrations and nesting (Bradley *et al.* 1999; Dunn & Winkler 1999; Gordo & Sanz 2006), insect diapause (Bradshaw *et al.* 2004; Tauber *et al.* 2007), fish spawning cycles (Gillet & Quétin 2006; Schneider *et al.* 2010) and anuran reproduction and tadpole emergence (Phillimore *et al.* 2010), among others.

Both theory and empirical data suggest that organisms utilize a wide array of environmental cues to time important life history events to maximize fitness. For example, many estuarine or coral reef fishes rely on lunar cycles to synchronize spawning to ensure that currents or tides are suitable for dispersal (Leatherland *et al.* 1992; Takemura *et al.* 2010) and spawning in many tropical freshwater fish is triggered by a decrease in pH and/or conductivity arising from arrival of the rainy season (Kirschbaum 1975, 1979). However, climate change affects some environmental cues differently than others, resulting in a possible decoupling of environmental variables. For example, local temperature and precipitation may increase under a warming climate, while photoperiod remains constant. In this scenario, organisms using temperature or precipitation as cues might have different phenological responses than organisms using photoperiod as a life history cue. Such a decoupling is of potential concern

because in some cases, gene expression is co-regulated by multiple environmental inputs. For example, expression of *Period4*, a gene important in circadian rhythms in zebrafish, is co-regulated by temperature and photoperiod (Lahiri *et al.* 2005). It is not known how expression of such genes will change if only one of the variables is altered, or what will be the likely phenotypic effects of such change. The potential effects these differential changes could have on organism-environment interactions is currently an area of intensive study.

In addition to affecting interactions between organisms and their physical environment, climate-driven changes in phenology are potentially important in altering biotic interactions (Yang & Rudolf 2010). Within a community, organisms often partition resources spatially via different habitat preferences. Spatial partitioning is facilitated by environmental heterogeneity or "patchiness". However, organisms also partition resources temporally by differential timing of key life history events including reproduction. Species utilizing different environmental cues will likely exhibit different shifts in phenology corresponding to how those cues are (or are not) altered in a changing climate. Consequently, phenological shifts can alter the type and strength of interactions within and among species (Yang & Rudolf 2010) or the magnitude of temporal niche partitioning. For example, a species which uses temperature as a cue for reproduction might reproduce earlier under warmer conditions than a species which relies primarily on photoperiod. If the two species are members of a community, then their species-specific responses to climate change could alter the ontogenetic stages at which they interact or whether they interact at all. Several such examples of possible changes in species interactions have been demonstrated, include shifting predator-prey dynamics (Both *et al.* 2009; Durant *et al.* 2007) and changes in the magnitude and dynamics of competition (Persson *et al.* 2004). Life

history theory predicts that organisms have evolved particular reproductive strategies to put their offspring into the environment at a time when key resources (e.g., prey items) are available (e.g., Match/Mismatch Hypothesis; Cushing 1969, 1990; Durant *et al.* 2007). If predators and prey exhibit differential phenological responses to climate change, then the dynamics of such a relationship could also be altered to the potential detriment (or benefit) of one or both species. In some cases these biotic effects of phenological shifts have been shown to be even greater than the abiotic effects of a changing environment (Parmesan 2006).

Environmental conditions can be highly variable over time, and this variation can be partitioned into stochastic events (e.g., flash floods) and predictable changes (e.g., seasonal patterns consistent across years). The stochastic portion of such variation can be an important driver in the relative success of reproduction, to the extent that conditions at time of reproduction and shortly thereafter match needs of resultant offspring (Match/Mismatch Hypothesis; Cushing 1969, 1990). This phenomenon is predicted to be common for species with type-III life history strategies (Schlosser 1987; Turner *et al.* 2010), or recruitment-dominated systems where adult mortality is comparatively low and the relative abundance of adults is largely a function of reproductive success in the previous year. For example, in most freshwater fish species the bulk of mortality occurs in young-of-year (YOY) fishes (reviewed in Sogard 1997). Much of this mortality is due to vulnerability of the offspring during the first few weeks of life, i.e, Critical Period Hypothesis (Hjort 1914). Thus, minor changes in survival rates of YOY fishes can have profound effects on relative and absolute abundance of adults of particular species, underscoring the importance of differences in reproductive phenology among species. This “sweepstakes” aspect of reproductive phenology can drive inter-annual

variation in relative abundance of species in that community, and thus play an important role in community assembly. For example, arid-land rivers in the southwestern U.S. often exhibit large fluctuations in discharge from year-to-year and inter-annual variation can play an important role in relative success of particular species or reproductive guilds (Pease *et al.* 2006; Turner *et al.* 2010). In the Rio Grande in New Mexico, relatively wet years favor reproductive success in Rio Grande silvery minnow (*Hybognathus amarus*), a representative of a reproductive guild with pelagic, semi-buoyant drifting eggs which spawns during flood pulses associated with spring snowmelt runoff (Platania and Altenbach 1998; Dudley and Platania 2007; Turner *et al.* 2010). Conversely, dry years favor red shiner (*Cyprinella lutrensis*), a crevice-spawning species with demersal–adhesive eggs which spawns later in the spring/summer during low flow conditions (Turner *et al.* 2010). Thus, inter-annual variation in environmental conditions, and concomitant effects on reproductive success, can dictate community assembly.

In this study, we assess inter-annual variation in reproductive phenology in fishes from a fragmented, arid-land river system in the southwestern United States to assess the dynamics of phenology in relation to environmental variation. The IPCC has predicted substantial regional climate change in the next century in the southwestern United States in terms of temperature and precipitation patterns, and concomitantly, river discharge (Intergovernmental Panel on Climate Change 2007). Climate change is of particular concern for temperate riverine fishes because the fragmented nature of many river systems (e.g., from dams or reservoirs) can limit dispersal (Jansson *et al.* 2000; Nilsson *et al.* 2005). Range shifts are one of the most common responses of organisms to climate change (Walther *et al.* 2002); however, fragmentation from dams and reservoirs and the dendritic shape of rivers overlain on environmental gradients can

prevent such migration or dispersal (Jansson *et al.* 2000). Thus, temporal rather than spatial responses to climate change might be necessary in some instances. For example, migration among headwater populations is often precluded by downstream barriers comprised of inappropriate (Krabbenhoft *et al.* 2008) or fragmented (Matthews & Marsh-Matthews 2007) habitats such as reservoirs. Additionally, flow regimes in many river systems are now regulated through dams and contemporary flow regimes often strongly differ from historical flows (Nilsson *et al.* 2005; Lytle & Poff 2004). The construction of impoundments have resulted in homogenized flows (Dudley & Platania 2007; Poff *et al.* 2007), with lower peak flows and higher low flows, a pattern largely consistent among regulated rivers across North America (Poff *et al.* 2007). Homogenization of flow regimes is of concern, as it can facilitate invasion of non-native and extirpation of native species, resulting in biotic homogenization (Dudley & Platania 2007; Rahel 2000; Rahel 2002; Xenopoulos & Lodge 2006).

In order to generate a predictive framework for the effects of climate change and direct anthropogenic impacts on arid-land riverine fish communities, we must understand proximate determinants of fish community composition and develop a framework for how organisms partition resources temporally based on seasonally varying environmental conditions. In this study, we look at the role timing and duration of reproduction play in structuring the fish community in the Rio Grande, a fragmented and highly-regulated arid-land river system in the western United States. We assess how inter-annual variation in key spawning cues affects reproductive timing. We used four years of spring and summer field collections of larval fishes to quantify fish spawning seasonality. We used these data to address the following questions: (1) Is the relative and absolute timing of spawning consistent across years? (2) To what extent

do environmental conditions (photoperiod, temperature, and discharge) predict reproductive timing across species? (3) How does spawning seasonality vary with respect to inter-annual variation in environmental conditions?

Materials and Methods

Sample locations.—Field collections of young-of-year fishes were made between April and July at five sites on the Rio Grande, New Mexico, U.S.A. (Fig. 1). We conducted field surveys over a three year time period (2008-2010) and compared these data with an existing dataset from 1995 (S.P. Platania, *unpublished*). In 1995, sampling was conducted at three sites: Sandia Pueblo, US Hwy 60 bridge, and Bosque del Apache National Wildlife Refuge (Table 1). From 2008-2010, sampling was conducted on the Rio Grande just north or upstream of Elephant Butte Reservoir, 7.4 river km upstream of the Sierra/Socorro County line (hereafter referred to as San Marcial site). In 2009 and 2010 sampling was also conducted at the Central Avenue bridge within the city limits of Albuquerque, New Mexico (hereafter referred to as the Albuquerque site). A complete list of sample dates and locality information are summarized in Appendix A.

Environmental data.—Data on putative environmental cues for spawning were compiled from a variety of sources. Mean daily discharge data ($ft^3\ s^{-1}$) were downloaded from the U.S. Geological Service (USGS) website (http://water.usgs.gov/) for the Floodway at San Marcial (USGS 08358400) and Albuquerque Gauges (USGS 08330000) and converted to m^3s^{-1}. Daily discharge data were grouped by water year, rather than calendar year. For our purposes, the water year was considered October 1st – September 30^{th} (Clow 2010). In order to quantify variation in the timing of spring flood-pulse from snowmelt runoff, we calculated the center of

mass date of the discharge data for a given water year (McCabe & Clark 2005; Stewart *et al.* 2004). Center of mass date, hereafter referred to as Q50, is the Julian date in a water year at which 50% of the annual discharge has occurred (McCabe & Clark 2005; Stewart *et al.* 2004). We calculated Q50 for all available years for the San Marcial (years 1950 – 2010) and Albuquerque gauges (years 1943 – 2010) in order to assess long-term variability in spring snowmelt.

Water temperature was recorded at time of sampling in the main channel at each site, as well as in off-channel backwater areas (where fish sampling took place). Additionally, Hobo electronic temperature loggers (Onset, Inc.) were used to record hourly water temperature at collection localities from April to July. In order to assess year-round temperature and degree days for a given year, air temperature data were garnered from the US Historical Climatology Network for the Los Lunas Station (Station no. 295150) (http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html) for years 1995 and 2008-2010 (see Fig. 1 for the location of this weather station). In order to assess long-term changes in temperature and to put conditions during this study into a broader climatic context, mean annual air temperature data for Albuquerque, New Mexico, for 1893 – 2010 were obtained from National Aeronautics and Space Administration's Goddard Institute for Space Studies website (http://data.giss.nasa.gov/gistemp/). Finally, photoperiod data for the sample sites were obtained from U.S. Navy tables (http://aa.usno.navy.mil/data/docs/RS_OneYear.php).

Autocorrelation analysis.—Environmental variables (i.e., photoperiod, air temperature, river discharge) were subjected to autocorrelation analysis using all available data from the aforementioned sources. The goal of this analysis was to determine whether current

environmental conditions predict future conditions. Good environmental cues for timing of spawning should predict what future conditions will be. For example, if conditions are favorable at the time of spawning, are they going to continue to be favorable in the near future when the offspring are exposed to the environment? Autocorrelations for lag times from 0 - 365 days were calculated for each of the variables. Long term data (e.g., several years) are required to accurately assess autocorrelation. Such data were unavailable for water temperature, which was only measured during the spawning season, and consequently we used air temperature as a proxy for these calculations. It is likely that autocorrelations would be even higher for water temperatures than air, however, because of the higher specific heat of water. Additionally, air and water temperature were highly positively cross-correlated over the reproductive seasons (r = 0.80 and 0.81 for air temperatures and San Marcial and Albuquerque water temperatures, respectively).

Young-of-year fish sampling.—Young-of-year fishes were sampled on 231 occasions from April – July of the years 1995 and 2008-2010 on the Rio Grande, New Mexico. Typically, three quatrefoil light traps were set just prior to sunset and contents were removed the following morning. Traps were set in low-velocity (approximately 0.1 m S^{-1} or less flow), backwater habitat, generally in less than 1.0 m water depth. At the Albuquerque site it was not possible, for logistical reasons, to use light traps (i.e., abundant human activity in the area), and so fishes were sampled with small mesh seines (1 m x 1 m x 0.8 mm) or a small mesh hand dip-net (20 cm x 10 cm x 0.8 mm). In our assessment, these sampling methods achieve similar results as light trapping, although specimens are in somewhat more degraded condition upon preservation due to handling. Larvae were fixed and preserved in 5% buffered formalin and

taken back to the laboratory where they were identified to species, measured (standard length (SL)), and staged according to terminology by Snyder (1976; i.e., proto-, meso-, meta-larvae, and juveniles; preflexion, flexion, postflexion). All specimens were deposited in the Museum of Southwestern Biology (MSB) (see Appendix A for MSB catalog numbers).

Larvae were enumerated according to species, date and locality. The eight most abundant species, which comprised 99.1% of the total number of YOY fish collected, were subjected to additional analyses. Rarer species were excluded from subsequent analyses because the infrequency of their collection led to uncertainty regarding timing of first spawning. In fact, several of these rarer species were only caught as later-stage larvae or juveniles. The Julian date of first appearance of a species for a given year and collection location was used as a proxy for onset of spawning season for that species. While there is some lag between the event of spawning and the first appearance of larvae in our collections, this is thought to be short and the two dates should be highly correlated. Absent direct spawning observations, which are difficult in turbid waters such as the Rio Grande, appearance of YOY fish is our best source of insight on timing of reproduction. The fish sampling data were compared with environmental data at the time of collection (i.e., photoperiod, water temperature, and river discharge).

All statistical analyses were conducted in R (http://www.r-project.org/).

Results

Environmental Variation

Temperature, photoperiod and river discharge data are presented in Fig. 2. The Rio Grande is a spring snowmelt-dominated system, where increasing photoperiod and

temperature in spring drive snowmelt in the mountainous headwaters and discharge increases concomitantly. Not surprisingly, the cross correlation of these three variables is strongly positive (Table 2). Additionally, these variables tend to be positively correlated spatially. For example, the correlation of discharge for Albuquerque and San Marcial from years 2008-2010 is 0.91 (Table 2). Finally, air temperature is a reasonably strong predictor of water temperature (r = 0.80 – 0.81; Table 2). This is important because long-term air temperature data are far more readily available than water temperature data. The three years were generally similar in air temperature, but 1995 had a substantially higher spring flood pulse than 2008 – 2010. This is likely due to the strong El Niño-Southern Oscillation climate pattern in early 1995, as compared to 2008 and 2009, which were weak to strong La Niña years (US National Oceanic and Atmospheric Administration; http://www.cpc.ncep.noaa.gov/). Strong El Niño conditions typically produce wetter conditions in the southern Rocky Mountains (Clark *et al.* 2001), as was the case in 1995. However, the year 2010 is anomalous in that it was a strong El Niño year, but had a relatively small spring flood pulse on the Rio Grande.

Autocorrelation of environmental variables.—For environmental variables to be important indicators of appropriate time to spawn, they must afford some information as to what environmental conditions will be like in the future. Figure 3 illustrates results of autocorrelation analyses for environmental variables. Air temperature and photoperiod exhibit positive, significant autocorrelation for up to 90 days lag time. Discharge also exhibits positive, albeit somewhat smaller autocorrelation for up to 90 days before it decreases to around 0. This is somewhat surprising given the perceived "flashiness" of arid-land river systems, relative to mesic streams. Together, autocorrelation plots suggest that conditions at time of spawning are

positively correlated to future conditions and thus could offer some information about what conditions will be experienced by larvae during their 'Critical Period' – e.g., the first 60 days of their lives for most freshwater fishes (Cushing 1969, 1990).

Young-of-Year Fish Data

A total of 19,838 young-of-year fishes were collected during the four sampling years spanning a sixteen-year timeframe and representing seventeen species (Table 3). More than 99.99% of individuals were identified to species, while only 11 of 19,838 specimens were unidentifiable. The relative abundances of these seventeen species were highly skewed, however, with the five and eight most abundant species comprising 96.4% and 99.1% of the total number of individuals, respectively. Relative abundances also varied greatly both spatially and temporally (Fig. 4), with colder water species such as white sucker (*Catostomus commersonii*) abundant in the more upstream sites (e.g., Sandia, Albuquerque), but rare or absent in the downstream sites (e.g., Bosque del Apache, San Marcial). Strong seasonal differences in reproductive periodicity among species were present, as well. For example, rank order of first appearance of YOY differed consistently (Table 4), suggesting reproductive seasonality differs among species. Additionally, the relative abundance of native versus non-native species was extremely variable, both temporally and spatially (Fig. 5), suggesting the intriguing possibility that conditions in the river (e.g., discharge) could be managed in a way that benefits native species to the detriment of non-natives.

Comparison across years: 1995 versus 2008–2010

The date of first appearance of each of the eight most abundant species, are plotted in Fig. 6. Young-of-year of these eight species appeared 4.3-28.1 days earlier (mean Julian date of

first appearance) in 2008-2010 than in 1995 (Table 4), suggesting earlier spawning phenology in the more recent years. Moreover, earlier spawning species (e.g., white sucker, common carp, Rio Grande silvery minnow, fathead minnow), shifted reproduction less (4.3-9.3 days earlier) in 2008-2010 versus 1995 than did later spawning species (flathead chub, red shiner, river carpsucker, western mosquitofish; 14.5-28.1 days) (Fig. 9). Next, we limited the analysis to protolarvae, as the lag time between time of spawning and capture is much less for protolarvae than later ontogenetic stages. Again, striking differences are apparent between 1995 and 2008-2010, with spawning apparently occurring much earlier in 2008-2010 (Fig. 7). This general pattern holds when species are plotted individually (Fig. 8), but earlier spawning species exhibit shorter advances in spawning time than later spawning species. As a result of these shifts, there was less temporal segregation among larvae of different species in recent years (2008-2010).

Strong differences in reproductive phenology among years begs the question: what drives shifts in spawning timing? Photoperiod is often considered to be the most important cue for seasonally-reproducing organisms. However, photoperiod does not differ across years and thus cannot drive inter-annual variation in reproductive timing. Similarly, in this study, air temperature and correspondingly, degree-days, did not differ appreciably across years (Fig. 10). In fact, air temperatures were on average slightly cooler in 2008-2010 than 1995, which should have resulted in later reproduction in these ectothermic organisms in 2008-2010 – the opposite of what was observed.

While air temperature and degree days were very similar for years 1995 versus 2008-2010, discharge patterns in the Rio Grande differed dramatically across the time spans (Fig. 11).

The timing of onset of snowmelt runoff was similar, but in 1995 the spring flood-pulse reached far greater magnitude and lasted substantially longer than in all years 2008-2010 (Fig. 11). The Q50 for daily discharge was Julian date 237 in year 1995 versus 214, 220, and 215 in years 2008-2010, respectively, resulting in an approximately 21 day advance relative to 1995. Interestingly, the mean difference of 21 days advance in discharge center of mass is similar, though generally greater, than the number of days phenology was earlier in 2008-2010 compared to 1995 (Table 4).

Discussion

Within biological communities, species often partition resources both spatially and temporally as a means of reducing competition. However, in arid-land rivers, reduced flows can drastically decrease habitat heterogeneity (Ward & Stanford 1995, Bunn & Arthington 2002). For example, dry years can result in loss of connectivity to flood-plain habitats, and concomitant reduction of spatial niche partitioning among fishes. This reduction in heterogeneity is exacerbated in fragmented systems such as the Rio Grande, because dams can constrict up- and downstream movements of fishes (Dudley & Platania 2007; Matthews & Marsh-Matthews 2007). Data presented in this paper suggest that temporal niche partitioning may also be affected by changes in flow regime, resulting in double jeopardy for the fish community.

It is possible that temporal niche partitioning facilitates the persistence of different species, or conversely, reduction of temporal partitioning could result in decreased reproductive success for certain species. In this study, all eight of the most abundant species spawned earlier in the drier years (2008-2010) as compared to 1995. However, the number of

days spawning advanced was greater for later spawning species. As a consequence, among-species differences in timing of the onset of the reproductive season are reduced in low flood-pulse years, with greater overlap in reproductive periodicity among species. This begs the important question of whether competition for resources among larvae of different species is likely to increase under such periods of greater temporal overlap. Carbon and nitrogen stable isotope data suggest that larval fishes in the Rio Grande in New Mexico have highly-overlapping patterns of resource use (Pease et al. 2006; Krabbenhoft et al. 2012; Turner *et al.* 2010).

The physiological or genetic mechanisms underlying shifts in reproductive timing are not known. In particular, it is not known what drives species-specific responses to among-year environmental variation. It is possible that organisms are able to plastically-shift phenology only to a certain threshold, beyond which they are unable to respond further without genetic adaptation (Phillimore *et al.* 2010). This scenario is of particular concern, as the rate of climate change is likely far greater than the maximal rate of molecular evolution in these organisms. One possibility for the smaller shift in timing in earlier spawning fishes in this study is that cold temperatures could act as a limit on how early reproduction can occur in these ectothermic organisms, either directly on acting on physiological performance, or indirectly on limiting food sources available. Additionally, survival of larval fishes is often greatly reduced in cold temperatures (Blaxter 1991), suggesting a possible trade-off to spawning earlier in years with low spring-runoff. In our opinion, unraveling the mechanistic underpinnings of reproductive phenology is key to understanding these dynamics.

There are several alternate explanations for the perceived shift in spawning periodicity across years. One possibility is that larvae were not detected until later in 1995 because of less

sampling effort or decreased sampling efficiency in that wetter year (i.e., a dilution effect). This scenario is unlikely, as the first larvae sampled in that year were very small protolarvae, suggesting that they were recently spawned. If there was a large lag between spawning time and appearance of larvae in our collections, then larvae would likely be larger or later ontogenetic stages (e.g., meso- or meta-larvae). In fact, our first samples contained exclusively individuals which were very early (=young) ontogenetic stages in all years. A second possible explanation is that apparent differences spawning time reflect different sample sites used in 1995 versus 2008-2010. This too is unlikely. For example, the Bosque del Apache and US HWY 60 sites (both sampled in 1995) are located between the two sites sampled in 2008-2010 (San Marcial and Albuquerque). The third site sampled in 1995, Sandia, is less than 10 km upstream of the Albuquerque site.

While 1995 was wetter year than 2008-2010, it is not outside the realm of typical variation present in this geographic area. For example, mean annual air temperatures in Albuquerque, New Mexico, have been rising over the past 120 years and substantial among year variation appears common (Fig. 10). While these data span a relatively long (16-year) time frame, there are only four years of data and only one 'wet' year (1995). Water year 1995 had a substantially higher spring flood-pulse than average, while 2008-2010 had only slightly lower spring flows than historically. These data suggest that a large flood pulse could act to delay reproduction. Additional long-term datasets on flow variability and fish reproductive periodicity in the Rio Grande are needed. Differences in flow between 1995 and 2008-2010 are likely not due to climate change over this 16-year time period, but rather reflect "normal" variation across years, e.g., such as that arising from fluctuations in the El Niño-Southern

Oscillation. However, climate change models consistently predict that the local and regional climate will continue to get hotter and drier (IPCC 2007), and that increased human demand for water resources will place additional strain on river systems in the southwestern United States. One possible outcome of this scenario is reduced and earlier flows and a continued shift in earlier spawning periodicity and greater overlap in reproductive timing among species. The effects this could have on competition and predation dynamics are not known, but adult fish community composition could be altered. Finally, we argue that local temperature alone is insufficient to determine the effects of climate change and additional regional environmental conditions could also play an important role. For example, earlier snowmelt in the headwaters of the Rio Grande could affect discharge patterns hundreds of kilometers away.

Studying phenology at the community scale affords not only information about individual species' responses, but also insight into shifts in interactions among species. For example, in this study we found varying degrees of overlap in reproductive periodicity in Rio Grande fishes among years. Shifts in the timing of interactions among species may have "profound ecological consequences" (Walther *et al.* 2002). Future research should focus on whether variation in temporal niche partitioning among fishes affects prey availability, magnitude of resource use overlap, and competition dynamics.

Acknowledgements

We sincerely thank American Southwest Ichthyological Research Foundation, specifically RK Dudley, WH Brandenburg, MA Farrington, A Barkalow, JL Hester, RC Keller, CM McBride, and M Brandenberg for field and laboratory assistance. WH Brandenburg and MA Farrington

assisted with identification of fish samples. AM Snyder kindly provided curatorial assistance at Museum of Southwestern Biology. Thanks to Y Lin for statistical discussions. CA Krabbenhoft, KG O'Malley, C Cunningham, and SL Collins provided valuable feedback on the manuscript. Fishes were collected under NM Game and Fish permit #3015 and US Fish and Wildlife permit #TE038055-0. This research was approved by Institutional Animal Care and Use Committee Protocol #10-100468-MCC and 10-100492-MCC.

Literature Cited

Balasubramanian, S., Sureshkumar, S., Agrawal, M., Michael, T.P., Wessinger, C., Maloof, J.N., *et al.* (2006) The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nature Genetics* 38:711-715.

Blaxter, J.H.S. (1991) The influence of temperature on larval fishes. *Netherlands Journal of Zoology* 42(2-3):336-357.

Both, C., Van Ash, M., Bijlsma, R.G., Burg Van Den, A.B., & Visser, M.E. (2009) Climate change and unequal phenological changes across four trophic levels: constraints or adaptations? *Journal of Animal Ecology* 78:73-83.

Bradley, N.L., Leopold, A.C., Ross, J., & Huffaker, W. (1999) Phenological changes reflect climate change in Wisconsin. *Proceedings of the National Academy of Sciences of the United States of America* 96:9701-9704.

Bradshaw, W.E., & Holzapfel, C.M. (2008) Genetic response to rapid climate change: it's seasonal timing that matters. *Molecular Ecology* 17(1):157-166.

Bradshaw, W.E., Zani, P.A., & Holzapfel, C.M. (2004) Adaptation to temperate climates. *Evolution* 58:1748-1762.

Bunn, S.E., & Arthington, A.H. (2002) Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental Management* 30(4):492-507.

Clark, M.P., Serreze, M.C., & McCabe, G.J. (2001) Historical effects of El Nino and La Nina events on the seasonal evolution of the montane snowpack in the Columbia and Colorado River Basins. *Water Resources Research* 37(3):741-757.

Clow, D.W. (2010) Changes in the timing of snowmelt and streamflow in Colorado: a response to recent warming. *Journal of Climate* 23:2293-2306.

Cushing, D.H. (1969) The regularity of the spawning season of some fishes. *Journal du Conseil International pour L'exploration de la Mer* 33:81-92.

Cushing, D.H. (1990) Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. *Advances in Marine Biology* 26:249-263.

Dudley, R.K., & Platania, S.P. (2007) Flow regulation and fragmentation imperil pelagic-spawning riverine fishes. *Ecological Applications* 17(7):2074-2086.

Dunn, P.O., & Winkler, D.W. (1999) Climate change has affected the breeding date of Tree Swallows throughout North America. *Proceedings of the Royal Society B: Biological Sciences* 266:2487-2490.

Durant, J.M., Hjermann, D.Ø., Ottersen, G., & Stenseth, N.C. (2007) Climate and the match or mismatch between predator requirements and resource availability. *Climate Research* 33:271-283.

Fujisawa, M., & Kobayashi, K. (2010) Apple (*Malus pumila var. domestica*) phenology is advancing due to rising air temperature in northern Japan. *Global Change Biology* 16:2651-2660.

Gillet, C., & Quétin, P. (2006) Effect of temperature changes on the reproductive cycle of roach in Lake Geneva from 1983 to 2001. *Journal of Fish Biology* 69:518-534.

Gordo, O., & Sanz, J.J. (2005) Phenology and climate change: a long-term study in a Mediterranean locality. *Oecologia* 146(3):484-495.

Gordo, O., & Sanz, J.J. (2006) Climate change and bird phenology: a long-term study in the Iberian Peninsula. *Global Change Biology* 12(10):1993-2004.

Hancock, A.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334:83-86.

Hjort, J. (1914) Fluctuations in the great fisheries of northern Europe. *Rapports Proc-Verb Reun Const Int Explor Mer* 20:1-228.

Intergovernmental Panel on Climate Change (2007) Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K., & Reisinger, A. (Eds.)]. IPCC, Geneva, Switzerland, 104 pp.

Jansson, R., Nilsson, C., & Renöfält, B. (2000) Fragmentation of riparian floras in rivers with multiple dams. *Ecology* 81(4):899-903.

Kirschbaum, F. (1975) Environmental factors control the periodical reproduction of tropical electric fish. *Cellular and Molecular Life Sciences* 31(10):1159-1160.

Kirschbaum, F. (1979) Reproduction of the weakly electric fish *Eigenmannia virescens* (Rhamphichthyidae, Teleostei) in captivity. *Behavioral Ecology and Sociobiology* 4:331-355.

Krabbenhoft, C.A., Burdett, A.S., & Turner, T.F. (2012) *In prep.*

Krabbenhoft, T.J., Rohde, F.C., Leibman, A.N., & Quattro, J.M. (2008) Concordant mitochondrial and nuclear DNA partitions define Evolutionarily Significant Units in the imperiled pinewoods darter, *Etheostoma mariae* (Pisces: Percidae). *Copeia* 2008(4):909-915.

Lahiri, K., Vallone, D., Babu Gondi, S., Santoriello, C., Dickmeis, T., & Foulkes, N.S. (2005) Temperature regulates transcription in the zebrafish circadian clock. *PLoS Biology* 3:2005-2016.

Leatherland, J.F., Farbridge, K.J., & Boujard, T. (1992) Lunar and semi-lunar rhythms in fishes. Pp. 83-107 in Ali, M.A. (Ed.) *Rhythms in Fishes.* Plenum Press, New York, NY.

Lytle, D.A., & Poff, N.L. (2004) Adaptation to natural flow regimes. *Trends in Ecology and Evolution* 19(2):94-100.

Matthews, W.J., & Marsh-Matthews, E. (2007) Extirpation of red shiner in direct tributaries of Lake Texoma (Oklahoma-Texas): a cautionary case history from a fragmented river-reservoir system. *Transactions of the American Fisheries Society* 136:1041-1062.

McCabe, G.J., & Clark, M.P. (2005) Trends and variability in snowmelt runoff in the western United States. *Journal of Hydrometeorology* 6:476-482.

Nilsson, C., Reidy, C.A., Dynesius, M., & Revenga, C. (2005) Fragmentation and flow regulation of the world's large river systems. *Science* 308:405-408.

Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* 37:637-669.

Parmesan, C., & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421:37-42.

Pease, A.A., Davis, J.J., Edwards, M.S., & Turner, T.F. (2006) Habitat and resource use by larval and juvenile fishes in an arid-land river (Rio Grande, New Mexico). *Freshwater Biology* 51:475-486.

Persson, L., Claessen, D., De Roos, A.M., Byström, P., Sjögren, S., Svanbäck, R., *et al.* (2004) Cannibalism in a size-structured population: energy extraction and control. *Ecological Monographs* 74(1): 135-157.

Phillimore, A.B., Hadfield, J.D., Jones, O.R., & Smithers, R.J. (2010) Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 107(18):8292-8297.

Poff, N.L., Olden, J.D., Merritt, D.M., & Pepin, D.M. (2007) Homogenization of regional river dynamics by dams and global biodiversity implications. *Proceedings of the National Academy of Sciences of the United States of America* 104(14):5732-5737.

Rahel, F.J. (2000) Homogenization of fish faunas across the United States. *Science* 288:854-856.

Rahel, F.J. (2002) Homogenization of freshwater faunas. *Annual Reviews in Ecology, Evolution, and Systematics* 33:291-315.

Schlosser, I. (1987) A conceptual framework for fish communities in small warmwater streams. Pp. 17-24 *In*: Matthews, W.J. & Heins, D.C. (Eds.) *Community and evolutionary ecology of North American stream fishes.* University of Oklahoma Press, Norman, OK, USA.

Schneider, K.N., Newman, R.M., Card, V., Weisberg, S., & Pereira, D.L. (2010) Timing of walleye spawning as an indicator of climate change. *Transactions of the American Fisheries Society* 139:1198-1210.

Snyder, D.E. (1976) Terminologies for intervals of larval fish development. Pages 41-60 *In:* Boreman, J. (Ed.) *Great Lakes fish egg and larvae identification, proceedings of a workshop*. U.S. Fish and Wildlife Service FWS/OBS-76/23.

Sogard, S.M. (1997) Size-selective mortality in the juvenile stage of teleost fishes: a review. *Bulletin of Marine Science* 60(3):1129-1157.

Stinchcombe, J.R., Weinig, C., Ungerer, M., Olsen, K.M., Mays, C., Halldorsdottir, S.S., *et al.* (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences of the United States of America* 101:4712-4717.

Stewart, I.T., Cayan, D.R., & Dettinger, M.D. (2004) Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. *Climatic Change* 62:217-232.

Takemura, A., Rahman, M.S., & Park, Y.J. (2010) External and internal controls of lunar-related reproductive rhythms in fishes. *Journal of Fish Biology* 76:7-26.

Tauber, E., Zordan, M., Sandrelli, F., Pegoraro, M., Osterwalder, N., Breda, C., *et al.* (2007) Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. *Science* 316:1895-1898.

Turner, T.F., Krabbenhoft, T.J., & Burdett, A.S. (2010) Reproductive phenology and fish community structure in an arid-land river system. *In: Community Ecology of Stream Fishes* (Gido, K., & Jackson, D., Eds.). *American Fisheries Society Symposium* 73:427-446.

Walther, G., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., *et al.* (2002) Ecological responses to recent climate change. *Nature* 416:389-395.

Ward, J.V., & Stanford, J.A. (1995) Ecological connectivity in alluvial river ecosystems and its disruption by flow regulation. *River Research and Applications* 11(1):105-119.

Wilczek, A.M., Burghardt, L.T., Cobb, A.R., Cooper, M.D., Welch, S.M., & Schmitt, J. (2010) Genetic and physiological bases for phenological responses to current and predicted climates. *Philosophical Transactions of the Royal Society, Series B* 365:3129-3147.

Xenopoulos, M.A., & Lodge, D.M. (2006) Going with the flow: using species-discharge relationships to forecast losses in fish biodiversity. *Ecology* 87(8):1907-1914.

Yang, L.H., & Rudolf, V.H.W. (2010) Phenology, ontogeny and the effects of climate change on the timing of species interactions. *Ecology Letters* 13:1-10.

Table 1. Collection locality information and number of young-of-year fishes sampled. Localities are mapped in Fig. 1. See Appendix 1 for additional locality information.

Sample location	Site Name	UTMs (Zone 13S)	Year sampled	# collections made	Gear type	# YOY collected
Sandia Pueblo	Sandia	353302E, 3897465N	1995	30	Light trap	3510
US Hwy 60 bridge	USHWY60	334593E, 3809728N	1995	30	Light trap	2435
Bosque del Apache	BDANWR	326882E, 3740925N	1995	27	Light trap	175
San Marcial	San Marcial	0305003E, 3711804N	2008-2010	41, 23, 43	Light trap	10407
Central Avenue bridge, Albuquerque	Albuquerque	0346608E, 3884164N	2009-2010	16, 21	Larval seine	3311

Table 2. Pearson correlation coefficients of environmental variables during years 2008-2010. See text for description of data sources.

	ABQ Discharge	San Marcial Discharge	Photoperiod	Air Temp	ABQ Temp
San Marcial Discharge	0.91				
Photoperiod	0.52	0.41			
Air Temp	0.29	0.19	0.88		
ABQ Temp	-0.27	-0.09	0.83	0.81	
San Marcial Temp	-0.25	-0.13	0.65	0.80	0.94

Table 3. Young-of-year fishes collected over four years spanning a 16-year timeframe. Taxon abbreviations are as follows: Hyb ama, *Hybognathus amarus*, Rio Grande Silvery Minnow; Car car, *Carpiodes carpio*, River Carpsucker; Cat com, *Catostomus commersonii*, White Sucker; Cyp car, *Cyprinus carpio*, Common Carp; Pim pro, *Pimephales promelas*, Fathead Minnow; Cyp lut, *Cyprinella lutrensis*, Red Shiner; Gam aff, *Gambusia affinis*, Western Mosquitofish; Pla gra, *Platygobio gracilis*, Flathead Chub; Ict pun, *Ictalurus punctatus*, Channel or Blue Catfish; Lep mac, *Lepomis macrochirus*, Bluegill; Rhi cat, *Rhinichthys cataractae*, Longnose Dace; Pom ann, *Pomoxis annularis*, White Crappie; Lep cya, *Lepomis cyanellus*, Green Sunfish; Per fla, *Perca flavescens*, Yellow Perch; San vit, *Sander vitreus*, Walleye; Dor cep, *Dorosoma cepedianum*, Gizzard Shad; Mic sal, *Micropterus salmoides*, Largemouth Bass; Unident, unidentified.

Sample Site	Hyb ama	Car car	Cat com	Cyp car	Pim pro	Cyp lut	Gam aff	Pla gra	Ict pun	Lep mac	Rhi cat	Pom ann	Lep cya	Per fla	San vit	Dor cep	Mic sal	Un-ident	Totals
2010 Albuquerque	4	14	1347	36	75	89	17	0	0	5	0	1	0	0	0	0	0	0	**1588**
2010 San Marcial	2059	583	18	286	33	17	4	28	13	0	0	0	0	0	1	0	0	0	**3042**
2009 Albuquerque	892	2	614	40	61	38	65	0	0	3	5	2	1	0	0	0	0	0	**1723**
2009 San Marcial	3605	137	10	494	85	18	57	1	0	5	0	0	0	1	0	0	0	0	**4413**
2008 San Marcial	882	1502	0	124	388	28	4	12	1	0	0	0	0	0	0	0	0	11	**2952**
1995 Sandia	155	1	2806	389	32	6	0	2	0	1	2	48	4	0	0	63	1	0	**3510**
1995 US60	391	72	475	14	1337	112	3	29	0	0	1	0	0	0	0	0	1	0	**2435**
1995 BDANWR	54	11	16	70	18	4	0	0	0	0	0	2	0	0	0	0	0	0	**175**
Totals	**8042**	**2322**	**5286**	**1453**	**2029**	**312**	**150**	**72**	**14**	**14**	**8**	**53**	**5**	**1**	**1**	**63**	**2**	**11**	**19838**
Relative abundance	**0.405**	**0.117**	**0.266**	**0.073**	**0.102**	**0.016**	**0.008**	**0.004**	**0.001**	**0.001**	**<0.001**	**0.003**	**<0.001**	**<0.001**	**<0.001**	**0.003**	**<0.001**	**0.001**	

Table 4. Mean and standard deviation of rank order, and mean Julian date of first appearance of YOY fishes from 1995 and 2008-2010. See Table 3 for list of species abbreviations. Only the eight most abundant species are presented here. Mean Julian date refers to the mean date at which young-of-year fish are first collected for a given species, averaged across sites for particular time spans. Note the correspondence between rank order first appearance (calculated across all years) and difference in Julian date across years: later spawning fishes advanced reproduction more than earlier spawning species.

	Cat com	***Cyp car***	***Hyb ama***	***Pim pro***	***Car car***	***Pla gra***	***Gam aff***	***Cyp lut***
Mean rank order 1st appearance	1.3	2.0	2.4	4.3	4.5	5.6	6.5	7.0
Standard deviation 1st appearance	0.8	0.8	1.1	0.7	2.4	0.5	1.0	1.2
Mean Julian date first appearance (1995)	124.3	132.7	140.7	146.3	165.7	153.5	171.0	178.3
Mean Julian date 1st appearance (2008-2010)	120.0	124.8	131.4	138.2	137.6	139.0	156.4	160.4
Difference in Julian date 1st appearance	4.3	7.9	9.3	8.1	28.1	14.5	14.6	17.9

Figure 1. Study area and sample locations on the middle Rio Grande, New Mexico. Years sampled are presented in the legend.

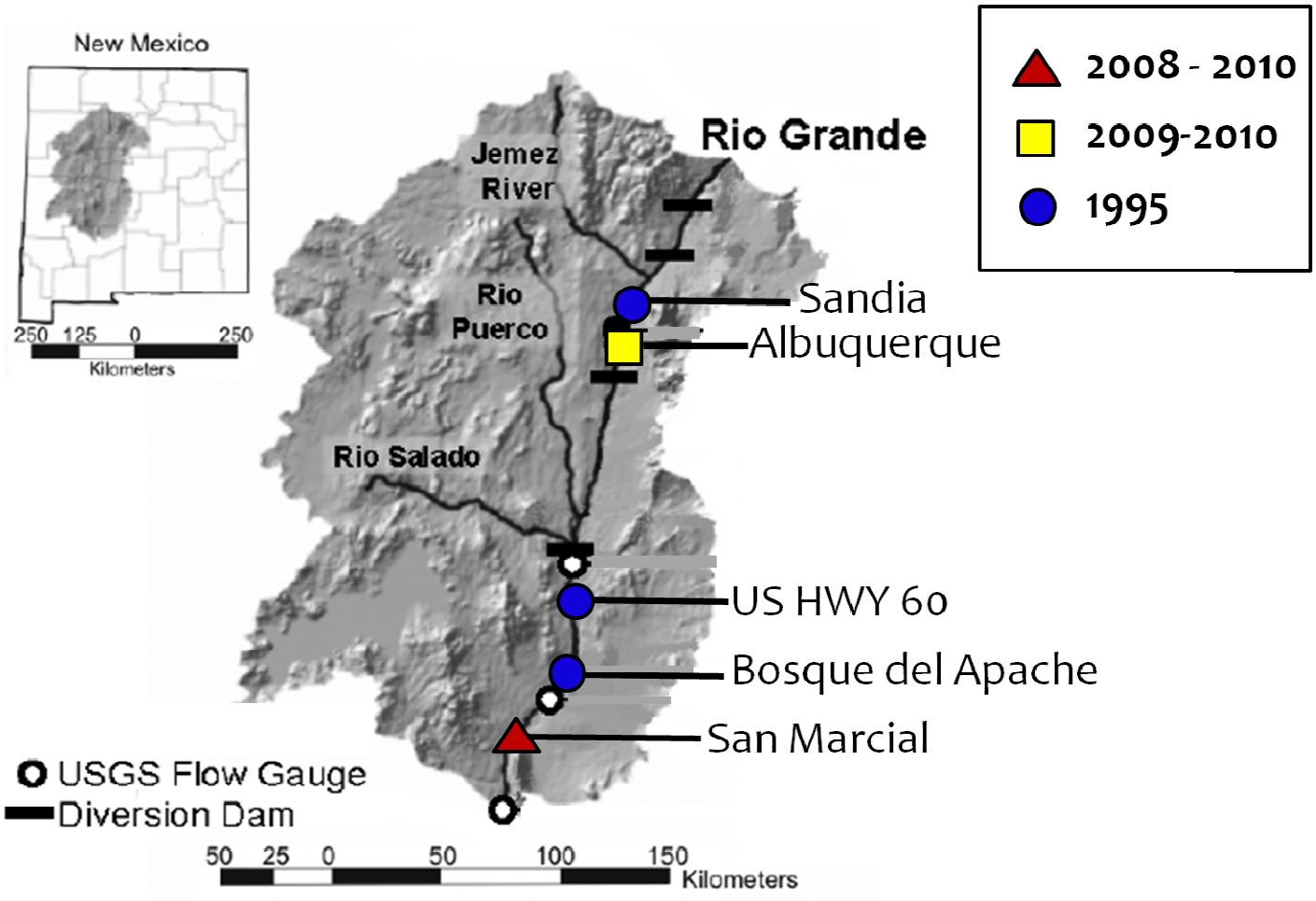


Figure 2. Environmental data for the Rio Grande for 1995 and 2008-2010. Top panel: mean daily discharge for the San Marcial (solid line) and Albuquerque Gauge (dotted line). Note the tight correlation between discharge at the two sites. Bottom panel: photoperiod (solid line) and mean daily air temperature data (points). These variables are highly cross-correlated. The reproductive seasons of most fishes in this study fall between April and July, which coincides with the most rapid environmental change.

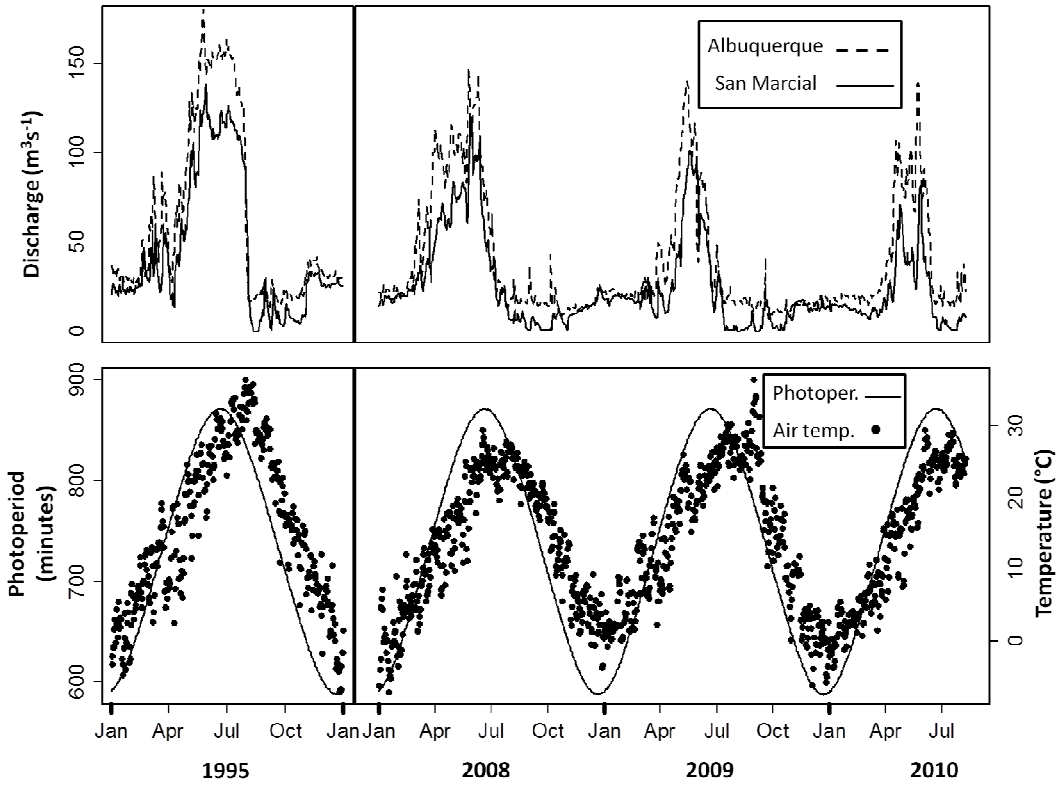


Figure 3. Flow autocorrelations for putative cues for spawning. The X-axis represents lag time (in days) and Y-axis represents Pearson correlation coefficient for a given lag time. Note the positive autocorrelation in all four cases for periods of up to ~100 days, suggesting these variables are fair predictors of future conditions after spawning.

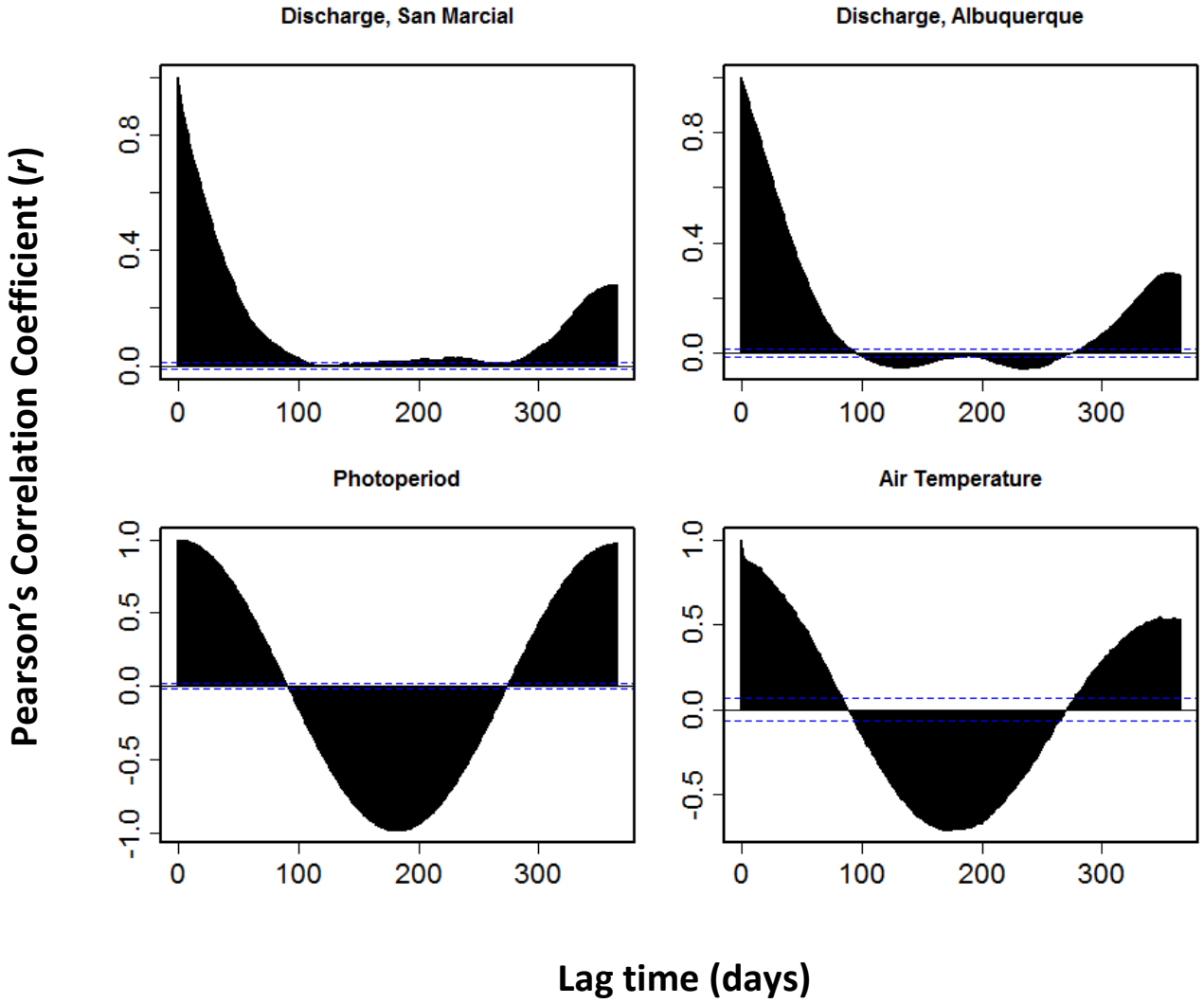


Figure 4. Percent abundance for YOY collected at each site for 1995 and 2008 – 2010. Relative abundance of species is highly variable, both across years and sites.

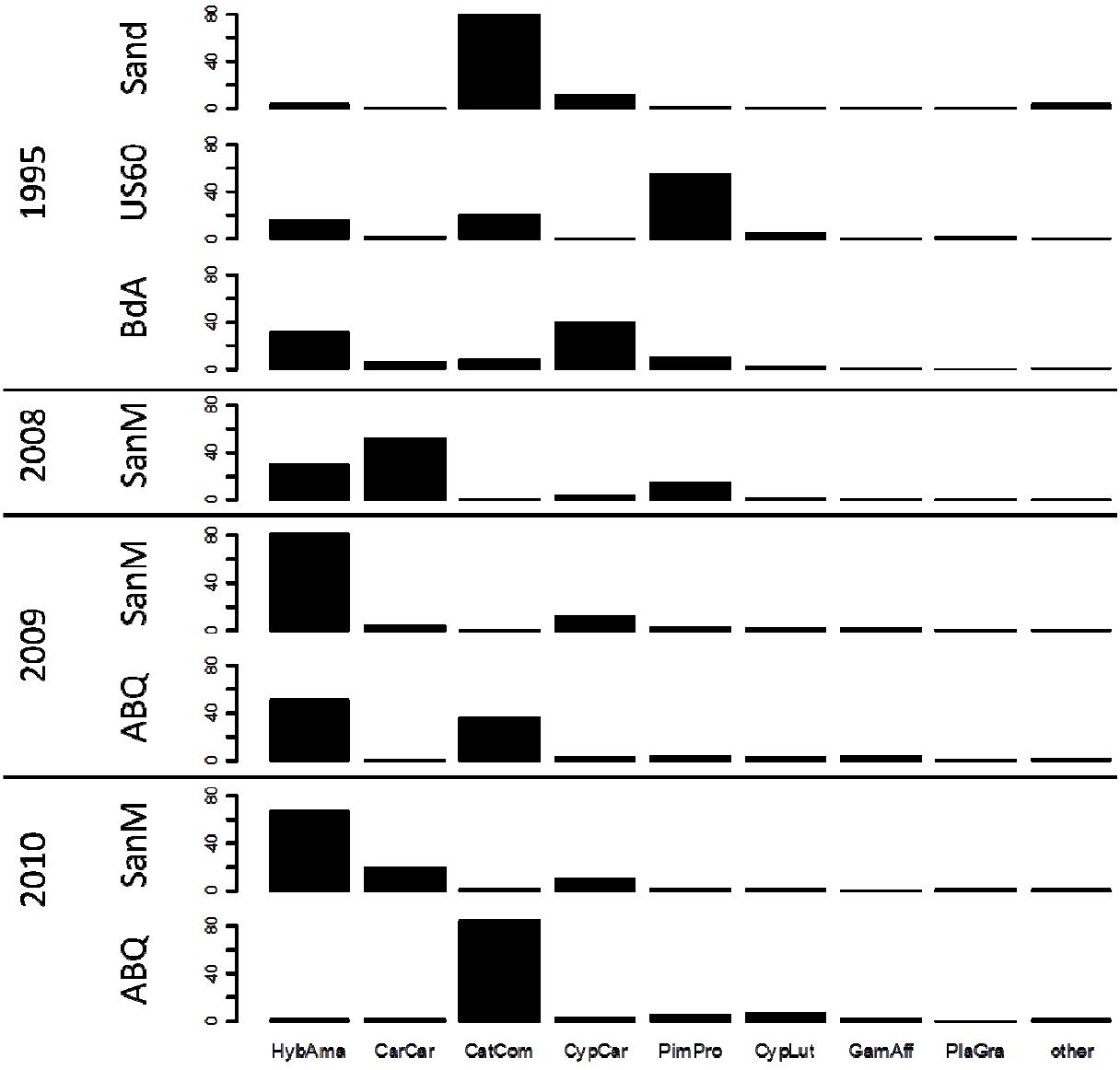


Figure 5. Percentage of young-of-year fishes that were introduced (yellow bars) versus native (blue bars) for each site and year. Note the high variability in the percent of native versus non-native species among years and sites. If these dynamics were better understood, perhaps flow conditions could be adaptively managed to shift the fish community toward native species.

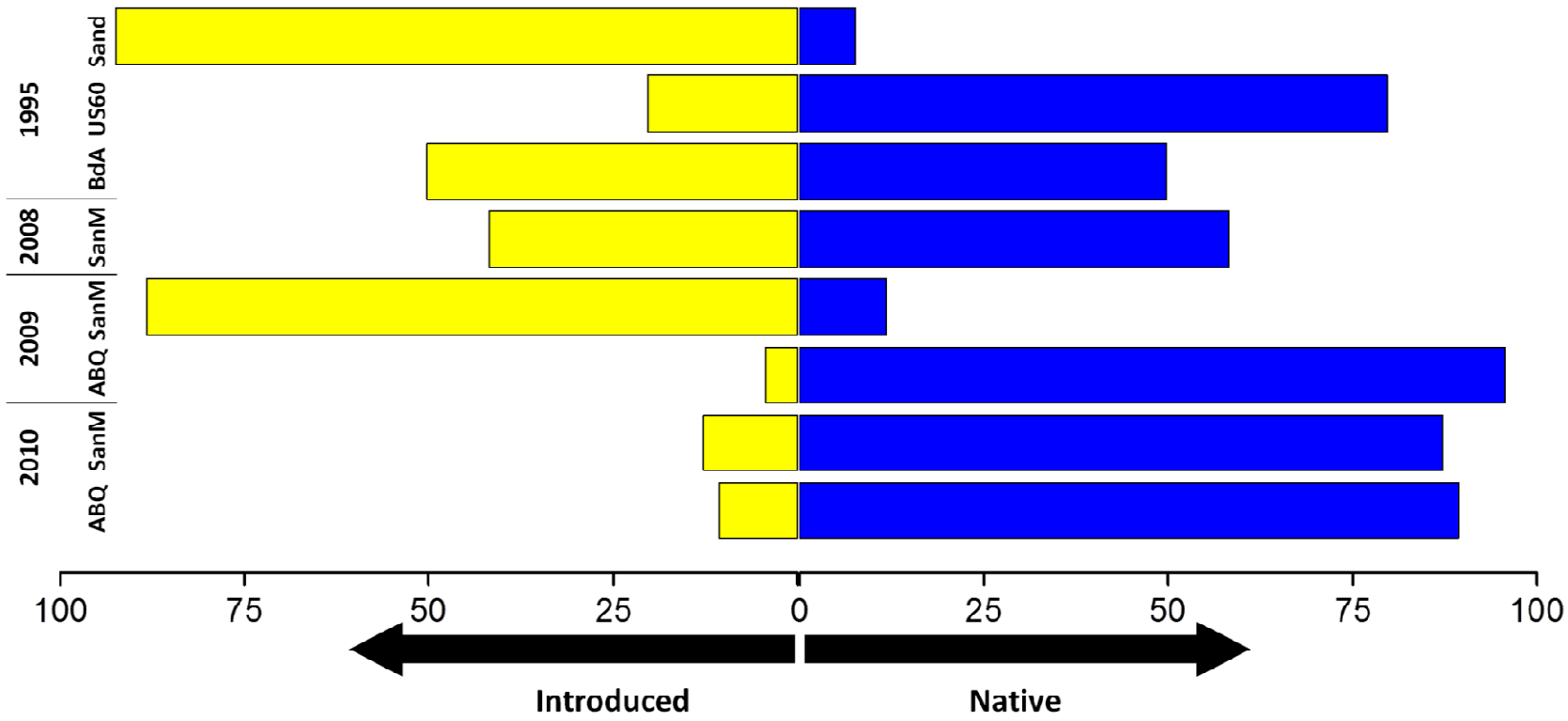


Figure 6. Date of first appearance of young-of-year of eight species in this study. Black circles represent samples from each location from 2008-2010, while red triangles represent the three localities sampled in 1995. Note the first appearance of larvae was generally later for each species in 1995, regardless of sample site.

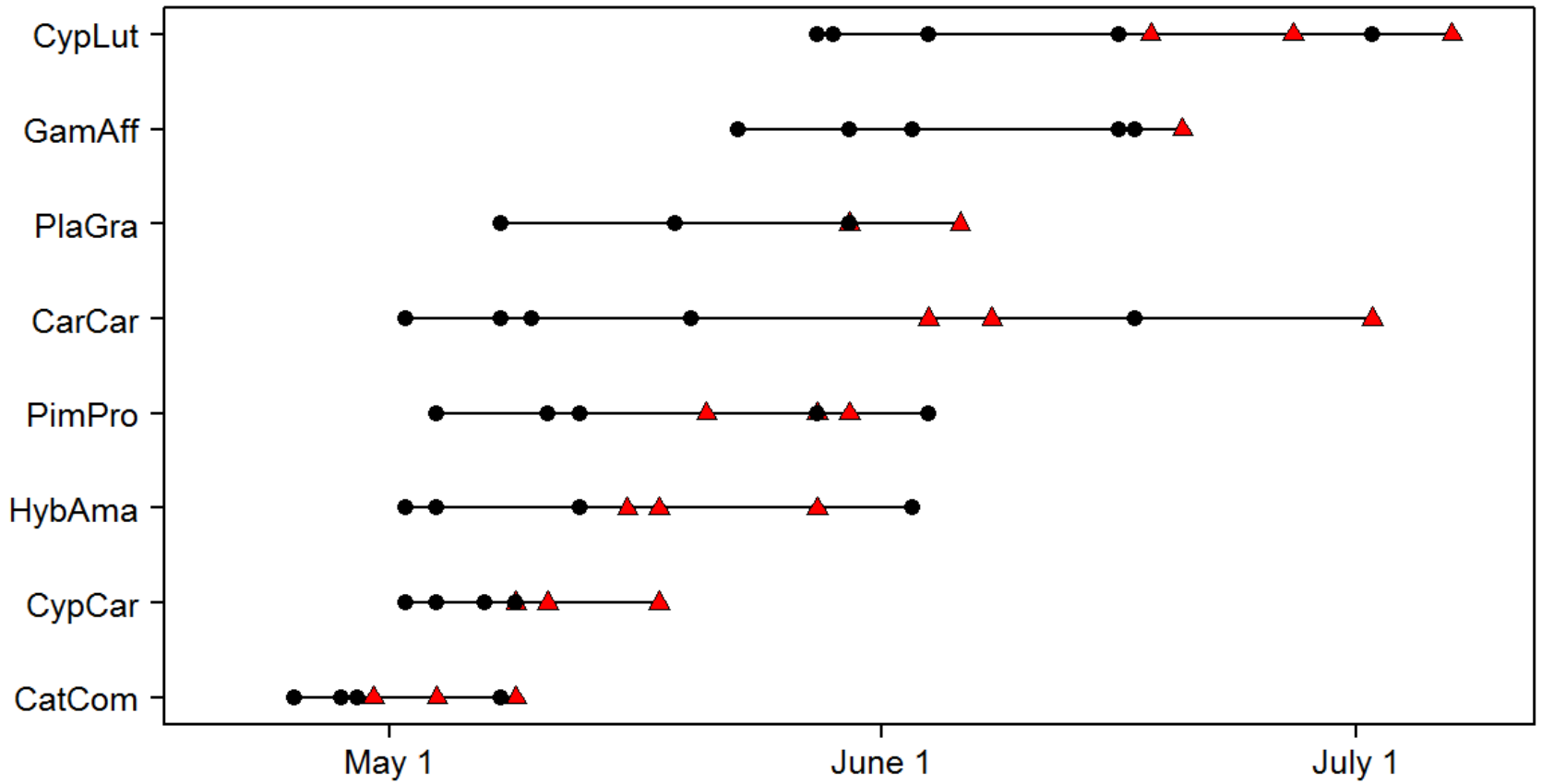


Figure 7. Sample date (x-axis) and standard length (y-axis, in mm) of 4,240 protolarvae collected from years 1995 and 2008-2010 for all species. Symbols represent individual specimens collected in 1995 (black triangles), 2008 (red inverted triangles), 2009 (blue diamonds) and 2010 (green circles). Note the obvious community wide advance of spawning time in 2008-2010 versus 1995.

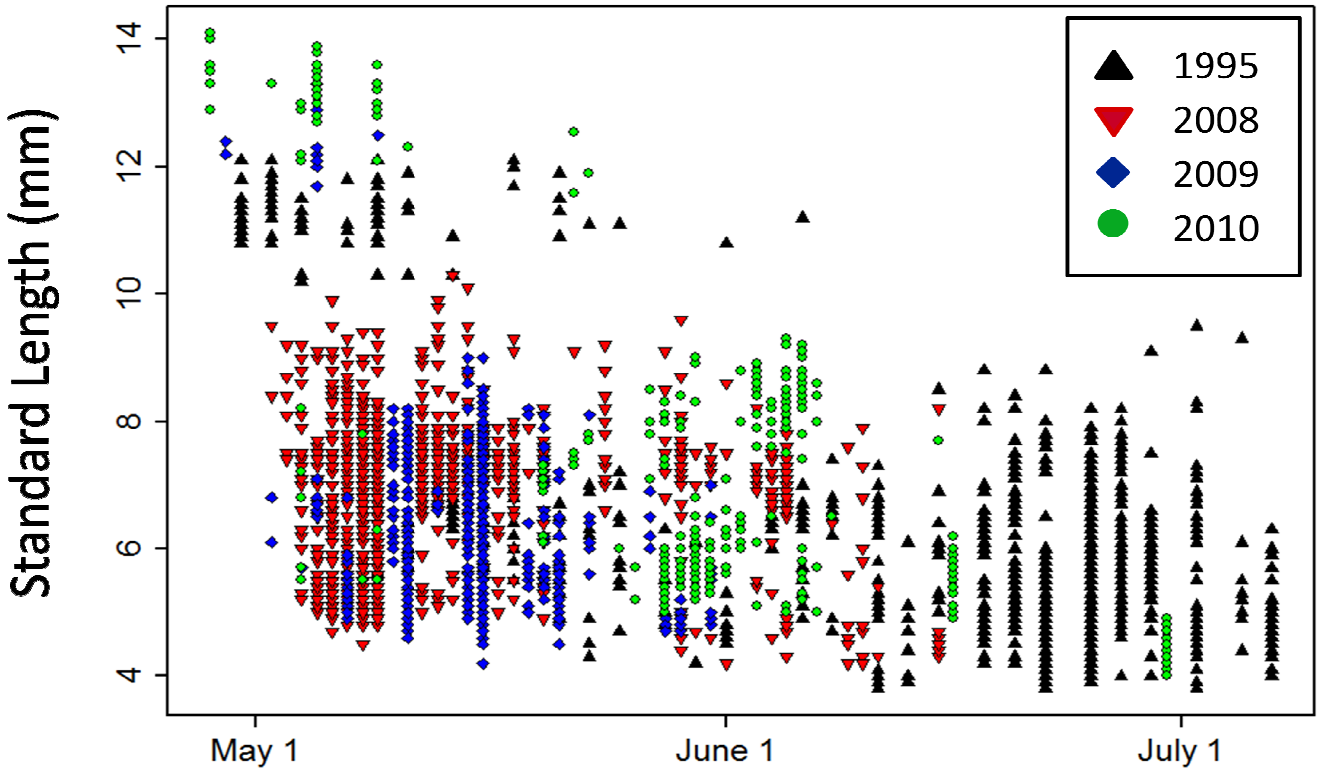


Figure 8. Standard length (y-axis, in mm) versus Julian date (x-axis) of 4,240 protolarvae collected from years 1995 and 2008-2010. Symbols are the same as in Figure 6. *Gambusia affinis* not shown because they are ovoviviparous and lack a distinct larval stage. Note the species-specific differences in magnitude of shifts in spawning time from 1995 to 2008-2010.

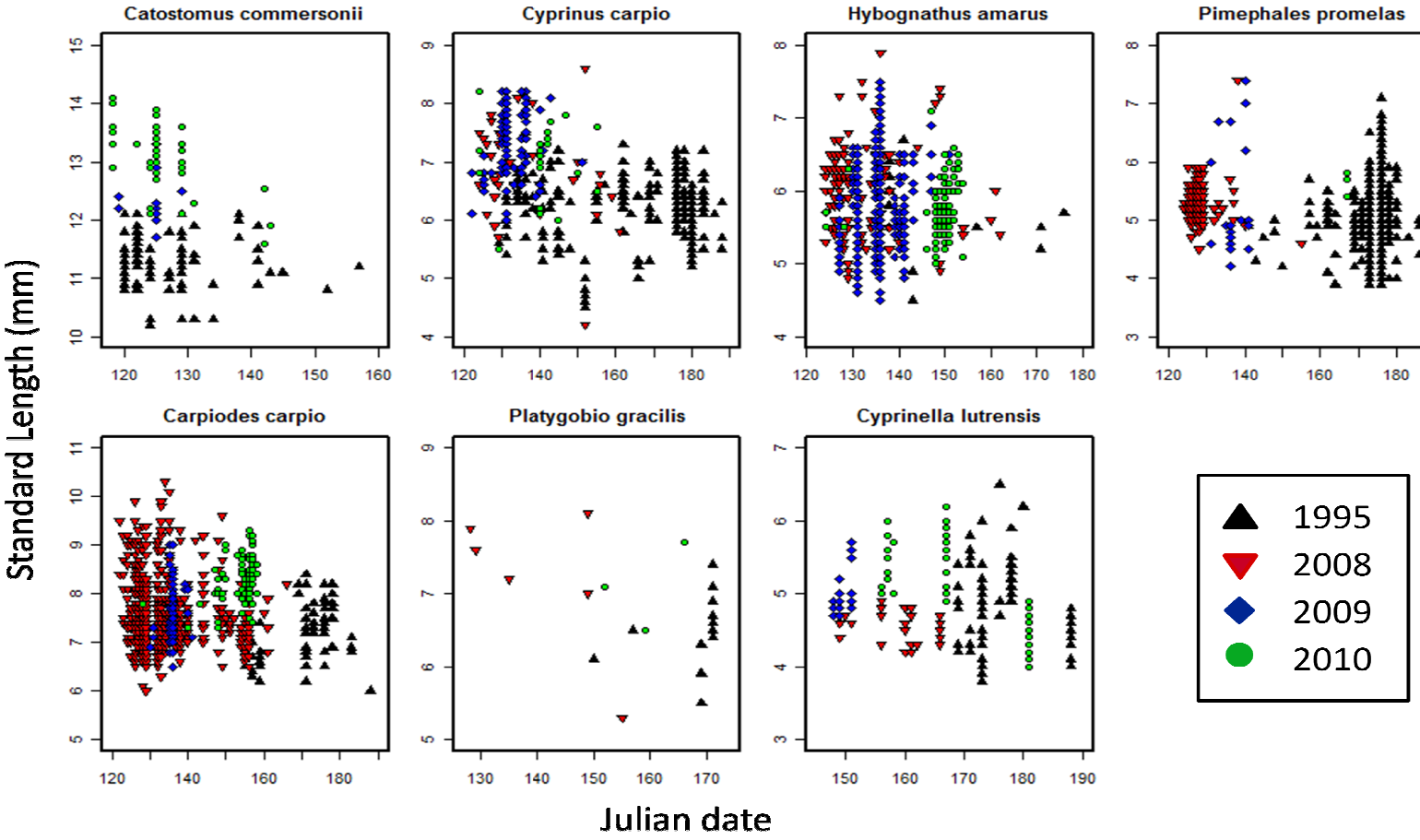


Figure 9. Magnitude of shifts in spawning season for 2008-2010 vs. 1995 versus rank order of spawning. Note that earlier spawning species advanced spawning less than later spawning species. These shifts result in greater overlap among species in reproductive periodicity in 2008-2010. The line represents least squares regression of days spawning advanced versus rank order of appearance.

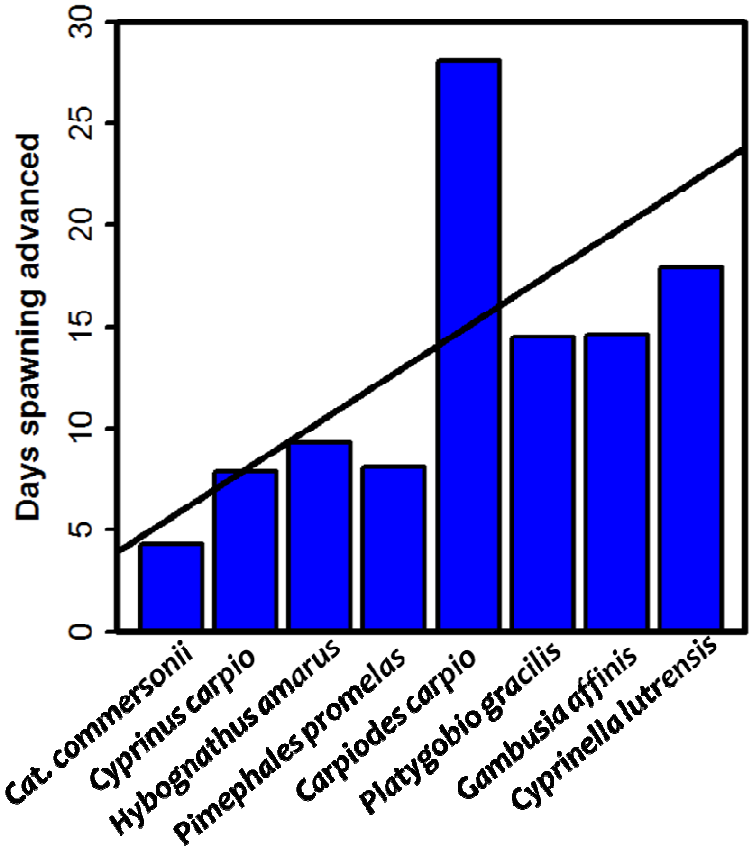


Figure 10. Air temperature for the four years of this study. Note the apparent lack of differences in temperature among years, including immediately prior to and during the spawning season.

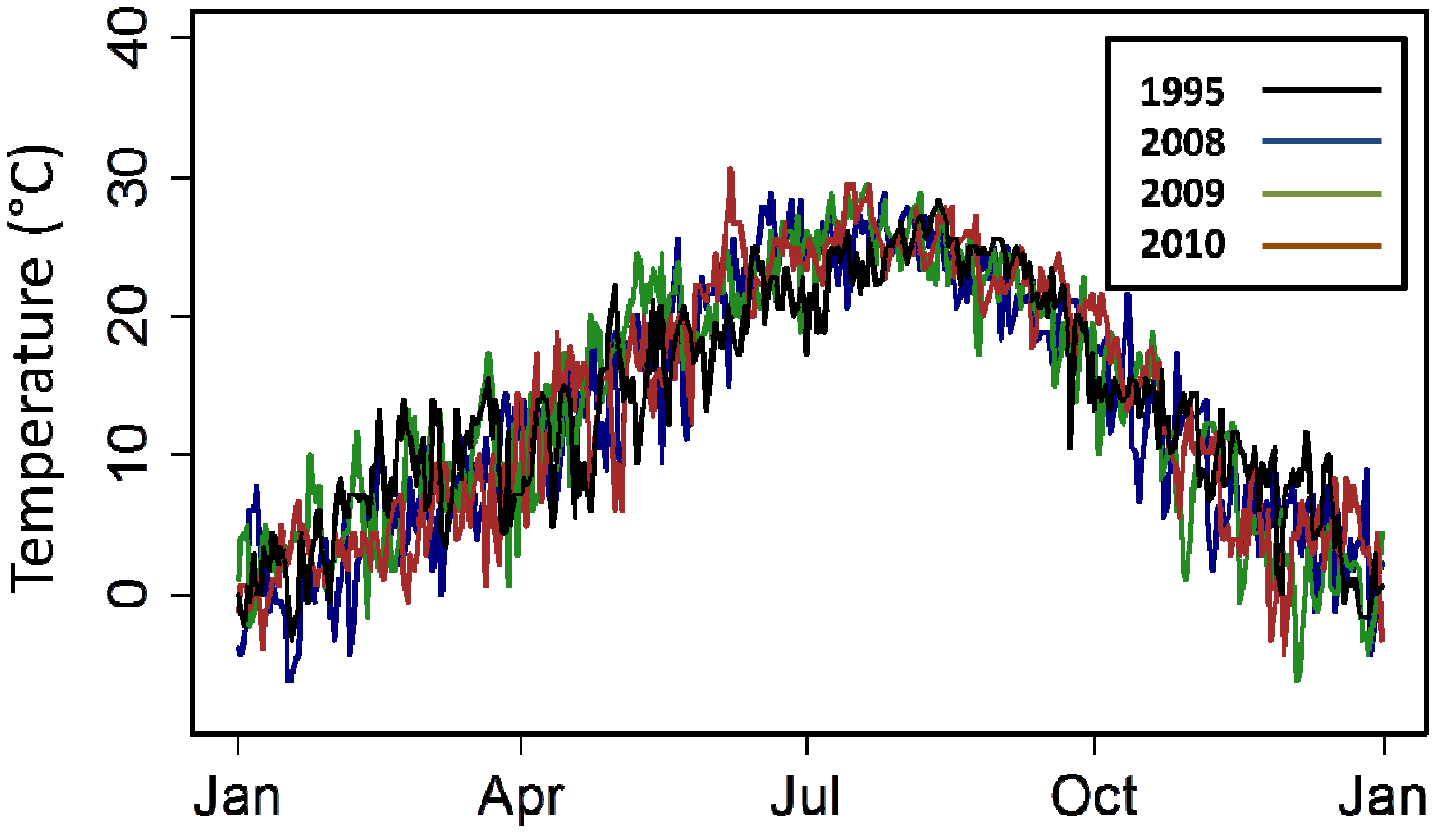


Figure 11. Discharge data for the Rio Grande for years 1995, 2008-2010, as well as mean historical values. Top panel illustrates discharge as a five day moving-average. Note the larger spring flood pulse runoff in 1995. Middle panel illustrates that 1995 was an overall wetter year than 2008 – 2010, as well as compared to the historical mean. Bottom panel: proportional discharge accumulation for water years 1995 and 2008-2010. The Q50 date is shown for each year. Q50 was approximately 21 days earlier in 1995 than 2008-2010.

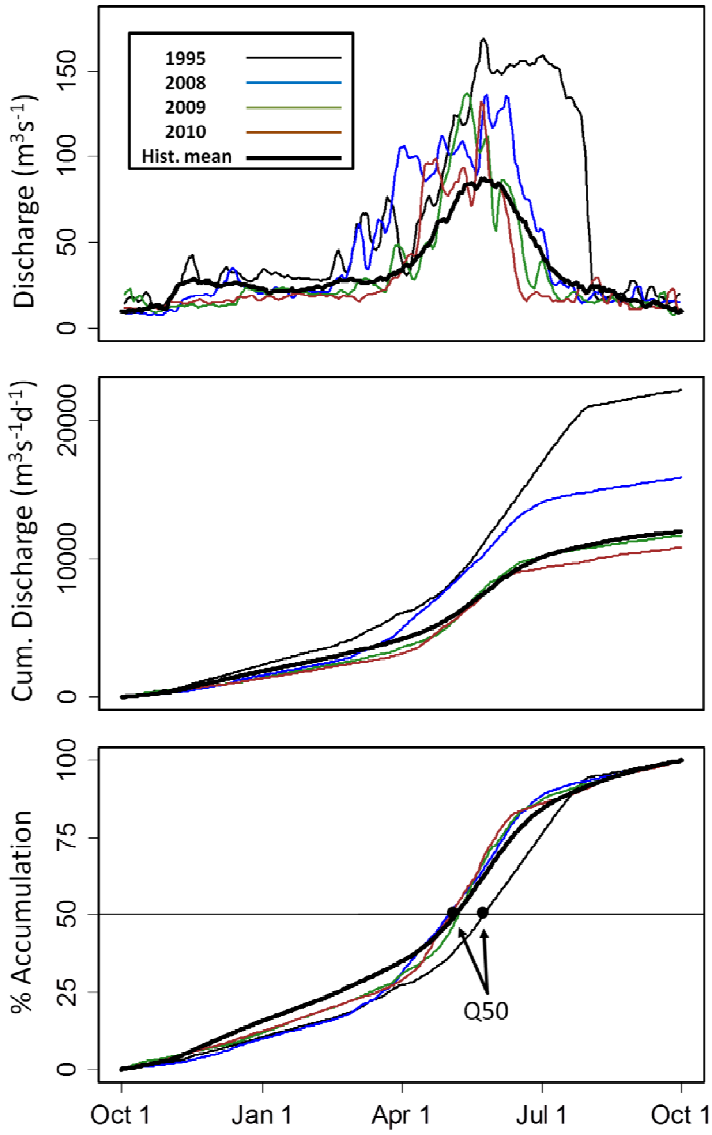
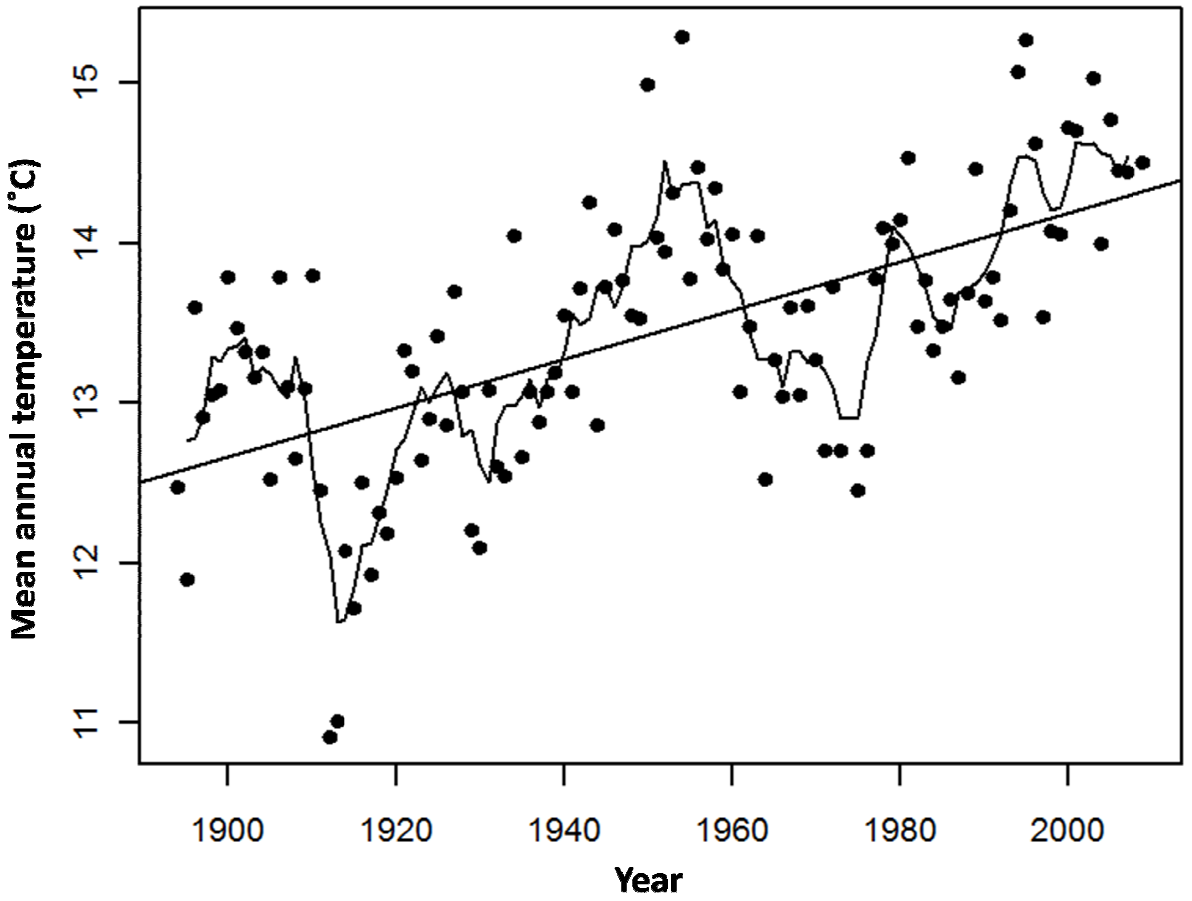


Figure 12. Mean annual air temperature data (˚C) for Albuquerque, New Mexico, from 1893 to 2010. Lines represent least-squares regression and five-year moving average.



Appendix A: Sample location information

Sandia Site.—NM: Bernalillo County, mouth of the North Diversion Canal at its confluence with the Rio Grande, Sandia Pueblo, Albuquerque UTM: Zone: 13S, Easting: 353302, Northing: 3897465

US HWY 60.—NM: Socorro County, Rio Grande, at U.S. HWY 60 bridge crossing, Bernardo. UTM: Zone: 13S, Easting: 334593, Northing: 3809728

Bosque del Apache.—NM: Socorro County, Rio Grande, directly east of Bosque del Apache National Wildlife Refuge Headquarters. UTM: Zone: 13S, Easting: 326882, Northing: 3740925

Albuquerque.—NM: Bernalillo County, Rio Grande, approximately 100 m upstream of Central Avenue Bridge in Albuquerque. UTM: Zone: 13S, Easting: 346597, Northing: 3884156

San Marcial.—NM: Socorro County, Rio Grande, 4.6 river miles upstream of Sierra County line. UTM: Zone: 13S, Easting: 304875, Northing: 3711741

Chapter 2: *CLOCK* GENE VARIATION: SEASONAL TIMING, PHYLOGENETIC SIGNAL, OR FUNCTIONAL CONSTRAINT?

Trevor J. Krabbenhoft and Thomas F. Turner

Abstract

Seasonal timing is a key aspect of reproductive strategies of organisms. Species often time reproduction differently, even within a common environment. This temporal partitioning of reproduction is one way in which organisms with overlapping resource needs can co-exist. Not surprisingly, in the present era of climate change research, much effort has been afforded toward trying to understand the genetic mechanisms underlying differences in reproductive phenology within and among species. Several recent studies have demonstrated latitudinal clines in allele lengths in a poly-glutamine (PolyQ) domain of a core circadian rhythm gene, *Clock*, that correlate with population-level migratory timing and reproductive seasonality in disparate organisms. In this study, we tested the hypothesis that PolyQ domain length of *Clock1a* correlates with differences in seasonal reproductive timing in five native and one introduced fish species (Teleostei: Cyprinidae) that co-occur in the Rio Grande, New Mexico, USA. We find that PolyQ mean allele length and most common allele length negatively correlate (r = -0.61 and -0.71, respectively) with reproductive seasonality among the five native, but not the introduced species. These values were not statistically significant, although there was low power associated with the small number of species in the test. We also found evidence of phylogenetic signal in PolyQ allele length, suggesting the intriguing possibility that phylogeny may drive seasonal reproductive timing in these species. Aside from PolyQ length variation in *Clock1a*, all other amino acids are conserved across the six species, despite more than 50 million years of divergence time - strong evidence of functional constraint in this gene. Finally, we tested the hypothesis that the federally-endangered Rio Grande silvery minnow (*Hybognathus amarus*) would exhibit less genetic variation in *Clock1a* than the five other (non-

imperiled) species. Consistent with this hypothesis, observed heterozygosity was approximately 2- to 8-fold lower in Rio Grande silvery minnow than each of the other five species, similar to previous observations of patterns of low genetic variation for putatively neutral markers that are thought to have resulted from recent population bottlenecks and range contraction. Alternatively, the low observed heterozygosity found in Rio Grande silvery minnow could be due to the relatively synchronous spawning season in this species compared to the other species which have more protracted spawning seasons. The correlation between spawning season length and observed heterozygosity is 0.55, partially supporting this hypothesis. In a broader context, these data beg the question of whether reduced genetic variation in this functionally-important gene may affect Rio Grande silvery minnow's capacity to respond to future environmental change.

Introduction

A nearly ubiquitous feature of long-lived organisms is the seasonal timing of life history events, such as reproduction. For many species, seasonal timing has shifted in recent years as a result of human-induced alterations to environmental conditions (e.g., climate change) (Walther et al. 2002; Parmesan 2006). Not surprisingly, the genetic mechanisms underlying seasonal timing of life history events have been a target of much recent empirical and theoretical work (Stinchcombe et al. 2004; Balasubramian et al. 2006; Tauber et al. 2007; Wilczek et al. 2010). Two key questions are: (1) how does natural selection shape seasonal timing, and (2) do populations possess the necessary genetic variation to respond to environmental change (Bradshaw and Holzapfel 2006; Reed et al. 2011). While phenotypic plasticity can mitigate some effects of environmental change, it is likely that genetic adaptation will also play a key role in whether populations are able to respond. For example, two frequent responses of organisms include altering seasonal timing or shifting latitudinal (or altitudinal) distribution. Over evolutionary time, both of these responses will likely require genetic adaptation in circadian and/or circannual rhythms.

A gene that has engendered much recent interest in studies of climate change and phenology is *Circadian Locomotor Output Cycles Kaput* (*Clock*), a key constituent of the core circadian oscillator. *Clock* encodes a protein, CLOCK, that is a member of the basic helix-loop-helix (bHLH)/Per-Arnt-Sim (PAS) family of transcription factors (Gekakis et al. 1998). CLOCK forms a heterodimer with a second protein, BMAL, which together act as a transcription-activating complex. CLOCK-BMAL positively regulates two genes, *Cryptochrome* and *Period*, by binding to E-boxes (CACGTG) present in their promoters, thus activating their expression (King

and Takahashi 2000). PERIOD and CRYPTOCHROME form a negative feedback loop by binding with CKIε/CKIδ and translocating to the nucleus to negatively inhibit transcription of CLOCK-BMAL (Reppert and Weaver 2002), thus completing the core cycle of circadian oscillation. Thus, the circadian clock allows organisms to anticipate, rather than respond to, daily environmental changes (Darlington et al. 1998).

The evolution of gene regulation is thought to occur largely by changes in either transcription factors (Hsia and McGinnis 2003; Lynch and Wagner 2008) or *cis*-regulatory elements (Prud'homme et al. 2007; Wray 2007). Thus, due to its role as a transcription factor, *Clock* is a potential target for natural selection to shape daily and perhaps seasonal rhythms. Changes in the amino acid sequence of *Clock* can have drastic phenotypic consequences. For example, a 51 amino-acid deletion in *Clock* in mice resulted in a 3-4 hour lengthening of circadian period (King et al. 1997). The transcription-activating potential of CLOCK is affected by a C-terminal polyglutamine repeat domain (PolyQ) present in most CLOCK proteins (Darlington et al. 1998). In *Drosophila*, a deletion in the PolyQ domain greatly reduced the affinity of *Clock* for its downstream targets, and thus resulted in longer circadian periodicity (Darlington et al. 1998), similar to that found in the mice mutants. The long series of amino acid repeats in *Clock* is significant in that long repeats are more likely to be polymorphic (Wren et al. 2000) and polymorphisms in functional genes are likely to have phenotypic consequences (Steinmeyer et al. 2009).

The functional importance of *Clock* is indicated by its general conservation across an extremely-broad phylogenetic spectrum, from corals to humans. Within vertebrates, the number of *Clock* paralogs is a product of whole-genome duplication events, and, in some cases,

lineage-specific losses of paralogs (Wang 2008). Human, mice, and chickens have two copies of *Clock*: *Clock1* and *Clock2* (also known as *Npas2*), using the nomenclature of Wang (2008). *Clock1* and *Clock2* are believed to be the result of a whole-genome duplication event early in the vertebrate lineage (Ohno 1970). Following the divergence of fish and tetrapod lines, an additional genome duplication event occurred in the teleost lineage (3R hypothesis, Amores et al. 1998; Postlewait et al. 1998) that is thought to have given rise to two paralogs of *Clock1*: *Clock1a* and *Clock1b* (Wang 2008). In zebrafish (*Danio rerio*), these paralogs were previously referred to as *zfClock1* and *zfClock3* (Ishikawa et al. 2002), prior to the understanding of their evolutionary origins. While *Clock1a* and *Clock1b* are both still present in zebrafish, fugu, and tetraodon, only *Clock1b* has been found in medaka and stickleback and *Clock1a* is presumed to have been secondarily lost in these species (Wang 2008). Because the functional similarities or differences of these paralogs are not well understood, the mechanisms facilitating persistence of each (e.g., subfunctionalization or neofunctionalization) in zebrafish, fugu and tetraodon are not known. Interestingly, the PolyQ domain is substantially longer in *Clock1a* than *Clock1b* in zebrafish, suggesting possible functional divergence. In fact, zebrafish *Clock1a* has among the longest *PolyQ* domain of known *Clock* genes (Saleem et al. 2001), begging the question of what this expansion means for the biology of these organisms.

Recently, several intriguing studies have examined the role of allele length polymorphism in circadian genes in shaping latitudinal clines in migration and reproductive seasonality. The earliest studies of clinal variation in circadian genes focused on another key gene, *Period*, in *Drosophila* (Costa et al. 1991, 1992). These studies revealed latitudinal variation in this gene, which was later linked to thermal compensation of circadian rhythms

(Sawyer et al. 1997). Weeks et al. (2006) expanded the search for clines in circadian genes to *Drosophila Clock*, but found no relationship between allele frequencies and latitude.

One of the first studies to tentatively link *Clock* and seasonal reproductive timing was by Leder *et al*. (2006), who mapped *Clock* to a quantitative trait locus in rainbow trout (*Oncorhynchus mykiss*) that explained up to 50% of the variance in spawning time in salmon. O'Malley and Banks (2008) subsequently demonstrated that *OtsClock1b* PolyQ length increases with latitude in populations of chinook salmon (*Oncorhynchus tshawytscha*), and correlates with migratory run and reproductive timing, while *OtsClock1a* is highly conserved among populations. Both *OtsClock1a* and *OtsClock1b* are homologous with zebrafish *Clock1a*, and arose from a salmon-specific genome duplication event (O'Malley and Banks 2008b; Paibomesai et al. 2010), sometimes referred to as the 4R hypothesis (Allendorf and Thorgaard (1984). Work in other salmon species (chum, coho, and pink; *O. keta*, *O. kitsutch*, *and O. gorbuscha*, respectively) suggested variable selection on *OtsClock1b* length across species, which corresponded to the extent of latitudinal variation in reproductive or migratory timing (O'Malley et al. 2010).

Studies of *Clock* polymorphism in birds have yielded mixed results. Johnsen et al. (2007) found clinal, latitudinal variation in *Clock* PolyQ domain length in non-migratory blue tit (*Cyanistes caeruleus*), but not in migratory bluethroat (*Luscinia svecica*) songbirds. Analysis of putatively neutrally-evolving microsatellite markers suggested that this pattern was not due to underlying demographic processes, but that variation in blue tit *Clock* results from positive selection. Within a single population of blue tit, Liedvogel et al. (2009) observed earlier reproductive seasonality in females, but not males, with shorter PolyQ domains. Additionally,

both sexes displayed shorter incubation time in individuals with shorter PolyQ domains. However, in a similar study, Liedvogel and Sheldon (2010) found low variability and no association between *Clock* polymorphism and variation in great tit (*Parus major*) inhabiting the same geographic area. These results were confirmed by Liedvogel et al. (2012), who also conducted quantitative genetic analyses that revealed very low additive genetic variation in timing of traits for these species. Finally, Dor et al. (2011) reported low levels of PolyQ variation among barn swallow populations, despite population genetic structure for microsatellite and mtDNA loci, which they attributed to possible stabilizing selection on *Clock*. Dor et al. (2012) expanded their work to five species of swallows (Genus *Trachycineta*), and found that *Clock* allele length did not correlate with latitude, nor did it correspond to date of clutch initiation or incubation duration.

Thus, there has been little consensus on the generality of *Clock* and reproductive timing across a diverse phylogenetic spectrum of organisms. However, this lack of generality could be due in part to the difficulty of detecting small, but significant, additive genetic variation effects on a quantitative character (Liedvogel et al. 2009). While these studies have greatly advanced our understanding of *Clock* in relation to latitude and seasonal timing, the picture is far from clear. Surprisingly-little attention has been paid, for example, to co-occurring communities of closely related organisms placed in a phylogenetic context.

Freshwater fish communities often exhibit strong temporal partitioning of reproductive seasonality, presumably as a means of minimizing competition among larvae during the critical period of early life. In the Rio Grande in New Mexico, USA, co-occurring fishes exhibit partially-overlapping, but distinct, spawning seasons (Turner et al. 2010; Krabbenhoft et al. 2012). A key

question is how human induced changes to the environment (e.g., alteration of flow regimes though dams or earlier spring flooding) will affect the phenology of reproduction. To answer this question, a firm understanding of the mechanisms controlling phenology is required. In this study, we tested three non-mutually exclusive hypotheses regarding variation in the PolyQ domain of *Clock1a* in six co-occurring Rio Grande cyprinid fishes (Teleostei:Cyprinidae): *(1) length variation corresponds to differences in reproductive phenology; (2) length variation is a result of phylogenetic inertia and reflects relationships among species; or (3) length variation reflects purifying selection for a certain circadian phenotype at the latitude where fishes were collected.*

If *Clock1a* plays a role in shaping seasonal reproductive timing, then we predict that earlier (colder) spawning species should have longer PolyQ domains. Alternatively, if *Clock1a* is not under divergent selection in these fishes but is instead under relaxed selective constraint, then length-polymorphism should track phylogenetic-relationships of these taxa. Finally, if *Clock1a* plays a role in non-reproductive circadian or circannual rhythms (e.g., adaptation to a particular latitude), then purifying selection could result in little among-taxon variation in these co-occurring fishes.

In addition to these hypotheses, we also tested whether the federally-endangered Rio Grande silvery minnow (*Hybognathus amarus*) has less genetic variation than the other five species. We base this hypothesis on two factors: (1) Rio Grande silvery minnow has experienced extensive population bottlenecks since European settlement and river regulation (Osborne et al 2005, 2012; Turner et al. 2006), and until recent reintroductions had occupied <10% of its native range in the Rio Grande (Bestgen and Propst 1996), and (2) Rio Grande silvery

minnow is a species with relatively synchronous spawning in response to flood pulses, versus the other species which have more protracted spawning seasons. We hypothesized that as a result of one or both of these factors, this species would possess less genetic variation in the functionally-important *Clock1a* gene than more widely-distributed, non-imperiled cyprinid fishes. In a broader context, the amount of genetic variation in functional genes (such as *Clock1a*) could play an important role in success or failure species to adapt to future environmental change (e.g., climate change).

Materials and Methods

Six species of Rio Grande fishes were included in this study: fathead minnow (*Pimephales promelas*), Rio Grande silvery minnow (*Hybognathus amarus*), red shiner (*Cyprinella lutrensis*), flathead chub (*Platygobio gracilis*), longnose dace (*Rhinichthys cataractae*) and common carp (*Cyprinus carpio*) (Table 1). Of these, the first five are native to the Rio Grande, while the last is introduced. The common carp is native to Asia and perhaps Europe, but was first transported to the U.S. in 1831 from European stock (McGeachin 1986) and by 1883 was introduced to New Mexico (Sublette et al. 1990). Individuals of these six species were collected with seine nets in the Rio Grande between Bernalillo and Los Lunas, New Mexico, USA. Fishes were sacrificed with an overdose of tricane methanesulfonate (MS-222) and brains were dissected and preserved in RNAlater (Ambion, Inc.) within five minutes of collection. RNA was extracted from brain tissues using TRIzol Reagent (Invitrogen, Inc.) and converted to complimentary DNA (cDNA) via reverse transcription PCR using Taqman Reverse Transcription Reagents (Applied Biosystems, Inc.) following the manufacturer's protocol.

Additionally, total genomic DNA was isolated from caudal-fin clips via phenol-chloroform extraction (Hillis et al. 1996).

PCR primers.—PCR primers were designed to amplify the PolyQ domain and adjacent sequence of *Clock1a* (Table 2, Fig. 1). First, degenerate PCR primers (named Clk-degAR and Clk-degEF) were designed based on conserved regions across vertebrate *Clock* genes based on the amino acid alignment in Wang (2008). NCBI's primer BLAST tool and Primer3 (Rozen and Skaletsky 2000) were used to assess primer quality, specificity, and melting temperature. In order to minimize degeneracy (e.g., below 250-fold), up to three inosines were incorporated in each degenerate PCR primer. Primers were designed to amplify *Clock1a* (i.e., to avoid amplifying *Clock1b*), while maintaining maximal conservation across fish taxa. These primers were used primarily for initial characterization of *Clock1a* in the six cyprinids. However, because these primers are complimentary to adjacent exons that span a long (>1000 bp) intron, they are only useful for amplifying cDNA. While we only tested the resulting primers in cyprinids, these are based on regions of *Clock* that are widely conserved across vertebrates and should be useful across a broad phylogenetic spectrum.

Once DNA sequences of *Clock1a* were obtained from degenerate primers, a second set of primers (Clk-GF and Clk-HR) was designed based specifically on the sequences of the six cyprinids in this study that spans the same region as the first primer set. Additionally, a third set of primers (Clk-MF and Clk-KR) was designed to amplify an approximately 260-290 bp region that encompasses the *Clock1a* PolyQ domain in these six species. Importantly, both the forward and reverse primers lie in the last exon of *Clock1a* (based on zebrafish genome sequence), and thus, amplify genomic DNA without interruption of introns. Additionally, for

fragment length analysis, a fluorescent HEX dye was attached to the 5' end of primer *Clk-MF* (see below).

Touchdown PCR.—Touchdown PCR was used to amplify *Clock1a* fragments. This procedure was necessary to prevent amplification of *Clock1b*, even with the non-degenerate primers. PCR was conducted in 10 µl reactions (1X buffer, 2 mM $MgCl_2$, 125 µM dNTPs, 5 pmol of each primer, 0.375 units of *Taq* DNA polymerase and approximately 375 ng template DNA) with the following conditions: 95˚C for 3 min, 20 cycles of 95˚C x 1 min, 59˚C x 1 min (decreasing by 0.5˚C each cycle), 72˚C x 1.5 min, then 21 cycles of 95˚C x 1 min, 53˚C x 1 min, 72˚C x 1.5 min and a final extension of 72˚C for 10 min.

Fragment length analysis.—In order to assess allele-length variation in *Clock1a*, fragment length analysis was conducted using 5' HEX dye labeled Clk-MF primer (and unlabeled CLK-KR primer). PCR product (1.0 µl) was combined with 10.0 µl formamide and 0.4 µl size standard (Genescan 400HD [ROX]), denatured at 95˚C for 5 min, and run out on an Applied Biosystems 3100 capillary sequencer. Fragment lengths were quantified using GeneMapper (Applied Biosystems). Based on this data, within-species mean allele length (MAL), most common allele (MCA), allele frequencies, number of unique alleles (k), and observed (H_o) and expected (H_e) heterozygosity were determined. Genotypic data were subjected to Guo and Thompson's (1992) test for deviation from Hardy-Weinberg Equilibrium (HWE), using Arlequin version 3.5 (Excoffier and Lischer 2010).

DNA sequencing.—Representative individuals (n=4) which were homozygous for the most common allele (MCA) for each species were selected for DNA sequencing. PCR products were purified using the E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, Inc.) and cycle sequenced using

ABI Prism Big Dye Terminator Cycle Sequencing Kit, Version 1.1 following the manufacturer's protocol. DNA sequencing was conducted on an Applied Biosystems 3100 capillary sequencer. Base calls were verified with Sequencer, version 4.9 (Gene Codes, Corp.), and sequences were aligned by ClustalW and by eye using BioEdit, version 7.0.9.0 (Hall 1999). Homology of these sequences with zebrafish *Clock1a* was confirmed with BLASTn searches against the nr and zebrafish RefSeq databases.

Mega5 (Tamura et al. 2011) was used to quantify relative synonymous codon usage rates (RSCU; Sharp et al. 1986) for all sequences to assess whether PolyQ domains are dominated by one glutamine codon over another (i.e., CAG versus CAA). Codon usage is important in determining whether PolyQ length is likely the result of DNA polymerase slippage, or is produced by other evolutionary forces (Wren et al. 2000; Johnsen et al. 2007).

Phylogenetic relationships.—Phylogenetic relationships for the six species of Rio Grande cyprinids, as well as the two other cyprinids for which *Clock1a* sequence data are available (zebrafish and Somalian cavefish, *Phreatichthys anduzzii*), were inferred from complete mitochondrial cytochrome *b* sequences obtained from Genbank. Cytochrome *b* was used because it is the only gene with sequence data available for all eight taxa and is generally well-suited for reconstructing inter-specific phylogenetic relationships. DNA sequences were aligned by eye and maximum likelihood analysis was conducted in Mega5 using the Tamura-Nei (1993) model of sequence evolution plus Γ-distributed rate heterogeneity. Tree support was assessed with 1000 bootstrap replicates.

Reproductive timing data.—Direct spawning observation is challenging, if not impossible, for fishes in turbid waters such as the Rio Grande. Consequently, we used the first

appearance of young-of-year (YOY) (e.g., protolarvae) as a proxy for onset of spawning season. Mean Julian date of first appearance of young-of-year was calculated based on data from Krabbenhoft et al. (2012). These data were based on 231 field collections over four years of young-of-year fish sampling on the Rio Grande, New Mexico, USA.

Relationship between PolyQ allele length, phylogeny, and reproductive timing.—We employed Mantel tests and phylogenetic comparative methods to assess whether reproductive timing (mean Julian date of first appearance of YOY) or phylogeny better explains PolyQ mean allele length differences among species. For Mantel tests, the cytochrome *b* sequences were used to calculate pairwise uncorrected p-distances between the six species (i.e., genetic distance). Similarly, we calculated Euclidian distance matrices for PolyQ MAL, MCA, and reproductive timing. Mantel tests were conducted on these four distance matrices using Ecodist package in R (http://www.r-project.org).

Results

Sequencing the Clock1a PolyQ domain.—We characterized a portion of *Clock1a* spanning the PolyQ domain in six cyprinid fishes. Homology with *Clock1a* was supported by BLASTn results, which revealed high similarity to other cyprinid *Clock1a* sequences. In particular, the Rio Grande cyprinids in this study were highly similar to Somalian cavefish and zebrafish (E-values < 10^{-5} in all cases), thus confirming these are orthologous genes. Conversely, DNA sequences in this study were far less similar to zebrafish *Clock1b* sequences (E-values > 1 in all comparisons). Translation of *Clock1a* DNA sequences revealed strong conservation at the amino acid level within and among species, as predicted, with all of the

amino acid differences arising from insertions and/or deletions of glutamines (Fig. 1). As expected, length variation among these sequences was due to glutamine indels in the PolyQ domain. There was only slight codon usage bias for the two glutamine codons, CAG and CAA: the CAG codon for glutamine was used approximately twice as frequently as CAA (Relative Synonymous Codon Usage = 1.32 and 0.68, respectively). Additionally, these codons tend to be interspersed, with a string of five consecutive CAGs representing the longest stretch without an alternate synonymous codon.

Fragment-length analysis: within-species variation.— Fragment-length analysis revealed little within-species variation in PolyQ allele length. Alelle length frequencies, MCA, MAL, and observed and expected heterozygosity are presented in Table 3 and Figure 2. Allelic richness ranged from just two to four in the six cyprinid species, with the MCA comprising between 50.0 and 96.8% of alleles in each species. Observed heterozygosity varied greatly among species: from 0.063 in Rio Grande silvery minnow to 1.000 in common carp. Interestingly, the federally-endangered Rio Grande silvery minnow, which has recently undergone extreme population bottlenecks and reduction in effective population size, has the lowest observed heterozygosity (H_o = 0.063) of all six species. Exact tests of HWE revealed a lack of deviation from Hardy-Weinberg expectation for all six species (p-value > 0.05 in all species).

Fragment-length analysis: among-species variation.— Fragment-length analysis revealed substantially longer PolyQ domains in the native Rio Grande cyprinids, with MAL ranging from 274.2 to 286.5 bp, as compared to zebrafish and Somalian cavefish, which have 257 and 237 bp alleles, respectively. Interestingly, common carp specimens collected in the Rio Grande are more similar in PolyQ length (MAL = 258.5) to zebrafish (an Asian native), than to

native Rio Grande species. Thus, carp have not converged on PolyQ allele lengths found in native Rio Grande cyprinids, but rather possess a more similar PolyQ length to other old-world cyprinids.

Phylogenetic analysis of mtDNA.—Results of a maximum likelihood analysis of cytochrome *b* mtDNA sequences are presented in Figure 3. While phylogenetic relationships within the family Cyprinidae are far from resolved, branching patterns in this tree generally agree with previous studies of cyprinid phylogenetics (e.g., Simons et al. 2003). Specifically, this tree revealed a reciprocally-monophyletic clade comprised of the five native North American species (fathead minnow, Rio Grande silvery minnow, red shiner, flathead chub, and longnose dace). Relationships within this North American clade are also congruent with the results of Simons et al. (2003).

Relation to environmental factors.— Mapping MCA for each taxon along the mtDNA phylogeny revealed strong congruence between mtDNA phylogeny and PolyQ allele length (Fig. 3, left). For example, old-world taxa (zebrafish, carp, and Somalian cavefish) formed a monophyletic grouping and all have short MAL (range 237 – 258.5 bp), whereas the North American species formed a clade and had longer MAL (range 274.2 – 286.5). Additionally, within the North American species there is a trend of longer allele lengths in more recently diverged species. Longnose dace, which is the most divergent species, has the shortest MAL of the North American species, while fathead minnow and Rio Grande silvery minnow have the longest. Finally, hierarchical cluster analysis of MAL resulted in a dendrogram that largely mirrored the mtDNA phylogeny, with some variation differences in internal branching patterns (Fig. 3, right), again suggesting phylogenetic signal in PolyQ allele lengths.

MAL and MCA do not correlate with reproductive timing (mean first appearance of larvae) in the six species in this study ($r = 0.156$, $p = 0.769$; and $r = 0.047$, $p = 0.930$, respectively). However, when the non-native common carp is removed from the analysis, this correlation becomes fairly strongly negative, although still not significant for either MAL ($r = -0.606$; $p = 0.279$) and MCA ($r = -0.705$; $p = 0.184$) versus reproductive timing (Fig. 4). Lack of significance in this instance may be due to lack of statistical power arising from the small sample size (n = 5 species). Thus, there appears to be a strong, but not statistically significant, trend of earlier spawning species having longer PolyQ domains than later spawning species.

Correlation between PolyQ allele length and MAL and MCA could be due to phylogenetic non-independence of these species, which we tested using Mantel tests. Mantel tests including all six species revealed a high, but not significant correlation for MAL versus genetic distance (uncorrected mtDNA p-distance) (Mantel $r = 0.756$; $p = 0.110$), and weak correlation for MAL versus reproductive timing (Mantel $r = -0.145$; $p = 0.823$). When common carp was removed from the analysis, correlations were still not significant (MAL vs. genetic distance: Mantel $r = 0.145$; $p = 0.703$ and MAL vs. reproductive timing: Mantel $r = 0.004$; $p = 1$).

As mentioned previously, Rio Grande silvery minnow had substantially lower observed heterozygosity than the other five species. While this may be due to population bottlenecks and low effective population size, this pattern is also consistent with this species having the most synchronous reproductive season relative to the other species which have more protracted spawning seasons (see Table 1). Indeed, the correlation between standard deviation of collection date of protolarvae (a proxy for spawning season length) and observed

heterozygosity is 0.54, suggesting lower heterozygosity in *Clock1a* could be partially due to differences in spawning season length among species.

Discussion

As researchers have raced to understand the effects of ongoing climate change, interest in phenology has burgeoned. Not surprisingly, the internal clocks of organisms have become a strong focus of functional genetics studies (Sawyer et al. 1997; Balasubramian et al. 2006; Bradshaw and Holtzapfel 2008). To that end, extensive progress has been made over the past decade toward understanding circadian pathways in laboratory and wild populations. However, substantive questions remain regarding the mechanistic connections between circadian pathways and other physiological processes, such as reproduction (Lincoln et al. 2006; Oliveira and Sánchez-Vázquez 2010). Despite circadian pathways being relatively well-studied, the molecular mechanisms linking circadian and seasonal rhythms are poorly understood. A central question for those studying the effects of climate change is how natural selection shapes reproductive phenology within and among species.

In this paper, we tested whether variation in the PolyQ domain in *Clock1a* is consistent with among-species differences in reproductive timing. Additionally, we tested whether PolyQ variation reflected phylogenetic relationships among species (i.e., phylogenetic inertia). We found a strong, negative correlation between reproductive timing and PolyQ allele length in the five native Rio Grande species. This correlation was not statistically significant, although this is likely due to the small number of species in the analysis and corresponding lack of statistical power. Surprisingly, the non-native common carp did not fit this trend, despite having been

introduced into New Mexico over a century ago. This contrasts with the findings of O'Malley et al. (2007) where PolyQ length in Chinook salmon introduced into New Zealand from the USA in the early 1900s appeared to be under divergent selection for different migratory timing. Three possibilities could explain the lack of carp convergence on native PolyQ lengths: (1) a lack of raw variation in the founding population in PolyQ allele length on which selection could act; (2) PolyQ length is a product of phylogenetic inertia, and not adaptation to local conditions or reproductive timing; or (3) other genes in the circadian pathway compensate for a shorter PolyQ domain length in *Clock1a* in carp; i.e., circadian rhythms in carp are "calibrated" differently. Common carp is tetraploid and has 2n = 100 chromosomes (versus 2n = 50 in the other cyprinids) as a result of a whole genome duplication event approximately 12 (David et al. 2003) or 5.6 to 11.3 million years ago (MYA) (Wang et al. 2012). Consequently, it is possible that dosage compensation or other factors could negate a need for convergence. At present, it is not known which of these possibilities is correct.

One additional caveat is that we compared PolyQ length with first appearance of offspring for a given year, rather than reproductive readiness. If *Clock1a* does play a role in reproduction, it most likely acts to trigger reproductive readiness (i.e., gonadal recrudescence) as photoperiod lengthens, rather than acting as a proximate driver of spawning. Thus, the relationship between *Clock1a* and reproductive timing might be blurred by species-specific differences in the lag between gonadal recrudescence and spawning and appearance of offspring. Regardless, the general trend is for longer PolyQ alleles in earlier spawning species. This trend is consistent with O'Malley and Banks' (2008) findings in Chinook salmon, where longer At first glance, this appears contradictory to O'Malley et al.'s (2008) findings of longer

PolyQ domains in later spawning fishes (in northern populations). In addition to differences in photoperiod, this pattern could also be driven by seasonally-varying water temperatures in the two systems. In the Rio Grande, earlier spawning species reproduce in colder temperatures than later spawning species. In Chinook salmon, the northern populations (which have longer PolyQ domains) presumably also reproduce in cooler temperatures (or lower degree days). It may be that in these ectothermic organisms, PolyQ allele length acts as a mechanism for temperature compensation. In ectotherms active in colder temperatures, longer PolyQ domains could compensate for temperature (and lower metabolic rates) by increasing the transactivation affinity of CLOCK for downstream targets (e.g., *Period*). Concordant with this hypothesis, in zebrafish, the amplitude of transcriptional activation by CLOCK is strongly temperature dependent (Lahiri et al. 2005). Future research is needed to determine how temperature and photoperiod interact to shape the evolution of *Clock* PolyQ domains. We argue that comparative studies of co-occurring taxa is one potentially fruitful avenue for addressing these issues.

PolyQ domain length and phylogeny. Mapping PolyQ allele length onto the mitochondrial DNA tree revealed an obvious correspondence between phylogeny and PolyQ length. However, the Mantel test for this relationship was not significant, perhaps due to the small number of taxa in this test. Given the apparent correspondence between PolyQ length and both reproductive timing and phylogeny, one intriguing possibility is that phylogeny could drive reproductive timing. In the native Rio Grande cyprinids in this study, for example, there is a general trend of earlier onset of spawning seasons in more derived species. Among-species differences in PolyQ length suggest that purifying selection at this latitude does not explain

overall patterns of length variation; e.g., the 278 bp allele is present in fathead minnow, but not dominant, as would be expected if there were purifying selection for a particular allele length at this latitude. However, aside from variation in the number of glutamine repeats, amino acid sequences in the PolyQ domain are strikingly conserved across very old phylogenetic splits. For example, aside from glutamine indels, all other amino acids were identical across the five native North American species in this study, despite a time to most recent common ancestor of approximately 56.9 million years for these species (Saito et al. 2011). This remarkable conservation strongly suggests functional constraint in this gene.

Data presented in this paper reveal longer PolyQ domains in North American species as compared to the more ancestral, old-world cyprinids. Additionally, there is an evolutionary trend toward longer PolyQ domains in more recently derived North American cyprinid fishes. Interestingly, this trend of increasing PolyQ length is opposite of the overall pattern of genome size reduction in more derived lineages of the cyprinid phylogeny. North American species in this study (or congeners) all have substantially smaller genomes than their old-world, more basal counterparts (data from Gregory (2012) and references therein). Patterns of relative synonymous codon usage for glutamines suggest that length variation in PolyQ in these species is probably not due to polymerase slippage over identical codons (Johnsen et al. 2007; Wren et al. 2000), and instead suggests positive selection for longer allele length in more-derived taxa.

Adaptation to climate change? While there is relatively little within-species variation in PolyQ allele length, this standing variation could prove important for future adaptation to climate change and shifting circadian rhythms. We tested the hypothesis that the federally-endangered Rio Grande silvery minnow has reduced amounts of allelic variation in *Clock1a*

relative to the other species. We predicted genetic diversity would be lower either because of population bottlenecks and range contraction in Rio Grande silvery minnow due in large part to human alterations of flow regimes over the past century, or because Rio Grande silvery minnow is a more synchronous spawner than the other species, and perhaps requires a tighter signal from *Clock1a*. As expected, observed heterozygosity in Rio Grande silvery minnow was substantially lower than the other five species. These data are concordant with previous studies of Rio Grande silvery minnow based on putatively-neutral genetic markers (Osborne et al 2005, 2012; Turner et al. 2006). However, low heterozygosity in Rio Grande silvery minnow is also consistent with the hypothesis that spawning season length is driven by among-individual variation in *Clock1a* allele length. Rio Grande silvery minnow has the most synchronous spawning season among the study species, and spawning season length is fairly strongly ($r = 0.54$) correlated with observed heterozygosity across species, which could suggest that variation in *Clock1a* is needed for differential reproductive phenology.

Regardless of whether *Clock1a* is important for reproductive timing, it is a functionally-important gene that plays a central role in circadian rhythms. As such, the amount of variation in *Clock* could be an important determinant in whether populations are able to respond to climate change. Several studies have suggested the importance of migration as a primary response to climate change (Parmesan and Yohe 2003; Lenoir et al. 2008; Sandel et al. 2011). However, it is possible that a lack of appropriate genetic variation in the circadian machinery (e.g., *Clock*) could act to constrain circadian adaptation or rapid shifts in migration to new latitudes, particularly in fragmented systems such as the Rio Grande. Rare alleles may be precisely those required for future response to climate change, and large populations may be

needed to prevent those rare alleles from being lost in the population, for example, via genetic drift. Analogously, in sticklebacks, certain *Ectodysplasin* alleles, present at very low frequencies in ancestral marine populations, rapidly spread through newly-derived freshwater populations where selection is thought to have favored their phenotypic products (reduced lateral armor plates) (Colosimo et al. 2005).

The relationship between *Clock1a* polymorphism, phylogeny, and reproductive timing observed in this study are correlative, and experimental manipulation, perhaps with inbred lines, will be necessary to test whether these patterns reflect causality. Regardless, in this study we demonstrate taxon-specific patterns of variation in a functionally important gene across co-distributed species.

Acknowledgements

We thank Han Wang for providing an amino acid alignment of *Clock* and for helpful discussions. Christiano Bertolucci kindly provided insight on sequencing *Clock*. This research benefitted from insight provided by Kathleen O'Malley, Charlie Cunningham, Scott Collins, Megan Osborne, Steven Platania, and Corey Krabbenhoft. George Rosenberg and the UNM Biology Molecular Facility provided valuable technical support. Jon Kawatachi assisted with genetic data collection and was supported by the NIH-funded Initiatives to Maximize Student Diversity. Financial support was provided by UNM Biology Graduate Research Allocations Committee (GRAC) funds to TJK. Samples were collected under New Mexico Department of Game and Fish permit #3015 and US Fish and Wildlife permit #TE038055-0. This research was

approved by Institutional Animal Care and Use Committee Protocol #10-100468-MCC and #10-100492-MCC.

Literature Cited

Allendorf FW, Thorgaard GH (1984) Chapter 1. Tetraploidy and the evolution of salmonid fishes. Pp. 1-46 *In*: Turner BJ (Ed.) *Evolutionary genetics of fishes.* Plenum Press, New York, NY, USA.

Amores A, Force A, Yan Y-L, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish *hox* clusters and vertebrate genome evolution. Science 282:1711-1714.

Balasubramian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, Weigel D (2006) The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. Nature Genetics 38:711-715.

Baird SF (1887) Report of the Commissioner for 1885. Part XIII. U.S. Commission of Fish and Fisheries, Washington, D.C.

Bestgen K, Propst DL (1996) Redescription, geographic variation, and taxonomic status of Rio Grande silvery minnow. Copeia 1996(1):41-55.

Bradshaw WE, Holzapfel CM (2008) Genetic response to rapid climate change: it's seasonal timing that matters. Molecular Ecology 17:157-166.

Costa R, Peixoto AA, Barbujani G, Kyriacou CP (1992) A latitudinal cline in a *Drosophila* clock gene. Proceedings of the Royal Society London, Series B: 250(1327):43-49.

Costa R, Peixoto AA, Thackeray JR, Dalgleish R, Kyriacou CP (1991) Length polymorphism in the threonine-glycine-encoding repeat portion of the *period* gene in *Drosophila*. Journal of Molecular Evolution 32(3):238-246.

Darlington TK, Wager-Smith K, Fernanda Ceriani M, Staknis D, Gekakis N, Steeves TDL, Weitz CJ, Takahashi JS, Kay SA (1998) Closing the circadian loop: CLOCK-induced transcription of its own inhibitors *per* and *tim*. Science 280:1599-1603.

David L, Blum S, Feldman MW, Lavi U, Hillel J (2003) Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. Molecular Biology and Evolution 20(9):1425-1434.

Doi M, Hirayama J, Sassone-Corsi P (2006) Circadian regulator CLOCK is a histone acetyltransferase. Cell 125:497-508.

Dor R, Cooper CB, Lovette IJ, Massoni V, Bulit F, Liljesthrom M, Winkler DW (2012) *Clock* gene variation in *Tachycineta* swallows. Ecology and Evolution 2(1):95-105.

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

Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacher LD, King DP, Takahashi JS, Weitz CJ (1998) Role of the CLOCK protein in the mammalian circadian mechanism. Science 280:1564-1569.

Gregory TR (2012) Animal genome size database. http://www.genomesize.com.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Hillis DM, Mable BK, Larson A, Davis SK, Zimmer EA (1996) Chapter 9. Nucleic acids IV: sequencing and cloning. Pp. 321-381, *In:* Hillis DM et al. (Eds), *Molecular Systematics*, 2nd Ed. Sinauer Associates, Sunderland, MA, USA.

Hsia CC, McGinnis W (2003) Evolution of transcription factor function. Current Opinion in Genetics and Development 13:199-206.

Ishikawa T, Hirayama J, Kobayashi Y, Todo T (2002) Zebrafish CRY represses transcription mediated by CLOCK-BMAL heterodimer without inhibiting its binding to DNA. Genes to Cells 7(10):1073-1086.

Johnsen A, Fidler AE, Kuhn S, Carter KL, Hoffmann A, Barr IR, Biard C, Charmantier A, Eens M, Korsten P, Siitari H, Tomiuk J, Kempenaers B (2007) Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies. Molecular Ecology 16(22):4867-4880.

King DP, Takahashi JS (2000) Molecular genetics of circadian rhythms in mammals. Annual Reviews in Neuroscience 23:713-742.

King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, Steeves TDL, Hotz Vitaterna M, Kornhauser JM, Lowrey PL, Turek FW, Takahashi JS (1997) Positional cloning of the mouse circadian *Clock* gene. Cell 89:641-653.

Krabbenhoft TJ, Platania SP, Turner TF (2012) Reduced flows in an arid-land river decrease temporal niche partitioning in the larval fish community. *In prep.*

Lahiri K, Vallone D, Babu Gondi S, Santoriello C, Dickmeis T, Foulkes NS (2005) Temperature regulates transcription in the zebrafish circadian clock. PLoS Biology 3(11):2005-2016.

Leder EH, Danzmann RG, Ferguson MM (2006) The candidate gene, *Clock*, localizes to a strong spawning time quantitative trait locus region in rainbow trout. Journal of Heredity 97(1):74-80.

Lenoir J, Gégout JC, Marquet PA, de Ruffray P, Brissse H (2008) A significant upward shift in plant species optimum elevation during the 20th century. Science 320:1768-1771.

Liedvogel M, Sheldon BC (2010) Low variability and absence of phenotypic correlates of *Clock* gene variation in a great tit *Parus major* population. Journal of Avian Biology 41:543-550.

Liedvogel M, Cornwallis CK, Sheldon BC (2012) Integrating candidate gene and quantitative genetic approaches to understand variation in timing of breeding in wild tit populations. Journal of Evolutionary Biology doi: 10.1111/j.1420-9101.2012.02480.

Liedvogel M*,* Szulkin M*,* Knowles SCL*,* Wood MJ*,* Sheldon BC *(*2009*)* Phenotypic correlates of *Clock* gene variation in a wild Blue Tit population: evidence for a role in seasonal timing of reproduction*.* Molecular Ecology 18*:*2444–2456*.*

Lincoln GA, Clarke IJ, Hut RA, Hazlerigg DG (2006) Characterizing a mammalian circannual pacemaker. Science 314:1941-1944.

Lynch VJ, Wagner GP (2008) Resurrecting the role of transcription factor change in developmental evolution. Evolution 62(9):2131-2154.

McDonald M (1887) Report on distribution of fish and eggs by the U.S. Fish Commission for the season of 1885-'86. Bulletin of the United States Fish Commission 6(1886):385-394.

McGeachin RB (1986) Chapter 3. Carp and buffalo. Pp. 43-55 *In:* Stickney RA (Ed.), *Culture of nonsalmonid freshwater fishes*. CRC Press, Boca Raton, FL.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3(5):418-426.

O'Brien C, Bradshaw WE, Holzapfel CM (2011) Testing for causality in covarying traits: genes and latitude in a molecular world. Molecular Ecology 20:2471-2476.

O'Malley KG, Banks MA (2008) A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) *Clock* gene: evidence for selection on PolyQ length variants. Proceedings of the Royal Society of London, B 275:2813-2821.

O'Malley KG, Banks MA (2008b) Duplicated *Clock* genes with unique polyglutamine domains provide evidence for nonhomologous recombination in Chinook salmon (*Oncorhynchus tshawytscha*). Genetica 132:87-94.

O'Malley KG, Camara MD, Banks MA (2007) Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). Molecular Ecology 16:4930-4941.

O'Malley KG, Ford MJ, Hard JJ (2010) *Clock* polymorphism in Pacific salmon: evidence for variable selection along a latitudinal gradient. Proceedings of the Royal Society of London, B 277:3703-3714.

Ohno S (1970) *Evolution by gene duplication*. Springer-Verlag, New York, NY, USA

Oliveira C, Sánchez-Vázquez FJ (2010) Reproduction rhythms in fish. Pp. 185-213 *In*: Kulczykowska E , Popek W, Kapoor BG (Eds.) *Biological clock in fish*. Science Publishers.

Osborne MJ, Benavides MA, Turner TF (2005) Genetic heterogeneity among pelagic egg samples and variance in reproductive success in an endangered freshwater fish, *Hybognathus amarus*. Environmental Biology of Fishes 73:463-472.

Osborne MJ, Carson EW, Turner TF (2012) Genetic monirotin gand complex population dynamics: insights from a 12-year study of the Rio Grande silvery minnow. Evolutionary Applications. DOI: 10.1111/j.1752-4571.2011.00235.x

Paibomesai MI, Moghadam HK, Ferguson MM, Danzmann RG (2010) Clock genes and their genomic distributions in three species of salmonid fishes: associations with genes regulating sexual maturation and cell cycling. BMC Research Notes 3:215.

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. Annual Review of Ecology, Evolution, and Systematics 37:637-669.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37-42.

Postlethwait JH, Yan Y-L, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z, Goutel C, Fritz A, Kelsh R, Knapik E, Liao E, Paw B, Ransom D, Singer A, Thomson M, Abduljabbar TS, Yelick P, Beier D, Joly J-S, Larhammar D, Rosa F, Westerfield M, Zon LI, Johnson SL, Talbot WS (1998) Vertebrate genome evolution and the zebrafish gene map. Nature Genetics 18:345-349.

Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. Proceedings of the National Academy of Sciences of the United States of America 104:8605-8612.

Reed TE, Schindler DE, Hague MJ, Patterson DA, Meir E, Waples RS, Hinch SG (2011) Time to evolve? Potential evolutionary responses of Fraser River sockeye salmon to climate change and effects on persistence. PLoS One 6(6):e20380.

Reppert SM, Weaver DR (2002). Coordination of circadian timing in mammals. Nature 418:935-941.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Pp. 365-386 *In*: Krawetz S, Misener S (Eds.) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, USA.

Saito K, Sado T, Doosey MH, Bart Jr. HL, Inoue JG, Nishida M, Mayden RL, Miya M (2011) Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Osteriophysi). Zoological Journal of the Linnean Society 161:633-662.

Saleem Q, Anand A, Jain S, Brahmachari SK (2001) The polyglutamine motif is highly conserved at the *Clock* locus in various organisms and is not polymorphic in humans. Human Genetics 109:136-142.

Sandel B, Arge L, Dalsgaard B, Davies RG, Gaston KJ, Sutherland WJ, Svenning J-C (2011) The influence of Lake Quaternary climate-change velocity on species endemism. Science 334(6056):660-664.

Sawyer LA, Hennessy JM, Peixoto AA, Rosato E, Parkinson H, Costa R, Kyriacou CP (1997) Natural variation in a *Drosophila* clock gene and temperature compensation. Science 278:2117-2120.

Sharp PM, Tuohy TMF, Mosurski KR (1986) Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Research 14(13):5125-5143.

Simons AM, Berendzen PB, Mayden RL (2003) Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Zoological Journal of the Linnean Society 139:63-80.

Steinmeyer C, Mueller JC, Kempenaers B (2009) Search for informative polymorphisms in candidate genes: clock genes and circadian behaviour in blue tits. Genetica 136:109-117.

Stearns SC (1992) *The evolution of life histories*. Oxford Press, New York.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. Proceedings of the National Academy of Sciences of the United States of America 101:4712-4717.

Sublette JE, Hatch MD, Sublette M (1990) *The fishes of New Mexico*. University of New Mexico Press, Albuquerque.

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599-607.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10:512-526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28:2731-2739.

Tauber E, Zordan M, Sandrelli F, Pegoraro M, Osterwalder N, Breda C, Daga A, Selmin A, Monger K, Benna C, Rosato E, Kyriacou CP, Costa R (2007) Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. Science 316:1895-1898.

Turner TF, Osborne MJ, Moyer GR, Benavides MA, Alò D (2006) Life history and environmental variation interact to determine effective population to census size ratio. Proceedings of the Royal Society London B 273:3065-3073.

Turner TF, Krabbenhoft TJ, Burdett AS (2010) Reproductive phenology and fish community structure in an arid-land river system. *In: Community Ecology of Stream Fishes* (Gido K, & Jackson D, Eds.). *American Fisheries Society Symposium* 73:427-446.

Walther G, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin J-M, Hoegh-Guldberg O, Bairlein F (2002) Ecological responses to recent climate change. Nature 416:389-395.

Wang H (2008) Comparative analysis of teleost fish genomes reveals preservation of different ancient clock duplicates in different fishes. Marine Genomics 1:69-78.

Wang J-T, Li J-T, Xiao-Feng Z, Xiao-Wen S (2012) Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). BMC Genomics 13:96.

Whitmore D (2010) Cellular clocks and the importance of light in zebrafish. Pp. 125-153 *in* E. Kulczykowska (Ed.), *Biological clock in fish*. Science Publishers.

Wilczek AM, Burghardt LT, Cobb AR, Cooper MD, Welch SM, Schmitt J (2010) Genetic and physiological bases for phenological responses to current and predicted climates. Philosophical Transactions of the Royal Society, Series B 365:3129-3147.

Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. Nature Reviews Genetics 8:206-216.

Wren JD, Forgacs E, Fondon JW, III, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Minna JD, Garner HR (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. American Journal of Human Genetics 67(2):345-356.

Table 1. Mean date of first spawn for Rio Grande cyprinids. Dates reflect mean Julian date of first appearance of young-of-year fish in samples from years 1995 and 2008-2010 (Krabbenhoft et al. 2012). The standard deviation column refers to the estimate of standard deviation for Julian date of collection of protolarvae and is used as a proxy for length of spawning season. Standard deviation was not estimated for *Rhinichthys cataractae* because of small numbers of protolarvae of that species collected.

Species	Common name	Mean Julian date, first appearance of YOY	St. dev. Julian date, protolarvae collection	Rio Grande native?
Cyprinus carpio	Common carp	126.4	20.6	Introduced
Hybognathus amarus	Rio Grande silvery minnow	132.5	7.2	Native
Pimephales promelas	Fathead minnow	138.5	21.2	Native
Platygobio gracilis	Flathead chub	142.6	14.4	Native
Cyprinella lutrensis	Red shiner	163.5	11.2	Native
Rhinichthys cataractae	Longnose dace	193.5	NA	Native

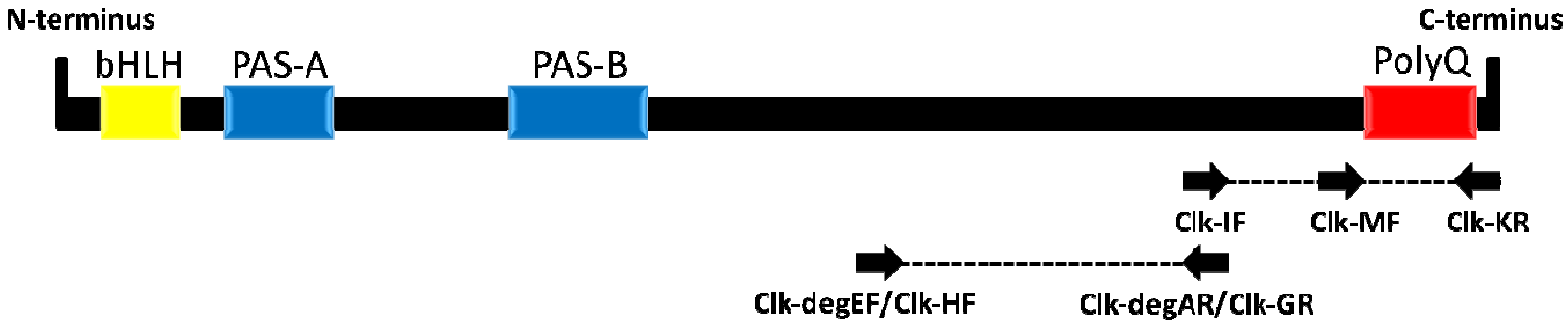
Table 2. PCR primers used in this study. For nucleotide sequences, the symbol "I" represents an inosine base.

Primer Name	Locus	Nucleotide Sequence (5' to 3')	Approx. Amplicon Size	Genbank Acc. no.
Clk-degAR	*Clock1a*(cDNA)	ACY TGI CCC ATR AAC ATI GTI GTN GGN ACC AT	630 bp	XXXX-XXXX
Clk-degEF	*Clock1a* (cDNA)	TIG ARC ARM GIA CNM GNA TGA T	630 bp	XXXX-XXXX
Clk-GR	*Clock1a* (cDNA)	ACC TGC CCC ATG AAC ATG GTG GTG GGA ACC AT	630 bp	XXXX-XXXX
Clk-HF	*Clock1a* (cDNA)	TGG AGC AGC GCA CAC GCA TGA T	630 bp	XXXX-XXXX
Clk-IF	*Clock1a*(cDNA)	ATG TGG GGC AGT TAT GGT KC	520 bp	XXXX-XXXX
Clk-MF	*Clock1a* (genomic)	CTC GGC TTT TGC ATG GCA ACC	260-290 bp	XXXX-XXXX
Clk-KR	*Clock1a* (genomic)	CTG TCK GAG CGA TGA GCT G	260-290 or 520 bp	XXXX-XXXX

Table 3. Summary of sample sizes, allele lengths, frequencies, and observed and expected heterozygosities in six species of cyprinid fishes in this study.

Species	N	k (unique alleles)	Mean allele length	Most common allele	H_o	H_e	Allele frequency						
							257	260	263	269	272	278	287
Fathead minnow	74	2	286.5	287	0.122	0.115						0.061	0.939
RG silvery minnow	190	4	277.8	278	0.063	0.062			0.003	0.003	0.003	0.968	
Red shiner	68	3	276.6	278	0.206	0.189			0.088		0.015	0.897	
Flathead chub	37	2	277.3	278	0.243	0.217					0.122	0.878	
Longnose dace	31	2	274.2	272	0.419	0.474					0.629	0.371	
Common carp	5	2	258.5	257/260	1.000	0.556	0.500	0.500					

Figure 1. Top Panel: Structure of *Clock1a* based on zebrafish (*Danio rerio*) amino acid sequence. Locations of conserved domains and regions complimentary to PCR primers are shown. The total length is 892 amino acids. Bottom Panel: Amino acid sequences of *Clock1a* PolyQ domain in cyprinid fishes and corresponding allele lengths (in bp). Glutamines are highlighted in grey. Amino acid conservation across the eight taxa is presented below the alignment.




```
Fathead minnow     287 AAFPLQQQGTFAAATQQQQQLQQQQQQLQQQHQQQQQQLQQHHQQQQQQLQQQQQQQLQQQQQQQQQQQQQQLQQQLAAHRSD
RG silvery minnow  278 AAFPLQQQGTFAAATQQQQQLQQQQQQLQQQHQQQQQQLQQHHQQQQQQLQQQQQQQLQQQQQQQQQQ---LQQQLAAHRSD
Red shiner         278 AAFPLQQQGTFAAATQQQQQLQQQQQQLQQQHQQQQQQLQQHHQQQQQQLQQQQQQQLQQQQQQQQQQ---LQQQLAAHRSD
Flathead chub      278 AAFPLQQQGTFAAATQQQQQLQQQQQQLQQQHQQQQQQLQQHHQQQQQQLQQQQQQQLQQQQQQQQQQ---LQQQLAAHRSD
Longnose dace      278 AAFPLQQQGTFAAATQQQQQLQQQQQQLQQQHQQQQQQLQQHHQQQQQQLQQQQQQQLQQQQQQQQQQ---LQQQLAAHRSD
Longnose dace      272 AAFPLQQQGTFAAATQQQQQLQQQQQQLQQQHQQQQQQLQQHHQQQQQQLQQQQQQQLQQQQQQQQQ-----LQQQLAAHRSD
Common carp        260 XAFPLQQQGTFAAATQQQQQQQLQQQQQQLQQQQQQQQQQQLQQQHQQQQQQLQQQQQ-QQQQQQQ--------XXXXXXXXXX
Zebrafish          257 AAFPLQQQGTFTTATQQQQQLHQQQQQLQQQQQLQQQQQQQQQQLQQQ-HQQQQQQ-LQQQHQQQ--------QQQLAAHRSD
Somalian cavefish  237 AAFPLQQQGTFATTTQQQQQHM------QQQHQQHHHQLQQQHHQQQQQLQQQQQQQLQQQ------------QQQLAAHRSD

CONSERVATION PLOT
```

Figure 2. *Clock1a* PolyQ allele frequencies for six Rio Grande fishes in this study.

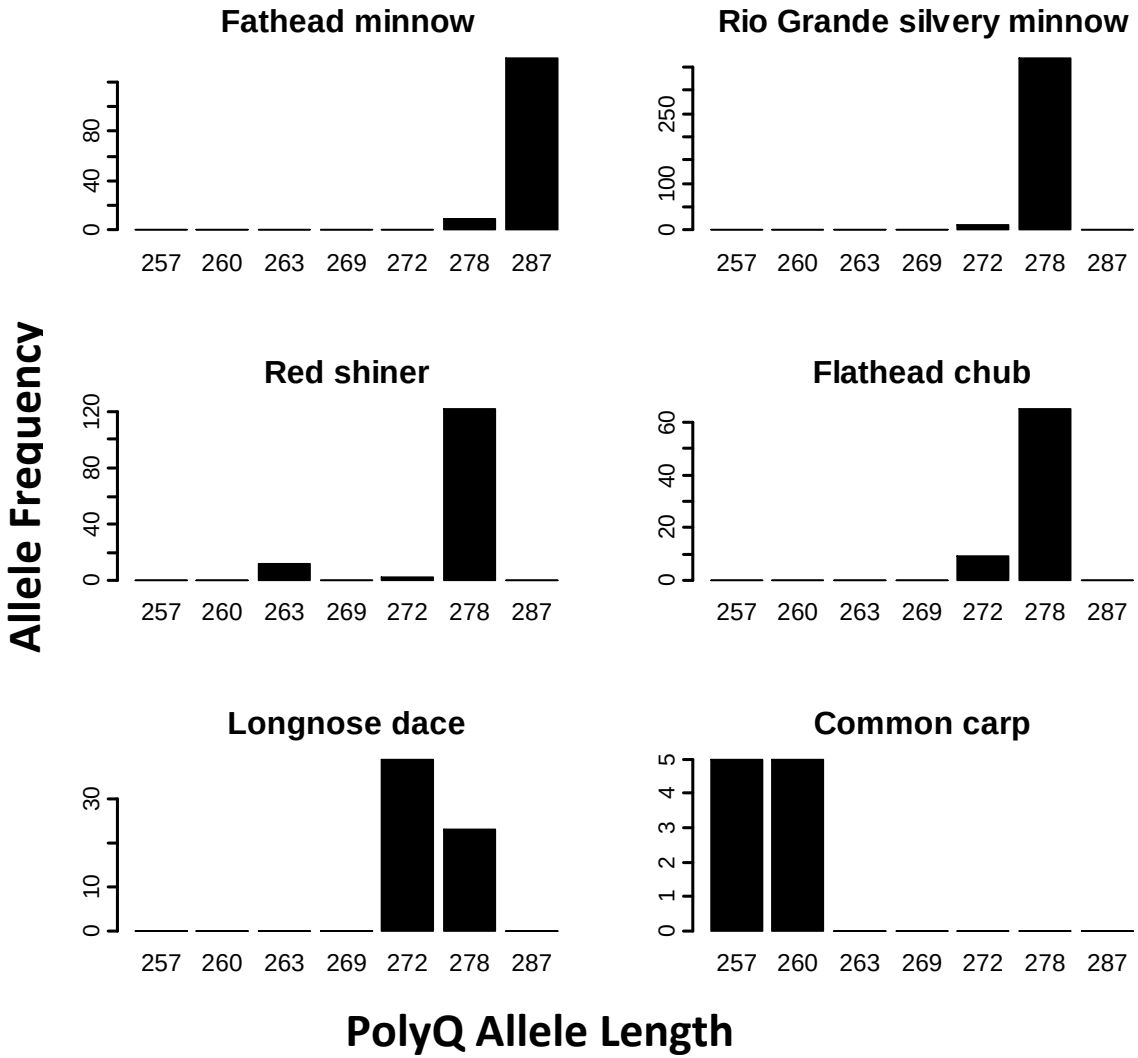


Figure 3. Left: Phylogenetic relationships of eight cyprinid fishes in this study as inferred from maximum likelihood analysis of complete mtDNA cytochrome *b* sequences, including the six Rio Grande focal species. Values represent bootstrap support (1000 replicates). The most common PolyQ allele length is plotted next to each taxon. Vertical bars represent hypothesized gains or losses of glutamines. Right: Dendrogram depicting relationships of taxa based on cluster analysis of PolyQ mean allele lengths. Note the general correspondence between the two trees, with some differences in internal branching patterns.

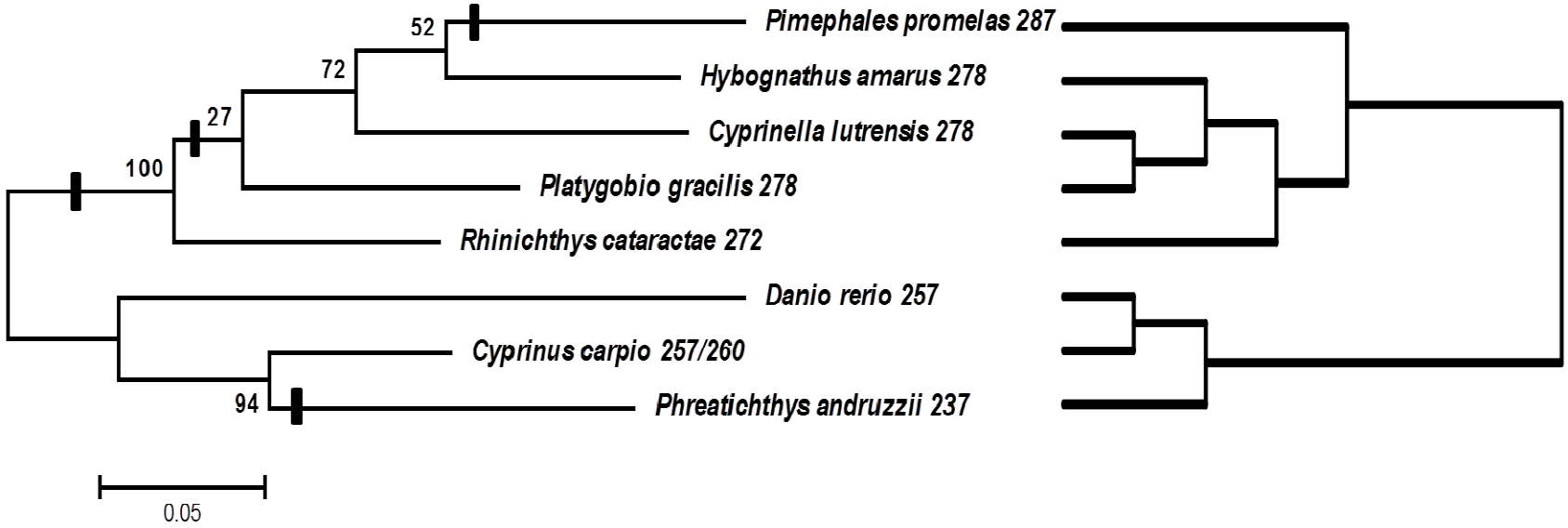
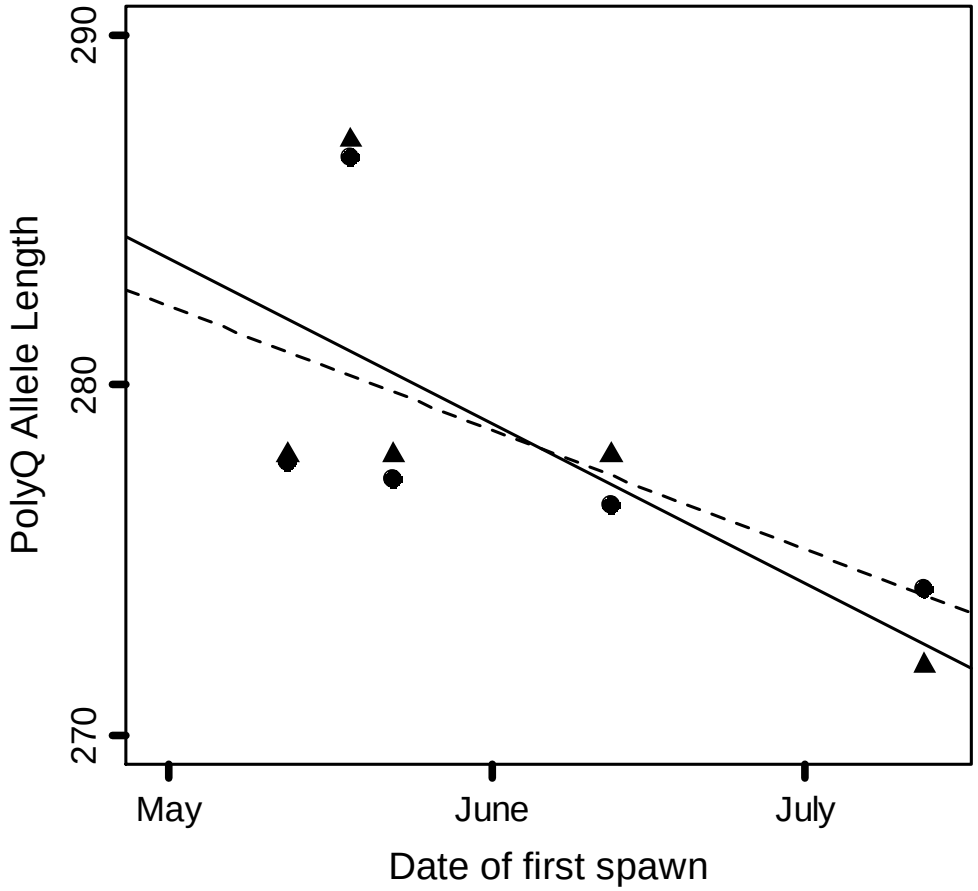


Figure 4. PolyQ mean allele length and most common allele versus mean date of first spawn in five native species of North American cyprinid fishes in this study. When the non-native common carp is removed, there is a strong, negative correlation between both PolyQ mean allele length (r = -0.61; dotted line) and most common allele (r = -0.71; solid line) versus first appearance of larvae (a proxy for reproductive timing).



Chapter 3: COMPARATIVE GENOMICS OF REPRODUCTION AND CIRCADIAN RHYTHMS IN FISHES: A NEXT-GENERATION TRANSCRIPTOME SEQUENCING APPROACH.

Trevor J. Krabbenhoft and Thomas F. Turner

Abstract

Next-generation DNA sequencing has paved the way for functional genomics studies in non-model organisms. We present results of an exploratory study aimed at identifying functional differences in genes associated with reproductive and circadian rhythms among three co-occurring fish species. We partially sequenced the transcriptomes of two cyprinid (Teleostei: Cypriniformes) fishes and combined sequences with an existing EST dataset for a third cyprinid fish, using zebrafish (*Danio rerio*) as a model reference organism. The resulting dataset of nearly two million sequence reads encompassed between 14,097 and 18,365 unique genes in each species, compared to approximately 30,651 genes in the zebrafish genome. Surprisingly, only between 42.91 and 76.30% of reads mapped to zebrafish, suggesting significant evolutionary divergence between these cyprinid fishes. In an effort to explore functional diversity in reproductive physiology and circadian rhythms, we identified and compared 86 genes of putative importance for reproductive physiology or circadian rhythms. A total of 43,213 reads mapped to these genes across the three species. Sequence level analysis of these genes revealed 342 variable amino acid sites among the three fish species. Of the 342 substitutions, 105 result in changes of electrical charge, while 63 resulted in change from polar to non-polar, or *vice versa*. Additionally, fifteen simple sequence amino acid repeat regions in candidate genes were identified as potential targets of divergent selection. Finally, red shiner RNA-seq data were compared with reproductive status of individual fish, and genes with expression patterns that most strongly (positively or negatively) correlated with reproductive activity were identified. Expression and sequence-level differences in candidate genes identified in this study are potential foci of selection that may help shape the reproductive and

circadian biology of these species. With over 2,400 species, the family Cyprinidae represents among the largest radiations of vertebrates. It is likely that functional genomic differences have played an important role in driving the diversification within this clade.

Introduction

Next-generation DNA sequencing has had rapid and transformative impact on functional genetics studies of non-model organisms. *De novo* transcriptome characterization in non-model species is now feasible (e.g., Abril et al. 2010; Parchman et al. 2010), including in large and complex vertebrate genomes (Hale et al. 2009; Yúfera et al. 2012; Seeb et al. 2011). Next-generation sequence data has informed a wide variety of ecological and evolutionary questions, from studies of local adaptation in heterogeneous environments (Renaut et al. 2010), to studies of pathogen resistance or susceptibility (Rosenblum et al. 2012), to unraveling genome duplication events (Wang et al. 2012). Once a considered a black box, the biochemical complexity of non-model organisms is being illuminated by the wealth of genomics data that is rapidly accumulating, offering profound insight into molecular adaptation. Functional genomics or transcriptomics studies of non-model, but ecologically-important organisms can benefit greatly from comparison with closely-related model species; e.g., those for which an annotated genome sequence is available. The primary advantage of having a closely-related model species with which to make comparisons is that many of the physiological pathways of interest will have already been studied in detail, and one can take advantage of existing knowledge of the sequences, structures and functions of those genes and physiological pathways. For the post-genomics revolution to reach full fruition, we must bridge the gap between existing genomics resources for model 'laboratory' organisms (i.e., zebrafish, fruit flies) and ecologically-important organisms in nature for which few resources currently exist. In this vein, comparative studies of closely-related taxa can yield powerful insight into how genome evolution has occurred across a phylogeny and offer empirical support for laboratory or *in silico*

models of gene function and physiological pathways (e.g., KEGG maps) in model species (Sarropoulou and Fernandes 2011). Additionally, by studying closely-related taxa in common environments, we can better understand how evolutionary processes have shaped genomes and the resulting trait-space to fill ecological niches. We view this approach as holding great promise for understanding evolutionary innovation and divergence of functional traits. Finally, patterns of gene expression, physiology, and behavior in nature can differ from findings from laboratory experiments, due perhaps to the oversimplification of laboratory experiments (by design) (e.g., Vanin et al. 2012). As a general rule, whenever possible gene functions should be validated in organisms in their natural environment.

In this study, we focus on two of the better-characterized physiological processes of importance to all multi-cellular organisms: reproduction and circadian rhythms. Not surprisingly, given their central importance to life, these broad physiological processes have long engendered focused research efforts by biologists and represent some of our most detailed understanding of organismal physiology (Dunlap 1999; King and Takahashi 2000; Reppert and Weaver 2002; Neill 2005). For example, the core circadian pathway has been well characterized across disparate organisms, from fruit flies to humans, and incredible conservation of key functional elements across hundreds of millions of years of evolution has been observed (e.g., PAS-domains of *Clock* genes; King et al. 1997). Similarly, reproductive pathways have been well studied and homologous genes often perform broadly similar functions across evolutionarily divergent organisms (Spargo and Hope 2003; Tsai 2006). Less well understood, however, are the genetic mechanisms underlying how organisms "tune" these systems to match their particular needs across heterogeneous landscapes or to fill new niches

over evolutionary time (Yerushalmi and Green 2009). For example, what are the evolutionary processes and genetic mechanisms that drive among-species differences in seasonal timing of reproduction (Johnsen et al. 2007; O'Malley and Banks 2008; Krabbenhoft and Turner 2012)? How do seasonal reproductive cycles cue off changing photoperiod information entrained in the circadian machinery? A necessary first step in addressing these questions is to characterize variation in genes and regulatory regions among species or populations that differ in key components of reproductive biology.

We used next-generation 454 DNA sequencing and existing expressed sequence tag (EST) data to characterize and compare the transcriptomes of three ecologically-important fish species from the family Cyprinidae (Teleostei: Cypriniformes). The family Cyprinidae is among the largest families of vertebrates, with well over 2,400 species (Nelson 2006), and whose members exhibit a wide array of reproductive behaviors and life history strategies (Johnston and Page, 1992; Johnston 1999). The three study species include red shiner (*Cyprinus lutrensis*), fathead minnow (*Pimephales promelas*), and Rio Grande silvery minnow (*Hybognathus amarus*). All three species are native to and co-occur in the Rio Grande, New Mexico, USA. Fathead minnow and red shiner are widely distributed and often among the most abundant fish taxa where they occur, while the Rio Grande silvery minnow is endemic to the Rio Grande and has highly variable population densities. The three study species vary in several key aspects of their life history strategies and breeding behaviors, and posses varying amounts of sexual dimorphism (Sublette et al. 1990). In particular, red shiner is a crevice spawning species with adhesive demersal eggs and no parental care (Gale 1986), fathead minnow exhibits male nest guarding and parental care of adhesive demersal eggs (Markus 1934), and Rio Grande silvery

minnow is a pelagic broadcast-spawner with drifting, semi-buoyant eggs and no parental care (Platania and Altenbach 1998). Additionally, there appear to be consistent differences in the reproductive phenology of these species (Turner et al. 2010; Krabbenhoft et al. 2012). In the Rio Grande, for example, Rio Grande silvery minnow in early spring as snowmelt runoff occurs and is followed shortly after by fathead minnow, while red shiner has a late and protracted spawning season relative to the other taxa (Krabbenhoft et al. 2012). A key question is how selection shapes the seasonal timing and reproductive biology of these three co-occurring, closely-related taxa.

Importantly, these fishes are relatively-closely related to another cyprinid fish – the zebrafish (*Danio rerio*), a model organism in developmental and neurobiology and a species with a completely sequenced and well-annotated genome (Sprague et al. 2006; Wellcome Trust Sanger Institute [http://www.sanger.ac.uk/Projects/D_rerio/]). We compared the transcriptomes of the three species using the zebrafish genome as a reference. In particular, we sought to delineate functional genetic variation that may underlie among-species differences in reproductive biology.

One of the first steps in understanding differences in complex phenotypic traits is characterizing the amount, distribution, and functional significance of variation present. We used a candidate gene approach to quantify among-species variation in genes involved in reproduction or circadian rhythms. Specifically, we assessed and compiled a list of amino-acid sites that differed among species, particularly amino acid substitutions resulting in change in electrical charge or polarity. While non-synonymous substitutions can either be due to relaxed purifying selection at a site, functional-neutrality or equivalency, or positive (divergent)

selection, in the present study our interest laid primarily in the latter. Additionally, we also sought simple sequence amino acid repeats in candidate genes. Such regions can be potential targets of selection and are often highly variable (Wren et al. 2000; O'Malley and Banks 2008; Steinmeyer et al. 2009). For example, a poly-glutamine repeat domain in the transcription factor *Clock* affects its DNA-binding affinity for downstream targets, and may be subject to divergent selection in a variety of organisms (Johnsen et al. 2007; O'Malley and Banks 2008, 2010; Caprioli et al. 2012).

While extensive research on model organisms such as zebrafish has greatly advanced our understanding of genome structure, function, and physiological pathways, it is clear that single species studies are insufficient for explaining the full breadth of ecological diversity in speciose clades that arises from divergent selection and evolutionary radiation.

Materials and Methods

We compared transcriptomic sequence data for three species of North American fishes (family Cyprinidae): red shiner (*Cyprinella lutrensis*), fathead minnow (*Pimephales promelas*), and Rio Grande silvery minnow (*Hybognathus amarus*) (Table 1). These species are of interest because they are widely distributed across North America (except Rio Grande silvery minnow) and frequently co-occur at the same locations. In the Rio Grande in New Mexico, USA, all three species co-occur and represent a substantial portion of the fish fauna in that river, in terms of both biomass and individuals. For fathead minnow, a total of 253,342 publically-available, expressed sequence tags (ESTs) were mined from NCBI's dbEST (Boguski et al. 1993). The

fathead minnow data were from several distinct EST projects and were based on a variety of tissue types, including brains, liver, and gonads. Data for the other two species (Red shiner and Rio Grande silvery minnow) were the result of two full 454 cDNA sequencing runs described below.

Tissue collection

For red shiner, fishes were collected with seine nets in the Rio Grande at Central Avenue bridge, in Albuquerque, New Mexico between March and April 2010, and transported live to aquaria. Red shiner specimens were held in 75 l. aquaria for three to eight weeks as part of a spawning experiment (Krabbenhoft and Turner, unpublished data). Twelve individual fish (ten adult females and two adult males) were captured from aquaria with dip nets. Fish were weighed (wet mass), measured (standard length), sacrificed and dissected within five minutes. Gonads and brains were also weighed (wet mass) and preserved separately in RNAlater (Ambion, Inc.).

Rio Grande silvery minnow samples were comprised of wild-caught fish held in aquaria at the Albuquerque BioPark Aquatic Research Facility as part of a captive breeding program. Fishes were processed as described for red shiner, except that Rio Grande silvery minnow tissue samples included liver instead of brain tissues.

RNA Extraction and cDNA Library Synthesis

Total RNA was extracted separately from brain and gonads of twelve red shiner individuals, including ten adult females and two adult males. Total RNA extractions were conducted with the RNA mini kit (Invitrogen, Inc.) following the manufacturers specifications. Ribosomal RNA was removed from total RNA samples using the MPG® mRNA Purification Kit

(Pure Biotech, LLC). RNA quality and amount of rRNA contamination were assessed using Pico kits for an Agilent 2100 Bioanalyzer. While Bioanalyzer results gave a rough estimate of mRNA concentration, samples were also quantified using a RiboGreen fluorometer. Because brain tissues did not yield sufficient quantities of mRNA for cDNA library construction, we pooled brain and gonad samples for each of the twelve samples prior to cDNA library construction. Complementary DNA synthesis and 454 FLX library construction were conducted by the UNM Department of Biology Molecular Facility following 454 Rapid Library preparation protocols. The twelve libraries were individually "barcoded" using standard 454 multiplex identifier adapters (MIDs). A full FLX + sequencing run was conducted on these 12 libraries, with six libraries loaded on each side of the sequencing plate. DNA concentration per library was quantified using the 454 FLX Library Quantification Kit (Kapa Biosystems, Inc.) and quality was assessed using High Sensitivity DNA Kits run on an Agilent Bioanalyzer.

Rio Grande silvery minnow tissues were shipped on dry ice to Ecoarray, Inc. (Ecoarray.com) for RNA isolation and normalized cDNA construction. A single normalized cDNA library was constructed based on equal amounts of RNA from six individuals and tissues. Total RNA was DNAse treated to minimize genomic DNA contamination. Full length double stranded cDNA was then produced using the Evrogen MINT cDNA Synthesis kit and Encyclo PCR buffer/polymerase. The library was normalized using Evrogen Trimmer Kit and was amplified and purified with QIAQuick PCR Purification Kit (Qiagen, Inc.). The resultant normalized cDNA library was shipped to Duke Institute for Genome Sciences and Policy, Sequencing Core Facility for 454 FLX sequencing on a full plate. At the time this sequencing was conducted, 454

titanium chemistry was not supported for cDNA samples, and thus, 454 FLX sequencing was conducted.

Mapping to Zebrafish Reference Transcriptome

Zebrafish (*Danio rerio*) genes and DNA sequences were downloaded from Ensembl (ensemble.org) release 60 via Biomart. In particular, cDNA sequences for each stable canonical transcript ID, along with gene names and gene ontology (GO) terms were downloaded. Thus, sequences include coding regions as well as 5' and 3' untranscribed regions (UTRs). This ENSEMBL dataset was used as the reference genome to which our sequence data were mapped. This dataset included 30,651 unique canonical transcripts, representing essentially the entire zebrafish transcriptome. For convenience, we will refer to these as "genes" throughout this manuscript.

Unassembled sequence reads were mapped to zebrafish canonical transcripts using 454 Refmapper software (version 2.6). Data for each species were mapped in separate Reference Mapper projects using default parameters, i.e., minimum overlap identity = 90, minimum overlap length = 40, seed length = 16. In the case of red shiner, each individual was mapped separately to zebrafish by first parsing reads out by MID tags, thus allowing individual-specific analyses. The resulting alignments were used to assess how many different genes were sequenced in different datasets, and to assess the efficacy of normalization in Rio Grande silvery minnow.

Next, we used a candidate gene approach to assess sequence-level variation in genes putatively associated with reproduction or circadian rhythms. Candidate genes were selected by (1) GO terms; i.e., genes with the terms "reproduction", "oocyte development", "circadian

rhythms", "photoperiodism", etc., were included; (2) genes identified in the literature as functionally-relevant to reproduction and circadian timing; and/or (3) genes which were expressed substantially higher in females than male red shiners in the aquarium experiment described above. We argue that this multi-pronged approach for identifying candidate genes is more robust than selecting candidate genes by a single feature (e.g., from GO terms alone).

Candidate genes were extracted from the zebrafish canonical transcripts dataset, and used as the reference sequences for a second round of 454 Reference Mapper alignments. Again, sequence reads for each of the three species were independently mapped to the zebrafish reference (in this case to the candidate genes) using default mapping parameters. All resultant contigs were extracted from Reference Mapper results files, renamed according to species, and pooled across species. Next, multiple alignments of candidate genes were manually constructed in Bioedit (Hall 1999). In cases where multiple, non-overlapping, contigs were present for a given gene within a species, a consensus sequence was made. DNA sequence alignments were translated and a list of non-synonymous DNA substitutions, i.e., substitutions which resulted in an amino-acid change, was compiled. Because the quality of 454 sequences frequently declines near the ends of reads, we excluded putative amino acid substitutions if they were within three residues from the end of the contig. Additionally, homopolymer runs longer than three nucleotides were also excluded from inferred amino acid substitutions, given high rates of error associated with 454 sequencing of homopolymers. Only substitutions within the three North American species were included in the list; i.e., amino acids that only differed between zebrafish versus all three North American species, were ignored. Indels, however, were included in the list. Changes in polarity and electrical charge resulting

from amino acid substitutions were also assessed, as these may have functional implications. Additionally, changes to- or from- "special case" amino acids were also quantified, i.e., proline and glycine changes, because these amino acids have unique molecular structures. Finally, we compiled a list of amino-acid repeat regions in candidate genes, because these regions are more likely to be variable and may have functional significance, and thus be a target for natural selection (Wren et al. 2000; Steinmeyer et al. 2009). Amino-acid repeats longer than five residues were included in the list.

Red Shiner Gonadosomatic Index and RNA-Seq

We compared the 454 sequence data for the twelve red shiner individuals with respect to reproductive status, as indicated by gonadosomatic index (GSI). Gonadosomatic index is calculated by the formula GSI = W / B * 100, where W is the mass of both gonads and B is the mass of the fish. Importantly, GSI is strongly tied to reproductive condition, particularly in females. For example, in red shiner, GSI differs significantly among histological stages of oocyte development (Brewer et al. 2008), and is thus a reliable indicator of reproductive status in female red shiner.

Sequence reads for each individual were summed according to the zebrafish canonical transcript to which they mapped and were divided by the total number of sequence reads per individual (i.e., relative "expression"). The ten females were rank ordered according to GSI and Spearman rank correlation coefficients (ρ) were calculated between transcript abundance and GSI. Only those genes for which at least ten total reads were present were included in the analysis. Genes which were highly positively or negatively correlated with GSI were identified. Spearman rank order correlation was used because it does not assume linear response

between transcript abundance and GSI. Additionally, Spearman correlations were also determined for those candidate genes for which at least ten red shiner sequences were present. Correlation coefficients were compared to the empirical distribution of correlations for all genes, and corresponding p-values were determined given the rank of each observed correlation relative to the total number of genes. No multiple-comparisons correction was made given the exploratory nature of this analysis and results should be viewed with caution. For example, this test is likely to overlook genes with a short window of expression during reproduction, i.e., those which are only transiently expressed.

Results

Sequencing and mapping results

A summary of sequencing results is provided in Table 2. In all, nearly two million sequence reads were included in this dataset. In addition to the fathead minnow EST sequences, we also obtained a total of 345,703 reads in Rio Grande silvery minnow, and 1,137,383 reads in red shiner (Table 2). A histogram depicting the distribution of lengths of sequence reads is presented in Figure 1. Differences in the distribution of sequence lengths among species are due to differences in sequencing platform and library preparation. For the three species, between 14,097 and 18,365 unique genes were partially or completely sequenced, as indicated by the Reference Mapper results. This likely represents approximately half to two thirds of the total number of genes in their genomes, given that zebrafish has 30,651 canonical transcripts and the same number of chromosomes as the North American species. Between 42.91 and 76.30% of reads mapped to zebrafish canonical transcripts. The

percent of reads mapped appears to be related to sequence length, with shorter 454 FLX reads for Rio Grande silvery minnow mapping less frequently than 454 FLX+ Titanium red shiner reads and fathead minnow Sanger ESTs. The high percentage of unmapped reads suggests either lack of evolutionary conservation or genomic DNA contamination.

Candidate genes: Reproduction and circadian rhythms

We selected 86 candidate genes with functions in zebrafish putatively associated with circadian rhythms or reproduction (Table 3). A total of 36,382, 5,162, and 1,669 reads mapped uniquely to these candidate genes in red shiner, fathead minnow, and Rio Grande silvery minnow, respectively (Table 2).

Amino acid substitutions

Sequence analysis of the 86 candidate genes yielded 342 putative amino acid substitutions in these genes among the three North American species (Appendix 1). The number of amino acid substitutions within each gene varied substantially across the 86 genes (Fig. 2). A total of 51 genes had between one and 43 variable amino acid sites. The 35 remaining genes had either no overlapping contiguous sequence across species, or were completely conserved at the amino acid level, i.e., no amino acid substitutions. *Vitellogenin* paralogs had high numbers of amino acid substitutions relative to other genes, perhaps suggesting relaxed selection in these egg proteins. Less unexpectedly, *MHC* genes also exhibited high numbers of amino acid substitutions among species.

Of the 342 variable amino acid positions, 105 substitutions resulted in changes in electrical charge between at least two of the North American species (Fig. 2, Appendix 1). Similarly, 63 result in changes in a change from polar to non-polar or vice versa between the

three in-group species (Appendix 1). Finally, 31 substitutions result in changes to or from proline or glycine, referred to as "special cases" due to their unique molecular structures. Thus, a total of 199 of the 342 amino acid substitutions (58.2%) result in assumed functional-property differences among species.

Simple Sequence Repeats

Sequence analysis of the candidate genes also revealed 15 simple sequence repeats in the 86 candidate genes (Table 4). Many of these repeats are located in core circadian rhythm genes for example, in *clock* and *period* paralogs. Additionally, *vitellogenin* paralogs (except *vitellogenin 3 phosvitinless*) all have a poly-serine domain. Sequence coverage in many of these repeat regions is lower than surrounding areas, perhaps indicating the difficulties of mapping reads to repeat regions due to k-mers aligning to many different locations, and when there is length variation between target and reference reads and rapid evolution in these regions may lead to insufficient conservation for mapping. However, despite this difficulty, there is evidence of among-species polymorphism in some of these regions. For example, *estrogen-related receptor alpha* has a poly-glycine repeat with the sequence "GGGGGGGG" in zebrafish, but "GGGGGGVGGGGGG" in red shiner. Additionally, there is a poly-glutamine repeat region in *cryptochrome 2a* that appears to differ between zebrafish and fathead minnow, similar to known variation in a poly-glutamine domain in *clock*, a positive transactivator of *cryptochrome*.

Red shiner gonadosomatic index and gene expression

GSI ranged from 4.6 to 14.1 in the ten red shiner females and 1.0 to 1.6 in the two males (Table 5). For perspective, Brewer et al (2008) found mid-vitellogenic (stage 3), late-vitellogenic (stage 4), and mature (stage 5) female red shiners had a mean GSI of approximately 5, 7, and

10, respectively, while post-reproductive individuals (stage 6) had GSI < 2. A total of 9,535 genes had more than ten total sequence reads for the ten females. Spearman rank correlations were calculated for GSI versus proportion of total reads represented by each gene (across the ten females). The distribution of Spearman rank correlations for these genes is presented in Fig. 3. Based on this empirical distribution of Spearman's rank correlations, the 95% confidence interval spanned correlations of -0.54 to 0.62 (two-tailed distribution, $\alpha = 0.05$). A total of 476 genes fell outside of this interval and were thus considered significantly positively or negatively correlated with GSI. This test is relatively liberal, however, and many of these 476 genes likely represent spurious correlations. Of the 86 candidate genes, four had correlations in the tails of this distribution or nearly so (Table 3), including *zona pellucia glycoproteins 2.3* and *2.5*, *aquaporin 3a*, and *estrogen-related receptor gamma b*.

Discussion

Next-generation DNA sequencing is paving the way for functional genomics studies in non-model species. In the present study, we employed (1) transcriptome-wide RNA-seq analysis and (2) a candidate gene approach for finding functional variation that takes advantage of detailed studies of model species on particular genes or functional pathways. For the former, we assessed the relationship between rate of transcription and reproductive status in female red shiner. In the latter case, we assessed variation candidate genes of known or presumed function across three species which fill different ecological roles and which differ in reproductive seasonality. Such variation could underpin some of the differences in reproduction and circadian rhythms among these species.

RNA-seq analyses of female red shiners of varying degrees of reproductive activity revealed which genes had expression patterns most strongly correlated with reproduction. Some genes known to be involved in reproduction had fairly low correlations with GSI. It is likely that expression of many genes is non-linear with respect to reproduction, and they are only 'turned on' transiently. Additionally, the range of GSI and reproductive status of red shiner in this study was relatively narrow (i.e., mid- to late-vitellogenic) and all females were approaching spawning readiness. The use of relatively high GSI individuals was necessitated by the need for sufficient amounts of mRNA for sequencing. Finally, RNA-seq data are based solely on pooled brain and ovary tissues, and reproductive genes expressed in other tissues (e.g., liver) would be missed by this analysis. The majority of genes listed in Figure 3 were not *a priori* predicted to correlate with reproduction, but nonetheless provide candidates for reproductive pathways. However, given the large number of genes in the analysis, some are expected to be false positives.

At the DNA sequence level, we identified more than 300 hundred single nucleotide polymorphisms (SNPs) that differ among relatively closely-related species. Importantly, the SNPs we identified are located in protein coding regions of reproductive and circadian rhythm genes and encode amino acid substitutions across species, and thus may have phenotypic consequences. Of the amino acid substitutions identified, nearly half (i.e., 168 of 342) resulted in either change in polarity or electrical charge and thus may have functional implications. While sequencing errors are of concern in resulting in false positives, coverage depths > 2 (as is generally the case in this study) are unlikely to repeatedly result in replicated sequencing errors

across multiple reads. Additionally, homopolymer runs were excluded from the list of variants due to high error rates of such sequences with 454 sequencing.

We also detected amino acid repeat regions that are candidates for length polymorphism and thus may be targets of divergent natural selection in shaping phenotypic traits. For example, previous studies have demonstrated functional variation in allele length of a poly-glutamine domain in *clock* (Johnsen et al. 2007; O'Malley and Banks 2008; Caprioli et al. 2012; Krabbenhoft and Turner 2012). In this study, in addition to observing the poly-glutamine domain in *clock* paralogs, we also identified a poly-glutamine domain in *cryptochrome 2a*. While only fathead minnow reads mapped to the zebrafish sequence for this region of this gene, the two species do differ in amino acid residues in this region. While there has been much recent effort in documenting and understanding drivers of variation in *clock*, we know of no studies that have examined poly-glutamine allele length variation in *cryptochrome 2a*. Because *cryptochrome* functions as a core element of the circadian oscillator by negatively regulating *clock* expression, this gene could be an additional target of natural selection in shaping circadian phenotypes. The observation of a poly-glutamine domain in *cryptochrome 2a* begs several questions. To what extent (if any) is this domain polymorphic within and among species? What are the phenotypic (circadian) implications of length polymorphism in this gene? Is allele length polymorphism in *cryptochrome 2a* the product of concerted evolution with *clock*? Future research should assess the extent and significance of variation in *cryptochrome* paralogs within and among species or populations, particularly in light of understanding variation in circadian rhythms.

With respect to functional variation, an important caveat is that many of the candidate genes have functions beyond reproduction or circadian rhythms (in some cases primary functions), and thus patterns may not be related to among-species differences in these broad physiological pathways. Regardless, along with evolution of gene regulatory regions, such changes ultimately generate the phenotypic variation that makes these species unique biological entities. Despite being confamilial, these three species have been on independent evolutionary trajectories since diverging between 43.2 and 66.4 million years ago, and diverged from the zebrafish lineage approximately 117.3 million years ago (Saito et al. 2011). Such long times to most recent common ancestor has facilitated evolutionary radiation and ecological diversification. Data presented in this paper reflect a necessary initial exploration into the amount and distribution of genetic variation among the three study species. Further study will be necessary to assess what variation is directly responsible for among-species differences in reproductive biology, and how natural selection shapes that variation over evolutionary time.

While sequence level variation likely has functional implications (evidenced by the high levels of amino acid conservation across these divergent species), it is also likely that evolution of reproductive biology of these species is also shaped by proximal control elements and small, non-coding regulatory RNAs that affect gene regulation. At this point, regulatory elements of many of these genes remain uncharacterized, and sequencing the non-coding portion of these genomes is beyond the scope of this study. In a genome-wide study in threespine sticklebacks, a large portion of the SNPs that differed between lake and stream ecomorphs were located in non-coding regions (presumed gene regulatory regions), but 17% were also in coding regions, suggesting changes in both may be important for local adaptation (Jones et al. 2012). If this

pattern represents a generality in vertebrates, as we suspect, then future sequencing of non-coding regions will likely yield interesting results. In the present study, there are single nucleotide polymorphisms in the 5' and 3' untranscribed regions (UTRs) of many of the genes examined; however, the functional implications of these changes is not known. We look forward to improved annotation of non-coding regions of the zebrafish genome to guide future investigation of gene regulation in the study species and the continued development of techniques to specifically target regulatory regions (e.g., CAGE-seq; reviewed in Lenhard et al. 2012).

Functional genomic investigations using next-generation DNA sequencing are facilitating comparative genomic studies of non-model species and are greatly advancing our understanding of genome evolution. Research in this vein can help elucidate the patterns and processes that generate functional diversity and ecological variation via divergent selection. Substantial future effort will be required to determine how genetic variation is associated with particular phenotypic traits.

Acknowledgements

We sincerely thank George Rosenberg for assistance with bioinformatics analysis and study design and Michelle Forys for laboratory assistance with the red shiner sequencing at all steps, from RNA extraction through sequencing. Thanks also to Rob Miller for his support of this research and helpful insight. The UNM Molecular Biology Facility and Center for Evolutionary and Theoretical Immunology were responsible for red shiner library construction and 454 sequencing. We thank Barbara Carter and Ecoarray, Inc. for producing the normalized cDNA

library for Rio Grande silvery minnow. Lisa Bukovnik at Duke conducted the 454 sequencing on Rio Grande silvery minnow. We appreciate insight provided by Megan Osborne, Charlie Cunningham, Scott Collins, and Kathleen O'Malley. Jon Kawatachi assisted with RNA isolation. Corey Krabbenhoft assisted with field collections and reviewed this manuscript. Thanks to Terina Perez and Kim Ward for assistance with Rio Grande silvery minnow tissue sampling at Albuquerque BioPark Aquatic Research Facility. Funding was provided by University of New Mexico Research Allocations Committee Funds to Thomas F. Turner, and by the Center for Evolutionary and Theoretical Immunology at University of New Mexico. Samples were collected under New Mexico Department of Game and Fish permit #3015 and US Fish and Wildlife permit #TE038055-0. This research was approved by Institutional Animal Care and Use Committee Protocol #10-100468-MCC and #10-100492-MCC.

Literature Cited

Abril JR, Cebria F, Rodriguez-Esteban G, Horn T, Fraguas S, Calvo B, Bartscherer K, Salo E (2010) Smed454 dataset: unravelling [*sic*] the transcriptome of *Schmidtea mediterranea*. BMC Genomics 11:731.

Boguski MS, Lowe TM, Tolstoshev CM (1993) dbEST—database for "expressed sequence tags". Nature Genetics 4(4):332-333.

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijó JA, Becker JD (2008) Comparative transcriptomics of Arabidopsis sperm cells. Plant Physiology 148:1168-1181.

Brewer SK, Rabeni CF, Papoulias DM (2008) Comparing histology and gonadosomatic index for determining spawning condition of small-bodied riverine fishes. Ecology of Freshwater Fish 17(1):54-58.

Caprioli M, Ambrosini R, Boncoraglio G, Gatti E, Romano A, Romano M, Rubolini D, Gianfranceschi L, Saino N (2012) *Clock* gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. PLoS ONE 7(4):e35140.

Crawford JE, Guelbeogo WM, Sanou A, Traoré A, Vernick KD, Sagnon N'F, Lazzaro BP (2010) *De novo* transcriptome sequencing in *Anopheles funestus* using Illumina RNA-seq technology. PLoS One 5(12):e14202.

Dunlap JC (1999) Molecular bases for circadian clocks. Cell 96:271-290.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1-15.

Gale WF (1986) Indeterminate fecundity and spawning behavior of captive red shiners-fractional, crevice spawners. Transactions of the American Fisheries Society 115(3):429-437.

Gohin M, Bobe J, Chesnel F (2010) Comparative transcriptomic analysis of follicle-enclosed oocyte maturational and developmental competence acquisition in two non-mammalian vertebrates. BMC Genomics 11:18.

Hale MC, McCormick CR, Jackson JR, DeWoody JA (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploidy lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction of gene discovery. BMC Genomics 10:203.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Hemmer-Hansen J, Nielsen EE, Meldrup D, Mittelholzer C (2011) Identification of single nucleotide polymorphisms in candidate genes for growth and reproduction in a nonmodel organism; the Atlantic cod, *Gadus morhua*. Molecular Ecology Resources 11(Suppl. 1):71-80.

Johnsen A, Fidler AE, Kuhn S, Carter KL, Hoffmann A, Barr IR, Biard C, Charmantier A, Eens M, Korsten P, Siitari H, Tomiuk J, Kempenaers B (2007) Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies. Molecular Ecology 16(22):4867-4880.

Johnston CE, Page LM (1992) The evolution of complex reproductive strategies of North American minnows (Cyprinidae). Pp. 600-621 *in:* RL Mayden (Ed.) *Systematics, historical*

ecology, and North American freshwater fishes. Stanford University Press, Stanford, CA, USA.

Johnston CE (1999) The relationship of spawning mode to conservation of North American minnows (Cyprinidae). Evolutionary Biology of Fishes 55:21-30.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55-61.

King DP, Takahashi JS (2000) Molecular genetics of circadian rhythms in mammals. Annual Reviews in Neuroscience 23:713-742.

King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, Steeves TDL, Hotz Vitaterna M, Kornhauser JM, Lowrey PL, Turek FW, Takahashi JS (1997) Positional cloning of the mouse circadian *Clock* gene. Cell 89:641-653.

Krabbenhoft TJ, Platania SP, Turner TF (2012) Reduced flows in an arid-land river decrease temporal niche partitioning in the larval fish community. *In prep.*

Krabbenhoft TJ, Turner TF (2012) *Clock* gene variation: seasonal timing, phylogenetic signal, or functional constraint? *In prep.*

Lenhard B, Sandelin A, Carnici P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Reviews Genetics 13:233-245.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078-2079.

Markus (1934) Life history of the blackhead minnow (*Pimephales promelas*) Copeia 1934(3):116-122.

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D (2010) Tablet – next generation sequence assembly visualization. Bioinformatics 26(3):401-402.

Neill JD (Ed.) (2006) *Knobil and Neill's Physiology of Reproduction. 3rd Ed.* Elsevier Press.

Nelson JS (2006) *Fishes of the World*. 4th Ed. John Wiley and Sons, Inc., Hoboken, New Jersey, USA.

Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics 11:180.

Platania SP, Altenbach CS (1998) Reproductive strategies and egg types of seven Rio Grande basin cyprinids. Copeia 1998:559-569.

Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Molecular Ecology 19(Suppl. 1):115-131.

Reppert SM, Weaver DR (2002) Coordination of circadian timing in mammals. Nature 418:935-941.

Rosenblum EB, Poorten TJ, Settles M, Murdoch GK (2012) Only skin deep: shared genetic response to the deadly chytrid fungus in susceptible frog species. Molecular Ecology DOI: 10.1111/j.1365-294X.2012.05481.x

Saito K, Sado T, Doosey MH, Bart Jr. HL, Inoue JG, Nishida M, Mayden RL, Miya M (2011) Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Osteriophysi). Zoological Journal of the Linnean Society 161:633-662.

Sarropoulou E, Fernandes JMO (2011) Comparative genomics in teleost species: knowledge transfer by linking the genomes of model and non-model fish species. Comparative Biochemistry and Physiology, Part D 6:92-102.

Seeb JE, Pascal CE, Grau ED, Seeb LW, Templin WD, Harkins T, Roberts SB (2011) Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. Molecular Ecology Resources 11:335-348.

Sprague J, Bayraktaroglu L, Clements D, Conlin T, Fashena D, Frazer K, Haendel M, Howe D, Mani P, Ramachandran S, Schaper K, Segerdell E, Song P, Sprunger B, Taylor S, Van Slyke C, Westerfield M (2006) The Zebrafish Information Network: the zebrafish model organism database. Nucleic Acids Research 34:D581-D585.

St-Cyr J, Derome N, Bernatchez L (2008) The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). Molecular Ecology 17:1850-1870.

Steinmeyer C, Mueller JC, Kempenaers B (2009) Search for informative polymorphisms in candidate genes: clock genes and circadian behaviour in blue tits. Genetica 136:109-117.

Sublette JE, Hatch MD, Sublette M (1990) *The fishes of New Mexico*. University of New Mexico Press, Albuquerque.

Turner TF, Krabbenhoft TJ, Burdett AS (2010) Reproductive phenology and fish community structure in an arid-land river system. *In: Community Ecology of Stream Fishes* (Gido K, & Jackson D, Eds.). *American Fisheries Society Symposium* 73:427-446.

Vanin S, Bhutani S, Montelli S, Menegazzi P, Green EW, Pegoraro M, Sandrelli F, Costa R, Kyriacou CP (2012) Unexpected features of *Drosophila* circadian behavioural rhythms under natural conditions. Nature doi:10.1038/nature10991

Wang J-T, Li J-T, Xiao-Feng Z, Xiao-Wen S (2012) Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). BMC Genomics 13:96.

Wheat CW (2010) Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica 138:433-451.

Wren JD, Forgacs E, Fondon JW, III, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Minna JD, Garner HR (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. American Journal of Human Genetics 67(2):345-356.

Yerushalmi S, Green RM (2009) Evidence of the adaptive significance of circadian rhythms. Ecology Letters 12:970-981.

Yúfera M, Halm S, Beltran S, Fusté B, Planas JV, Martínez-Rodríguez G (2012) Transcriptome characterization of the larval stage in gilthead seabream (*Sparus aurata*) by 454 pyrosequencing. Marine Biotechnology. DOI 10.1007/s10126-011-9422-3

Table 1. Species in this study and reproductive characteristics, as well as sequencing platform and source of transcriptomics data.

Common Name	Scientific Name	Eggs	Spawning substrate	Parental care	Platform	Data Source	Acc. No.
Rio Grande silvery minnow	*Hybognathus amarus*	Pelagic, semi-buoyant	Main-channel, open-water	None	454 FLX	Original	XXXX-XXXX
Red shiner	*Cyprinella lutrensis*	Demersal, adhesive	Broadcast-crevice spawner	None	454 FLX+	Original	XXXX-XXXX
Fathead minnow	*Pimephales promelas*	Demersal, adhesive	Prepared nest on hard substrate	Male nest-guarding	Sanger	NCBI dbEST	XXXX-XXXX

Table 2. Sequencing results and reference mapping summary for three species of cyprinid fishes in this study. Zebrafish canonical transcripts (N= 30,651) were used as the reference. Note the correspondence between N50 and percent reads aligned.

	Reads	N50	Reads aligned	% Reads aligned	Unique matching reads	Unique "genes"	Candidate genes	
							Mapped reads	Unique matching reads
Rio Grande silvery minnow	345,703	247	148,332	42.91	140,075	14,097	1,896	1,669
Red shiner	1,137,383	500	684,841	60.21	561,283	18,365	65,668	36,382
Fathead minnow	253,342	776	193,288	76.30	192,885	16,040	5,731	5,162

Table 3. Candidate genes (n=86) in this study with putative functions associated with reproduction or circadian rhythms, along with associated Ensembl ID for zebrafish canonical transcripts, GO Terms and Spearman's correlation coefficients between proportion of reads and gonadosomatic index.

Gene	Ensembl ID (ENSDART)	GO Term	Spearman's ρ	P
clock	00000025575	Circadian rhythm	NA	
clock3	00000133959	Circadian rhythm	0.242	0.651
cryptochrome 1a	00000130692	DNA photolyase activity	0.261	0.604
cryptochrome 1b	00000128609	DNA photolyase activity	NA	
cryptochrome 2a	00000129210	DNA photolyase activity	-0.079	0.541
cryptochrome 2b	00000125347	DNA photolyase activity	NA	
cryptochrome 3	00000123497	DNA photolyase activity	NA	
cryptochrome 4	00000105873	DNA photolyase activity	NA	
cryptochrome 5	00000023831	DNA photolyase activity	-0.036	0.633
melatonin recep. 1A a	00000054674	G-protein coupled receptor activity	0.497	0.153
melatonin recep. 1C	00000128781	G-protein coupled receptor activity	NA	
neuronal PAS domain protein 2	00000014806	photoperiodism	NA	
period 1a	00000121968	entrainment of circadian clock by photoperiod	NA	
period 1b	00000011082	entrainment of circadian clock by photoperiod	0.497	0.153
period 2	00000148788	entrainment of circadian clock by photoperiod	NA	
period 3	00000024304	entrainment of circadian clock by photoperiod	NA	
estrogen receptor 1	00000087844	sequence-specific DNA binding	NA	
estrogen receptor 2a	00000131069	sequence-specific DNA binding	NA	
estrogen receptor 2b	00000131800	sequence-specific DNA binding	NA	
follicle stimulating hormone recep.	00000105751	follicle-stimulating hormone receptor activity	NA	
gonadotropin rel. hormone recep. 1	00000061346	gonadotropin-releasing hormone receptor activity	NA	
gonadotropin rel. hormone recep. 4	00000055567	gonadotropin-releasing hormone receptor	NA	

Gene	Ensembl ID (ENSDART)	GO Term	Spearman's ρ	P
		activity		
gonadotropin rel. hormone 3	00000078672	hormone activity	NA	
luteinizing hormone / choriogonadotropin receptor	00000097459	G-protein coupled receptor activity	NA	
oogenesis-related gene	00000112722	cell cortex	-0.006	0.735
vitellogenin 1	00000050238	lipid transport	0.012	0.749
zona pellucida glycoprotein 2, like 1	00000077745	binding, cellular component	-0.164	0.396
zona pellucida glycoprotein 2, like 2	00000065475	binding, cellular component	0.109	0.994
zona pellucida glycoprotein 2.2	00000047101	binding, cellular component	-0.176	0.377
zona pellucida glycoprotein 2.3	**00000074085**	**binding, cellular component**	**-0.539**	**0.051**
zona pellucida glycoprotein 2.5	**00000024598**	**binding, cellular component**	**-0.600**	**0.030**
zona pellucida glycoprotein 2.6	00000121661	binding, cellular component	-0.200	0.344
zona pellucida glycoprotein 3.1	00000128797	binding, cellular component	0.370	0.353
zona pellucida glycoprotein 3.2	00000074173	binding, cellular component	-0.176	0.377
aquaporin 3a	**00000009182**	**transmembrane transporter activity**	**-0.545**	**0.049**
bone morphogenetic protein 15	00000149793	negative regulation of oocyte development	-0.115	0.473
anti-Müllerian hormone	00000013803	gonad development	0.055	0.859
Fanconi anemia, complementation group L	00000020946	oocyte development	0.550	0.104
follistatin a	00000146237	oocyte differentiation	NA	
HEN1 methyltransferase homolog 1	00000015000	oocyte development, oogenesis	-0.300	0.217
MAD homolog 2	00000121892	positive regulation of oocyte development	0.455	0.208
ornithine decarboxylase antizyme 1	00000105532	enzyme inhibitor activity	-0.224	0.311
protein arginine N-methyltransferase 7	00000073609	DNA methylation involved in gamete generation	-0.152	0.415
sperm adhesion molecule 1	00000130678	sequence-specific DNA binding	0.261	0.604
vasa homolog	00000128866	gonad development	-0.309	0.204
tudor domain containing 1	00000066249	DNA methylation involved in gamete generation	0.273	0.573
tudor domain containing 9	00000124958	DNA methylation involved in gamete generation	-0.067	0.568
aryl-hydrocarbon rec. nuc. translocator-	00000148660	photoperiodism	0.006	0.725

Gene	Ensembl ID (ENSDART)	GO Term	Spearman's ρ	P
like 1a				
aryl-hydrocarbon rec. nuc. translocator-like 1b	00000098259	photoperiodism	0.055	0.868
aryl-hydrocarbon rec. nuc. translocator-like 2	00000099849	photoperiodism	0.315	0.469
cryptochrome 2 (photolyase-like)	00000028606	DNA photolyase activity	0.515	0.139
cytochrome P450 fam. 19 subfam. A, polypep. 1a	00000129828	response to estradiol stimulus	NA	
cytochrome P450 fam. 19 subfam. A, polypep. 1b	00000130307	response to estradiol stimulus	-0.376	0.142
cystolic phospholipase a2	00000099913	ovarian follicle development; reproduction	NA	
daz-like gene	00000137590	oogenesis	-0.224	0.307
estrogen receptor binding site assoc. antigen 9	0000058546	receptor activity	-0.018	0.676
estrogen-related recep. Alpha	00000100658	sequence-specific DNA binding	0.267	0.587
estrogen-related recep. Beta	00000007600	sequence-specific DNA binding	0.364	0.359
estrogen-related recep. gamma a	00000126966	sequence-specific DNA binding	NA	
estrogen-related recep. gamma b	**00000122258**	**sequence-specific DNA binding**	**0.673**	**0.023**
follistatin a	00000146237	oocyte differentiation	NA	
hydroxysteroid (17-beta) dehydrogenase 1	00000020376	androgen metabolic process	NA	
hydroxysteroid (17-beta) dehydrogenase 8	00000001331	binding	-0.261	0.264
hydroxysteroid (17-beta) dehydrogenase 10	00000016891	binding	0.285	0.530
hydroxysteroid (17-beta) dehydrogenase 12a	0000005299	estradiol 17-beta-dehydrogenase activity	-0.467	0.080
hydroxysteroid (17-beta) dehydrogenase 12b	00000098842	estradiol 17-beta-dehydrogenase activity	0.430	0.247

Gene	Ensembl ID (ENSDART)	GO Term	Spearman's ρ	P
invariant chain-like protein 1	00000026021	MHC class II protein binding	-0.224	0.312
MHC class I UBA gene	00000009689	MHC class I protein complex	0.436	0.229
MHC class I ZE	00000039465	MHC class I protein complex	-0.200	0.342
MHC class I ZE gene	00000105659	MHC class I protein complex	0.152	0.880
MHC class II DAB gene	00000111240	MHC class II protein complex	0.345	0.398
MHC class I antigen	00000114093	MHC class I protein complex	-0.006	0.703
Nuclear factor, interleukin 3 reg.	00000138821	circadian rhythm	0.048	0.843
ornithine decarboxylase 1	00000105532	carboxy-lyase activity	0.012	0.754
ornithine decarbox. antizyme 2a	00000007748	enzyme inhibitor activity	0.018	0.759
*piwi-like 2 (*Drosophila*)*	00000134274	oogenesis	-0.158	0.403
progestin & apipoQ rec. fam. member Va	00000060311	oogenesis	0.224	0.698
progestin & apipoQ rec. fam. member VII b	00000144069	oogenesis	NA	
retinol dehydrogenase 8	00000032788	estrogen biosynthetic process	NA	
vitellogenin 2	00000061165	lipid transport	0.100	0.964
vitellogenin 3 (phosvitinless)	00000014979	lipid transport	0.133	0.928
vitellogenin 4	00000136837	lipid transport	0.109	0.998
vitellogenin 5	00000078225	lipid transport	-0.055	0.592
vitellogenin 7	00000078216	lipid transport	0.079	0.920
zona pellucida protein C	00000082324	cellular component	0.406	0.292
zygote arrest 1-like	00000046599	sodium ion transport	0.091	0.941

Table 4. Amino acid repeats in candidate reproductive and circadian rhythm genes. Sequences are from zebrafish. Only repeat regions longer than five amino acid residues are reported.

Gene	Amino acid sequence
clock	QQQQQLHQQQQQLQQQQQLQQQQQQQQQLQQQHQQQQQQLQQQHQQQQQQ
*clock homolog 3 (*mouse*)*	QQQQQQQQ
neuronal PAS domain protein 2	QQQQQQQQ
cryptochrome 2a	QQQHQQQQQQQQQQ
*period homolog 1a (*Drosophila*)*	SGSGSGSGS
*period homolog 1b (*Drosophila*)*	SGSGSSGTGSSGSGSGSSGSGSNGCSSSGSG
*period homolog 2 (*Drosophila*)*	$(ESHGN)_4$ [motif]
period homolog 3	HHRHHHHHRHHHHHHHH
estrogen-related receptor alpha	GGGGGGGG
HEN1 methyltransferase homolog 1	EEEEEEEEEE
vitellogenin 1	SSESSSSRNSRSSSSRSTSTSTSSSSSSSSSSSSSSSSSMSSS
vitellogenin 2	SSGSSSSRSSSSSRSSSSSSSTSNSTSSSSSSSSSSSSMSSS
vitellogenin 4	SSESSSSSSRNSRSTSSSSTSTSSSGSSSSSSSSSSSMSSS
vitellogenin 5	SSESSSSSRNSRSSSSSSSSTSTSSSSSSSSSSSSSSLSSS
vitellogenin 7	SSESSSSRNSRSSSSRSHSTSSSSSSRSSSSSSSSMSSS

Table 5. Gonadosomatic index (GSI), standard length (SL), sex, and number of sequences mapped to zebrafish (*Danio rerio*) genes for twelve red shiner (*Cyprinella lutrensis*) individuals sequenced.

Sample	GSI	SL (mm)	Sex	Total Reads	Mapped Reads
CL01	6.2	48	Female	98,930	55,050
CL02	10.8	32	Female	143,814	68,836
CL03	11.3	45	Female	87,718	41,171
CL04	6.9	46	Female	74,153	53,781
CL05	8.0	46	Female	106,230	44,572
CL06	4.6	57	Female	98,436	58,886
CL07	11.0	61	Female	118,097	33,942
CL08	13.6	34	Female	73,733	44,328
CL09	14.1	48	Female	91,194	49,630
CL10	9.2	53	Female	75,298	35,182
CL11	1.0	56	Male	106,345	33,757
CL12	1.6	51	Male	63,790	42,148

Figure 1. Sequence length histograms for three cyprinid fishes in this study.

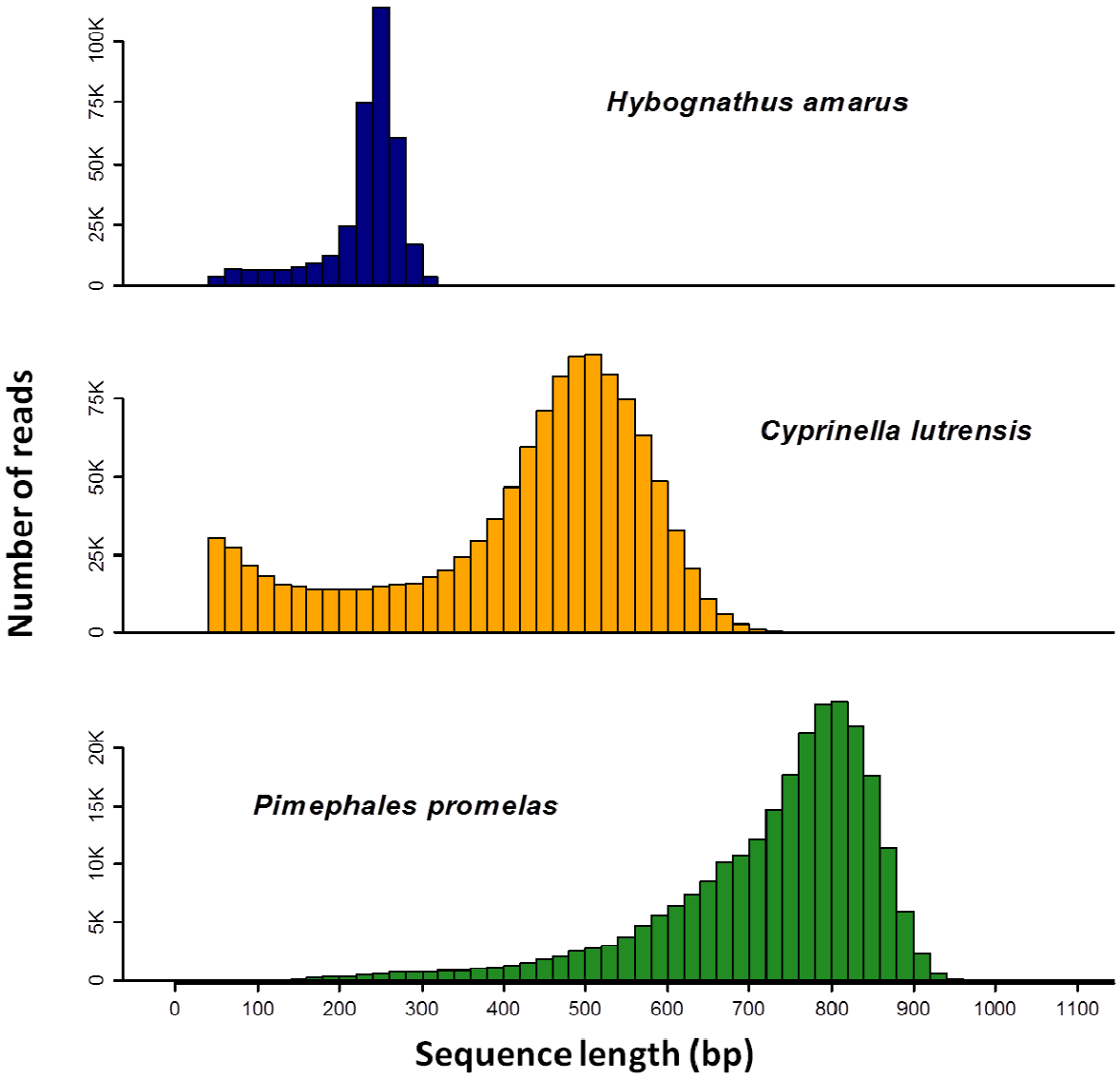


Figure 2. Number of amino acid substitutions per gene among three species of North American cyprinid fishes. Amino acid substitutions resulting in changes in electrical charge and polarity are also shown. "Property changes" refers to the number of substitutions resulting in changes in electrical charge, polarity, and special case changes. "Special case" amino acid changes are those to or from proline or glycine, two structurally unique amino acids. Note the large number of substitutions found in vitellogenin paralogs. The candidate genes with no substitutions (n = 35) are not shown.

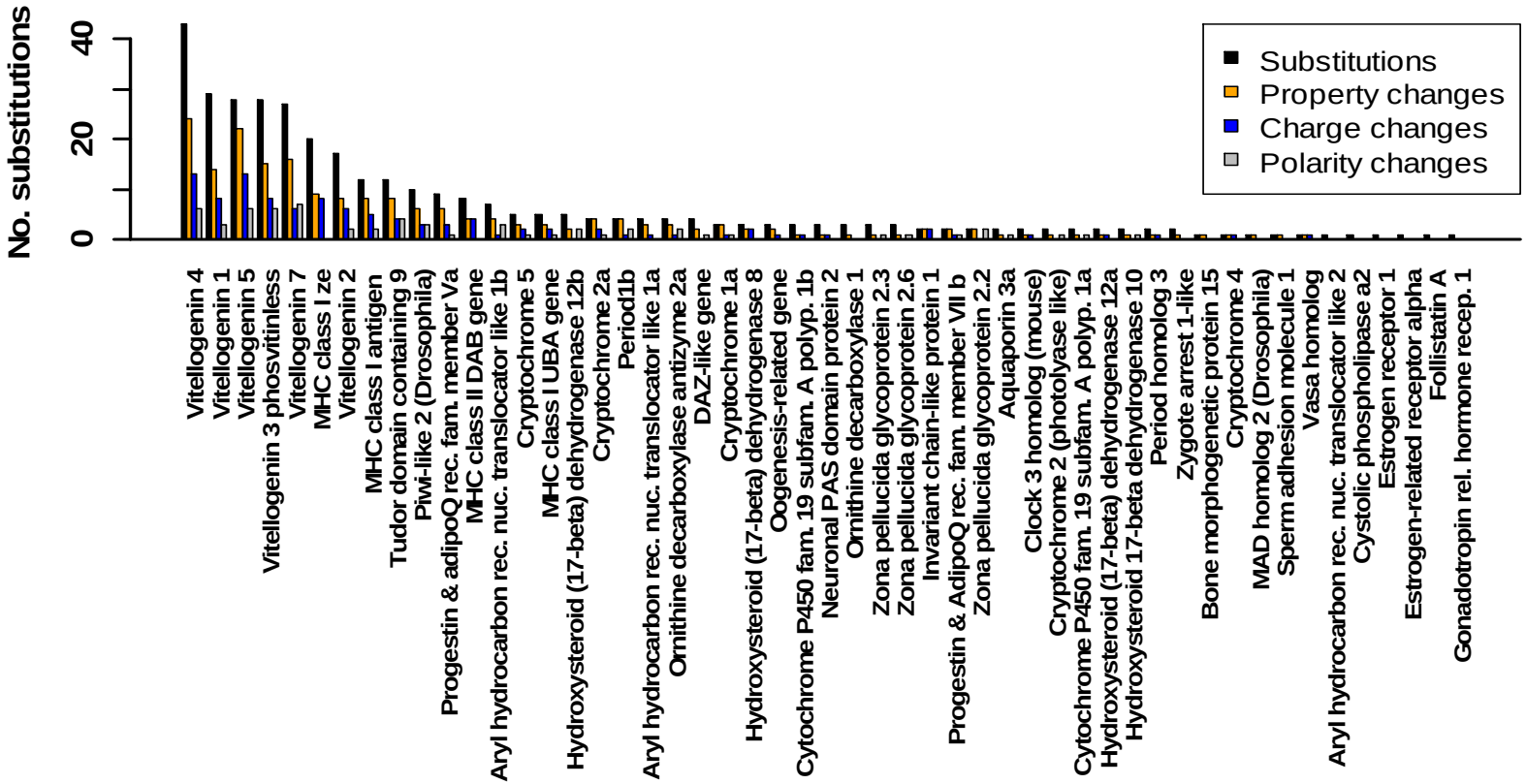
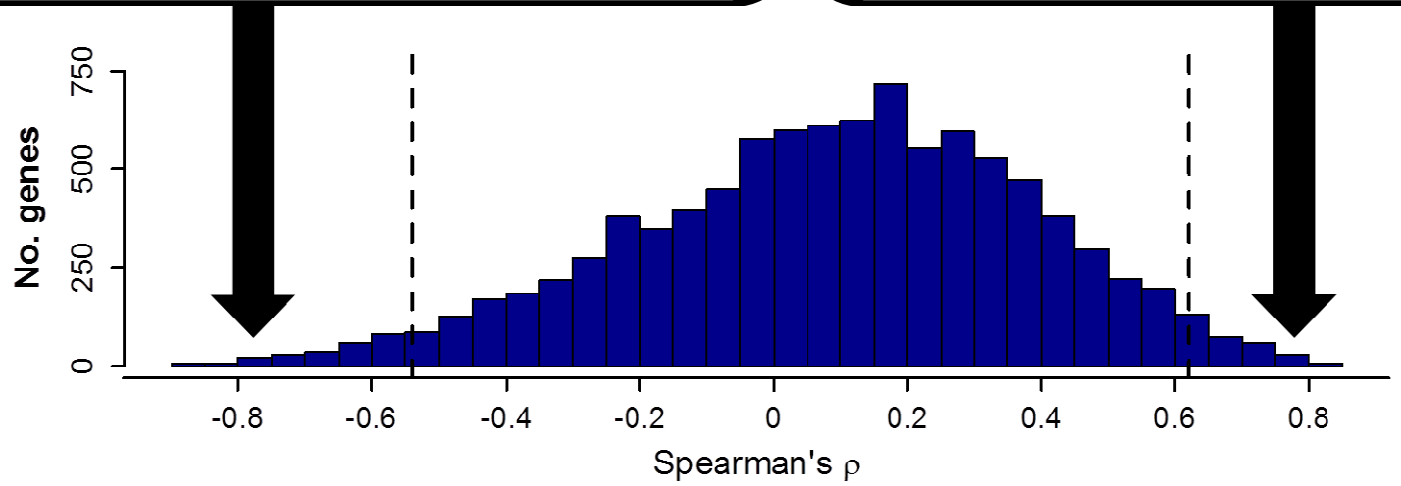


Figure 3. Histogram depicting the distribution of Spearman rank correlation coefficients between transcript abundance and gonadosomatic index in ten adult female red shiner (*Cyprinella lutrensis*). Only genes with 10 or more sequence reads total were analyzed (n = 9,535). Based on this empirical distribution, the 95% confidence interval ranges from -0.54 to 0.62 (dotted lines). Genes with correlations falling outside of this interval are considered to have significant association with female GSI.

Gene	ρ
mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltr.	-0.879
adaptor-related protein complex 1, gamma 2 subunit	-0.867
methyl-CpG binding domain protein 1	-0.855
tyrosine 3-monooxygenase/tryptophan 5-monooxygenase act. prot.	-0.842
fascin homolog 1, actin-bundling protein	-0.842
DDB1 and CUL4 associated factor 12	-0.842
vacuolar protein sorting 35 (yeast)	-0.818
interferon-related developmental regulator 2	-0.806
ceroid-lipofuscinosis, neuronal 6, late infantile, variant	-0.806
arachidonate 5-lipoxygenase	-0.794

Gene	ρ
MYST histone acetyltransferase 2	0.848
nucleotide binding protein 2 (MinD homolog, E. coli)	0.842
ArfGAP with SH3 domain, ankyrin repeat and PH dom. 2a	0.818
family with sequence similarity 86, member A	0.806
regulator of calcineurin 1	0.806
mitochondrial ribosomal protein S18B	0.806
FERM, RhoGEF (ARHGEF) and pleckstrin dom. prot. 1	0.800
adenosine deaminase, RNA-specific, B1	0.794
mitochondrial ribosomal protein 63	0.794
chromosome 11 open reading frame 30	0.784



Appendix 1. Non-synonymous amino acid substitutions in coding regions of candidate genes. Only amino acids which vary among the North American species are shown. Position refers to the nucleotide positions in zebrafish from the cDNA start site. Gaps in sequence coverage are denoted as "No Data".

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Aquaporin 3a	400-402	Leu	Leu	NO DATA	Ile			
Aquaporin 3a	490-492	Leu	Thr	NO DATA	Val		x	
Aryl hydrocarbon rec. nuc. translocator like 1a	934-936	Asp	Glu	NO DATA	Asp			
Aryl hydrocarbon rec. nuc. translocator like 1a	946-948	Pro	Pro	NO DATA	Ser			x
Aryl hydrocarbon rec. nuc. translocator like 1a	1108-1110	Leu	Leu	NO DATA	Pro			x
Aryl hydrocarbon rec. nuc. translocator like 1a	1735-1737	Pro	Pro	NO DATA	His	x		
Aryl hydrocarbon rec. nuc. translocator like 1b	214-216	Asn	His	NO DATA	Asn	x		
Aryl hydrocarbon rec. nuc. translocator like 1b	259-261	Met	Leu	NO DATA	Met		x	
Aryl hydrocarbon rec. nuc. translocator like 1b	331-333	Ser	Leu	NO DATA	Ser		x	
Aryl hydrocarbon rec. nuc. translocator like 1b	463-465	Met	Met	NO DATA	Ile		x	
Aryl hydrocarbon rec. nuc. translocator like 1b	712-714	Lys	Arg	NO DATA	Lys			
Aryl hydrocarbon rec. nuc. translocator like 1b	967-969	Arg	Arg	Lys	Arg			
Aryl hydrocarbon rec. nuc. translocator like 1b	1039-1041	Asp	Glu	Asp	NO DATA			
Aryl hydrocarbon rec. nuc. translocator like 2	106-108	Ser	Ala	NO DATA	Val			
Bone morphogenetic protein 15	448-450	Pro	Ser	Pro	Pro			x
Clock 3 homolog (mouse	937-939	Ile	Ile	NO DATA	Leu			
Clock 3 homolog (mouse	961-963	Leu	Gln	NO DATA	His	x		
Cryptochrome 1a	889-891	Leu	Leu	Leu	Pro			x
Cryptochrome 1a	2020-2022	Tyr	Tyr	Cys	NO DATA		x	
Cryptochrome 1a	2230-2232	Asn	Asn	NO DATA	His	x		
Cryptochrome 2 (photolyase like	244-246	Asp	Glu	Asp	Asp			
Cryptochrome 2 (photolyase like	298-300	Tyr	Tyr	Asn	Tyr		x	
Cryptochrome 2a	256-258	Gln	Gln	Arg	Gln	x		
Cryptochrome 2a	613-615	Leu	Leu	Leu	Gln		x	
Cryptochrome 2a	1024-1026	Lys	Gln	NO DATA	Lys	x		

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Cryptochrome 2a	1744-1746	Pro	Pro	NO DATA	Gln			x
Cryptochrome 4	1528-1530	Lys	Lys	NO DATA	Gln	x		
Cryptochrome 5	1180-1182	Leu	Leu	NO DATA	Cys		x	
Cryptochrome 5	1183-1185	Arg	Arg	NO DATA	Val	x		
Cryptochrome 5	1186-1188	Gln	Arg	NO DATA	Gln	x		
Cryptochrome 5	1291-1293	Leu	Trp	NO DATA	Leu			
Cryptochrome 5	1387-1389	Ile	Ile	NO DATA	Val			
Cystolic phospholipase a2	322-324	Ile	Ile	NO DATA	Val			
Cytochrome P450 fam. 19 subfam. A polyp. 1a	352-354	Ser	Phe	NO DATA	Ser		x	
Cytochrome P450 fam. 19 subfam. A polyp. 1a	385-387	Ser	Asn	NO DATA	Ser			
Cytochrome P450 fam. 19 subfam. A polyp. 1b	472-474	Glu	Glu	NO DATA	Gln	x		
Cytochrome P450 fam. 19 subfam. A polyp. 1b	1342-1344	Asn	Asn	NO DATA	Ser			
Cytochrome P450 fam. 19 subfam. A polyp. 1b	1486-1488	Ile	Ile	NO DATA	Val			
DAZ-like gene	583-585	Ser	Ser	NO DATA	Asn			
DAZ-like gene	658-660	Met	Val	NO DATA	Ile			
DAZ-like gene	766-768	Met	Ile	NO DATA	Met		x	
DAZ-like gene	784-786	Ser	Pro	NO DATA	Ser			x
Estrogen receptor 1	550-552	Asn	Ser	Asn	Asn			
Estrogen-related receptor alpha	1012-1014	Ala	Ala	NO DATA	Val			
Follistatin A	418-420	Ser	Ser	NO DATA	Asn			
Gonadotropin rel. hormone recep.1	916-918	Leu	Leu	NO DATA	Phe			
Hydroxysteroid (17-beta dehydrogenase 12a	121-123	Ala	Val	Ala	Val			
Hydroxysteroid (17-beta dehydrogenase 12a	418-420	Asp	Asp	Phe	Asp	x		
Hydroxysteroid (17-beta dehydrogenase 12b	181-183	Leu	Val	NO DATA	Leu			
Hydroxysteroid (17-beta dehydrogenase 12b	418-420	Ile	Ile	Ile	Asn		x	
Hydroxysteroid (17-beta dehydrogenase 12b	451-453	Pro	Ser	Thr	Thr			
Hydroxysteroid (17-beta dehydrogenase 12b	667-669	Ala	Ala	Ser	Ala		x	
Hydroxysteroid (17-beta dehydrogenase 12b	802-804	Phe	Ile	Phe	NO DATA			
Hydroxysteroid (17-beta dehydrogenase 8	448-450	Arg	Lys	Asn	NO DATA	x		
Hydroxysteroid (17-beta dehydrogenase 8	457-459	Glu	Val	Asp	NO DATA	x		
Hydroxysteroid (17-beta dehydrogenase 8	910-912	Ile	Ile	Phe	NO DATA			
Hydroxysteroid 17-beta dehydrogenase 10	619-621	Ile	Val	NO DATA	Ile			

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Hydroxysteroid 17-beta dehydrogenase 10	790-792	Ser	Ser	Ser	Ala		x	
Invariant chain-like protein 1	751-753	Ala	Gln	NO DATA	Glu	x		
Invariant chain-like protein 1	775-775	Met	Leu	Met	Lys	x		
MAD homolog 2 (Drosophila	1156-1158	Ser	Ser	Pro	Ser			x
MHC class I antigen	217-219	Lys	Lys	Arg	Lys			
MHC class I antigen	220-222	Ile	Val	Ile	Ile			
MHC class I antigen	232-234	Gln	His	Gln	Arg	x		
MHC class I antigen	253-255	Gln	Gln	His	Pro	x		
MHC class I antigen	292-294	Ser	Arg	Ser	Ser	x		
MHC class I antigen	301-303	Gln	Leu	Gln	Gln		x	
MHC class I antigen	445-447	Gly	Ser	NO DATA	Gly			x
MHC class I antigen	478-480	Glu	Asp	Asp	Glu			
MHC class I antigen	481-483	Thr	Arg	Lys	Arg			
MHC class I antigen	508-510	Lys	Asp	Asp	Phe	x		
MHC class I antigen	511-513	Glu	Glu	Ala	Ser	x		
MHC class I antigen	514-516	Ala	Ala	Ala	Ser		x	
MHC class I UBA	208-210	Ile	Ile	Val	NO DATA			
MHC class I UBA	247-250	Asn	Asn	Lys	NO DATA	x		
MHC class I UBA	295-297	Glu	Glu	Asp	NO DATA			
MHC class I UBA	307-309	Gln	Arg	Gln	NO DATA	x		
MHC class I UBA	310-312	Gln	Gln	Phe	NO DATA		x	
MHC class I ze	199-201	Asp	Asp	Glu	Glu			
MHC class I ze	202-204	Gln	Gln	Gln	Lys	x		
MHC class I ze	205-207	Lys	Arg	Lys	Lys			
MHC class I ze	211-213	Ile	Val	Ile	Ile			
MHC class I ze	247-249	Glu	Glu	Glu	Ala	x		
MHC class I ze	325-327	Asp	Glu	Asp	Glu			
MHC class I ze	334-336	Arg	Arg	Gly	Gly	x		
MHC class I ze	346-348	Lys	Glu	Asp	Asp			
MHC class I ze	355-357	Leu	Ile	Val	Val			
MHC class I ze	478-480	Asp	Asp	Glu	Glu			
MHC class I ze	502-504	Asp	Asp	Val	Val	x		

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
MHC class I ze	511-513	Leu	Leu	Val	Glu	x		
MHC class I ze	523-525	Arg	Arg	Thr	Arg	x		
MHC class I ze	535-537	Asn	Asn	Asn	Met			
MHC class I ze	598-600	Asn	Thr	Asn	Asn			
MHC class I ze	601-603	Lys	Arg	Lys	Lys			
MHC class I ze	607-609	Arg	Arg	Lys	Arg			
MHC class I ze	610-612	Glu	Glu	Gly	Glu	x		
MHC class I ze	616-618	Gly	Gly	Val	Gly			x
MHC class I ze	619-620	Asn	Asn	Asp	Asp	x		
MHC class II DAB gene	442-444	Glu	Glu	Asp	Glu			
MHC class II DAB gene	454-456	Lys	Gln	Phe	Lys	x		
MHC class II DAB gene	457-459	Lys	His	Arg	Lys			
MHC class II DAB gene	490-492	Glu	Glu	Val	Leu	x		
MHC class II DAB gene	493-495	Val	Val	Val	Ile			
MHC class II DAB gene	499-501	Ser	Ser	Thr	Ser			
MHC class II DAB gene	517-519	Met	Met	Met	Glu	x		
MHC class II DAB gene	559-561	His	His	Phe	His	x		
Neuronal PAS domain protein 2	649-651	Val	Val	Ala	NO DATA			
Neuronal PAS domain protein 2	682-684	Ile	Ile	Val	NO DATA			
Neuronal PAS domain protein 2	1297-1299	Glu	NO DATA	Phe	Asp	x		
Oogenesis-related gene	595-597	Ala	Ala	Gly	Gly			x
Oogenesis-related gene	619-621	Tyr	His	Tyr	Tyr	x		
Oogenesis-related gene	655-657	Val	Ile	Leu	Ile			
Ornithine decarboxylase 1	1228-1230	Val	Met	NO DATA	Gly			x
Ornithine decarboxylase 1	1231-1233	Ser	Thr	NO DATA	Thr			
Ornithine decarboxylase 1	1636-1638	Arg	Arg	NO DATA	His			
Ornithine decarboxylase antizyme 2a	229-231	Ala	Ala	Ala	Ser		x	
Ornithine decarboxylase antizyme 2a	454-459	Leu	Leu	Ser	NO DATA		x	
Ornithine decarboxylase antizyme 2a	457-459	Glu	Glu	Gly	NO DATA	x		
Ornithine decarboxylase antizyme 2a	460-462	Phe	Phe	Val	NO DATA			
Period homolog 3	1768-1770	Thr	Met	NO DATA	Thr			
Period homolog 3	1828-1830	His	Tyr	NO DATA	His	x		

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Period1b	1033-1035	Ala	NO DATA	Ala	Thr		x	
Period1b	1549-1551	Ser	Ser	Pro	Ser			x
Period1b	2698-2700	Ala	Ala	Thr	Ala		x	
Period1b	2704-2706	Gly	Leu	Leu	Arg	x		
Piwi-like 2 (Drosophila	358-360	Leu	Leu	NO DATA	Phe			
Piwi-like 2 (Drosophila	409-411	Thr	Ala	NO DATA	Thr		x	
Piwi-like 2 (Drosophila	1657-1659	Ser	Thr	NO DATA	Ser			
Piwi-like 2 (Drosophila	2374-2376	Leu	Met	NO DATA	Leu		x	
Piwi-like 2 (Drosophila	2452-2454	Val	Val	NO DATA	Ile			
Piwi-like 2 (Drosophila	2911-2913	Asp	Asp	NO DATA	Asn	x		
Piwi-like 2 (Drosophila	3067-3069	Asp	Asn	NO DATA	Asp	x		
Piwi-like 2 (Drosophila	3154-3156	Thr	Ser	NO DATA	Thr			
Piwi-like 2 (Drosophila	3178-3180	Cys	Arg	NO DATA	Cys	x		
Piwi-like 2 (Drosophila	3229-3231	Asn	Asn	NO DATA	Tyr		x	
Progestin & AdipoQ rec. fam. member VIIb	451-453	Met	Leu	NO DATA	Met		x	
Progestin & AdipoQ rec. fam. member VIIb	781-783	Cys	Cys	Arg	Cys	x		
Progestin & adipoQ rec. fam. member VA	481-483	Arg	Gln	NO DATA	Lys	x		
Progestin & adipoQ rec. fam. member VA	685-687	Asn	Asn	Lys	Asn	x		
Progestin & adipoQ rec. fam. member VA	688-690	Gly	Ser	Ser	Gly			x
Progestin & adipoQ rec. fam. member VA	712-714	Ser	Ser	Pro	Pro			x
Progestin & adipoQ rec. fam. member VA	772-774	Leu	Val	Leu	Leu			
Progestin & adipoQ rec. fam. member VA	919-921	Ala	Ala	Ala	Thr		x	
Progestin & adipoQ rec. fam. member VA	952-954	Thr	Thr	Thr	Met			
Progestin & adipoQ rec. fam. member VA	1021-1023	Arg	Arg	Thr	Arg	x		
Progestin & adipoQ rec. fam. member VA	1078-1080	Phe	Phe	Val	Phe			
Sperm adhesion molecule 1	1036-1038	Ser	Gly	NO DATA	Ala			x
Tudor domain containing 9	679-681	Val	Val	NO DATA	Ile			
Tudor domain containing 9	1324-1326	Arg	Arg	NO DATA	Cys	x		
Tudor domain containing 9	1327-1329	Ile	Ile	NO DATA	Met		x	
Tudor domain containing 9	1507-1509	Asp	Asp	NO DATA	Glu			
Tudor domain containing 9	2884-2886	Met	Ile	NO DATA	Met		x	
Tudor domain containing 9	2887-2889	Thr	Thr	NO DATA	Ser			

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Tudor domain containing 9	3220-3222	Ser	Asn	NO DATA	Ser			
Tudor domain containing 9	3238-3240	Leu	Met	NO DATA	Val		x	
Tudor domain containing 9	3253-3255	Glu	Glu	NO DATA	Val	x		
Tudor domain containing 9	3277-3279	Gln	Glu	NO DATA	Lys	x		
Tudor domain containing 9	3286-3288	Leu	Arg	NO DATA	Gln	x		
Tudor domain containing 9	3313-3315	Ala	Ala	NO DATA	Thr		x	
Vasa homolog	1003-1005	His	Tyr	NO DATA	His	x		
Vitellogenin 1	283-285	Val	NO DATA	Ala	Val			
Vitellogenin 1	349-351	Ala	NO DATA	Val	Ala			
Vitellogenin 1	424-426	Leu	NO DATA	Pro	Leu			x
Vitellogenin 1	514-516	Val	NO DATA	Ile	Ala			
Vitellogenin 1	520-522	Asn	NO DATA	Ser	Asn			
Vitellogenin 1	529-531	Pro	NO DATA	Pro	Thr			x
Vitellogenin 1	790-792	Lys	Arg	Lys	Lys			
Vitellogenin 1	829-831	Val	Ile	Val	Val			
Vitellogenin 1	880-882	Phe	Leu	Leu	Phe			
Vitellogenin 1	916-918	Met	Met	Ile	Met		x	
Vitellogenin 1	934-936	Pro	Pro	Pro	Arg	x		
Vitellogenin 1	946-948	Val	Ile	Ile	Thr		x	
Vitellogenin 1	1276-1278	Thr	Ile	NO DATA	Met		x	
Vitellogenin 1	1357-1359	Val	Ile	NO DATA	Val			
Vitellogenin 1	1363-1365	Val	Asp	NO DATA	Val	x		
Vitellogenin 1	1414-1416	Glu	Glu	NO DATA	Asp			
Vitellogenin 1	1867-1869	Arg	His	NO DATA	Arg			
Vitellogenin 1	2215-2217	Phe	Leu	NO DATA	Phe			
Vitellogenin 1	2233-2235	Tyr	Tyr	Phe	Tyr			
Vitellogenin 1	2311-2313	Glu	Arg	Leu	Arg	x		
Vitellogenin 1	2437-2439	Ala	Gly	NO DATA	Ala			x
Vitellogenin 1	2977-2979	His	Arg	NO DATA	Gln	x		
Vitellogenin 1	2995-2997	Arg	Ser	NO DATA	Arg	x		
Vitellogenin 1	3010-3012	Phe	Tyr	NO DATA	Phe			
Vitellogenin 1	3091-3093	Val	Ile	NO DATA	Val			

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Vitellogenin 1	3703-3705	Phe	Val	NO DATA	Leu			
Vitellogenin 1	3892-3894	Arg	Ser	NO DATA	Arg	x		
Vitellogenin 1	3946-3948	Arg	Thr	NO DATA	Arg	x		
Vitellogenin 1	3976-3978	Glu	Glu	NO DATA	Gln	x		
Vitellogenin 2	1288-1290	Leu	Ile	Leu	Met		x	
Vitellogenin 2	1315-1317	Ala	Ala	Val	Ala			
Vitellogenin 2	1369-1371	Val	Ile	NO DATA	Val			
Vitellogenin 2	1375-1377	Val	Asp	NO DATA	Val	x		
Vitellogenin 2	2095-2097	Ala	Ala	Val	Ala			
Vitellogenin 2	2104-2106	Lys	Gln	Gln	Lys	x		
Vitellogenin 2	2119-2121	Asp	Glu	Asp	Asp			
Vitellogenin 2	2125-2127	Ser	Asn	Ser	Ser			
Vitellogenin 2	2182-2184	Ala	NO DATA	Ala	Asp	x		
Vitellogenin 2	2923-2925	Ala	Thr	Ala	Ala		x	
Vitellogenin 2	2980-2982	Val	Val	Ile	Val			
Vitellogenin 2	2995-2997	His	His	His	Gln	x		
Vitellogenin 2	3007-3009	Phe	Phe	Leu	Phe			
Vitellogenin 2	3013-3015	Arg	Ser	Arg	Arg	x		
Vitellogenin 2	3028-3030	Phe	Tyr	Phe	Phe			
Vitellogenin 2	3109-3111	Val	Ile	Val	Val			
Vitellogenin 2	3154-3156	Asn	Arg	Ser	Ser	x		
Vitellogenin 3 phosvitinless	178-180	Ile	Ile	Val	Ile			
Vitellogenin 3 phosvitinless	304-306	Ile	Leu	Phe	Phe			
Vitellogenin 3 phosvitinless	355-357	Arg	Trp	Arg	Arg	x		
Vitellogenin 3 phosvitinless	379-381	Thr	Ala	Thr	Ala		x	
Vitellogenin 3 phosvitinless	394-396	Val	Val	Ala	Val			
Vitellogenin 3 phosvitinless	619-621	Lys	Arg	Lys	NO DATA			
Vitellogenin 3 phosvitinless	703-705	Ala	Ala	Val	Ala			
Vitellogenin 3 phosvitinless	730-732	Ser	Ser	Ser	Thr			
Vitellogenin 3 phosvitinless	760-762	Arg	Arg	Gln	Arg	x		
Vitellogenin 3 phosvitinless	811-813	Asp	Asp	Gly	Asp	x		
Vitellogenin 3 phosvitinless	817-819	Val	Val	Ile	Val			

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Vitellogenin 3 phosvitinless	820-822	Val	Val	Val	Met		x	
Vitellogenin 3 phosvitinless	835-837	Gln	Glu	Gln	Gln	x		
Vitellogenin 3 phosvitinless	862-864	Thr	Thr	Ile	NO DATA		x	
Vitellogenin 3 phosvitinless	1222-1224	Glu	Asp	Phe	NO DATA	x		
Vitellogenin 3 phosvitinless	1345-1347	Thr	Ser	Thr	NO DATA			
Vitellogenin 3 phosvitinless	1420-1422	Asp	Glu	Lys	NO DATA	x		
Vitellogenin 3 phosvitinless	1768-1770	Leu	NO DATA	Ile	Leu			
Vitellogenin 3 phosvitinless	1783-1785	Lys	Lys	Lys	Met	x		
Vitellogenin 3 phosvitinless	1822-1824	Cys	Tyr	Cys	Cys		x	
Vitellogenin 3 phosvitinless	1852-1854	Lys	Arg	Lys	Arg			
Vitellogenin 3 phosvitinless	1921-1923	Phe	Phe	Ile	Phe			
Vitellogenin 3 phosvitinless	2419-2421	Val	Met	Ile	Met		x	
Vitellogenin 3 phosvitinless	2422-2424	Val	Ile	Leu	Val			
Vitellogenin 3 phosvitinless	2428-2430	Ala	Gly	Ala	Gly			x
Vitellogenin 3 phosvitinless	2512-2514	Gln	Asp	Arg	His	x		
Vitellogenin 3 phosvitinless	2695-2697	Val	NO DATA	Val	Ala			
Vitellogenin 3 phosvitinless	2719-2721	Ser	NO DATA	Phe	Ser		x	
Vitellogenin 4	190-192	Leu	Gln	NO DATA	His	x		
Vitellogenin 4	217-219	Tyr	Phe	NO DATA	Phe			
Vitellogenin 4	232-234	Met	Val	NO DATA	Met		x	
Vitellogenin 4	250-252	Glu	Asp	NO DATA	Glu			
Vitellogenin 4	277-279	Ser	Gln	NO DATA	Pro			x
Vitellogenin 4	370-372	Val	Leu	Val	Val			
Vitellogenin 4	385-387	Gly	Glu	Gly	Gly	x		
Vitellogenin 4	451-453	Lys	Asn	Lys	Lys	x		
Vitellogenin 4	514-516	Val	Ala	Ile	Ala			
Vitellogenin 4	520-522	Asn	Ser	Ser	Asn			
Vitellogenin 4	529-531	Pro	Thr	Pro	Thr			x
Vitellogenin 4	541-543	His	Gln	His	His	x		
Vitellogenin 4	601-603	Val	Ile	Val	Val			
Vitellogenin 4	676-678	Asn	Ser	NO DATA	Asn			
Vitellogenin 4	730-732	Glu	NO DATA	Asp	Glu			

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Vitellogenin 4	790-792	Lys	Arg	Phe	Arg	x		
Vitellogenin 4	826-828	Pro	Pro	Ser	Pro			x
Vitellogenin 4	880-882	Phe	Leu	NO DATA	Phe			
Vitellogenin 4	934-936	Pro	Pro	NO DATA	Arg	x		
Vitellogenin 4	958-960	Lys	Lys	NO DATA	Arg			
Vitellogenin 4	1006-1008	Leu	Gln	Leu	Gln		x	
Vitellogenin 4	1015-1017	Val	Ile	Val	Ile			
Vitellogenin 4	1057-1059	Ile	Ile	Ile	Thr		x	
Vitellogenin 4	1531-1533	Ala	Ala	NO DATA	Gly			x
Vitellogenin 4	1534-1536	Ala	Asp	NO DATA	Ala	x		
Vitellogenin 4	1636-1638	Val	Phe	NO DATA	Val			
Vitellogenin 4	1696-1698	Ala	Ala	NO DATA	Ser		x	
Vitellogenin 4	2092-2094	Lys	Lys	Gln	Lys	x		
Vitellogenin 4	2171-2173	Ala	Asp	Ala	Asp	x		
Vitellogenin 4	2174-2176	Leu	Leu	Phe	Phe			
Vitellogenin 4	2206-2208	Val	Val	NO DATA	Ile			
Vitellogenin 4	2215-2217	Phe	Leu	NO DATA	Phe			
Vitellogenin 4	2275-2277	Met	Ile	NO DATA	Met		x	
Vitellogenin 4	2356-2358	Lys	Asn	NO DATA	Lys	x		
Vitellogenin 4	2437-2439	Ala	Gly	NO DATA	Ala			x
Vitellogenin 4	2665-2667	Val	Val	NO DATA	Met		x	
Vitellogenin 4	2668-2670	Ala	Val	NO DATA	Phe			
Vitellogenin 4	2701-2703	Lys	Arg	NO DATA	Lys			
Vitellogenin 4	2725-2727	Leu	Leu	NO DATA	Val			
Vitellogenin 4	2782-2784	Glu	Glu	NO DATA	Ala	x		
Vitellogenin 4	3940-3942	Lys	Thr	Arg	Arg	x		
Vitellogenin 4	3952-3954	Val	Ile	Val	Val			
Vitellogenin 4	3970-3972	Glu	Glu	Glu	Gln	x		
Vitellogenin 5	208-210	Glu	Gly	Phe	Glu	x		
Vitellogenin 5	229-231	Leu	Ala	Leu	Leu			
Vitellogenin 5	232-234	Met	Leu	Met	Ile		x	
Vitellogenin 5	250-252	Glu	Asp	Phe	Glu	x		

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Vitellogenin 5	271-273	Arg	Lys	Phe	Lys	x		
Vitellogenin 5	277-279	Pro	Gln	Pro	Pro			x
Vitellogenin 5	370-372	Val	Leu	Val	Val			
Vitellogenin 5	451-453	Lys	Asn	NO DATA	Lys	x		
Vitellogenin 5	529-531	Pro	Thr	NO DATA	Pro			x
Vitellogenin 5	541-543	His	Gln	NO DATA	His	x		
Vitellogenin 5	700-702	Asn	Ser	NO DATA	Ala		x	
Vitellogenin 5	724-726	Thr	Thr	Lys	Thr	x		
Vitellogenin 5	796-798	Thr	Thr	Ile	Thr		x	
Vitellogenin 5	829-831	Val	NO DATA	Val	Ile			
Vitellogenin 5	1276-1278	Thr	Ile	NO DATA	Met		x	
Vitellogenin 5	1357-1359	Thr	Ile	NO DATA	Val			
Vitellogenin 5	1363-1365	Val	Asp	NO DATA	Val	x		
Vitellogenin 5	1534-1536	Ala	Asp	NO DATA	Ala	x		
Vitellogenin 5	1648-1650	Ala	Thr	NO DATA	Ala		x	
Vitellogenin 5	1696-1698	Ala	Ala	NO DATA	Ser		x	
Vitellogenin 5	2437-2439	Ala	Gly	NO DATA	Ala			x
Vitellogenin 5	2551-2553	Glu	Glu	Ala	Glu	x		
Vitellogenin 5	2629-2631	Ala	Ala	Asp	Ala	x		
Vitellogenin 5	2659-2661	Gly	Gly	Arg	Gly	x		
Vitellogenin 5	2668-2670	Ala	Val	NO DATA	Phe			
Vitellogenin 5	2782-2784	Glu	Glu	NO DATA	Ala	x		
Vitellogenin 5	3550-3552	Gln	Gln	His	His	x		
Vitellogenin 5	3697-3699	Phe	Val	Phe	Val			
Vitellogenin 7	57-60	Ser	Cys	Ala	Cys		x	
Vitellogenin 7	208-210	Glu	Gly	Glu	Glu	x		
Vitellogenin 7	250-252	Glu	Asp	Glu	Glu			
Vitellogenin 7	277-279	Pro	Gln	Pro	Pro			x
Vitellogenin 7	370-372	Val	Leu	Val	Val			
Vitellogenin 7	451-453	Lys	Asn	Lys	Lys	x		
Vitellogenin 7	514-516	Val	Ala	Ile	Ala			
Vitellogenin 7	520-522	Asn	Ser	Ser	Asn			

Gene	Position	Zebrafish	Red shiner	RG silvery minnow	fathead minnow	Char.	Polar.	Sp. Case
Vitellogenin 7	529-531	Pro	Thr	Pro	Thr			x
Vitellogenin 7	541-543	His	Gln	His	His	x		
Vitellogenin 7	700-702	Asn	Ser	Ala	Ala		x	
Vitellogenin 7	1534-1536	Ala	Asp	NO DATA	Ala	x		
Vitellogenin 7	1648-1650	Ala	Thr	NO DATA	Ala		x	
Vitellogenin 7	1696-1698	Ala	Ala	NO DATA	Ser		x	
Vitellogenin 7	1804-1806	Ile	Ile	NO DATA	Asn		x	
Vitellogenin 7	1867-1869	Arg	His	NO DATA	Arg			
Vitellogenin 7	2311-2313	Glu	Arg	Leu	Arg	x		
Vitellogenin 7	2314-2316	Ala	Ala	Val	Ala			
Vitellogenin 7	2437-2439	Ala	Gly	Ala	Ala			x
Vitellogenin 7	2647-2649	Thr	Thr	Ile	Thr		x	
Vitellogenin 7	2668-2670	Ala	Val	Val	Phe			
Vitellogenin 7	3682-3684	Thr	Thr	Ile	Thr		x	
Vitellogenin 7	3691-3693	Phe	Val	Phe	Val			
Vitellogenin 7	3874-3876	Phe	NO DATA	Phe	Leu			
Vitellogenin 7	3934-3936	Lys	Thr	Arg	Arg	x		
Vitellogenin 7	3946-3948	Phe	Ile	Val	Val			
Vitellogenin 7	4030-4032	Phe	NO DATA	Leu	Phe			
Zona pellucida glycoprotein 2.2	715-717	Ala	Ala	NO DATA	Ser		x	
Zona pellucida glycoprotein 2.2	1174-1176	Ser	Ser	NO DATA	Ala		x	
Zona pellucida glycoprotein 2.3	859-861	Ala	Ala	NO DATA	Ser		x	
Zona pellucida glycoprotein 2.3	925-927	Thr	Thr	NO DATA	Asn			
Zona pellucida glycoprotein 2.3	955-957	Asn	His	NO DATA	Arg			
Zona pellucida glycoprotein 2.6	709-711	Ala	Ser	NO DATA	Thr			
Zona pellucida glycoprotein 2.6	754-756	Val	Met	NO DATA	Val		x	
Zona pellucida glycoprotein 2.6	934-936	Asn	His	NO DATA	Arg			
Zygote arrest 1-like	805-807	Cys	Ser	Cys	Cys			
Zygote arrest 1-like	898-900	Gly	Ser	Gly	Ser			x

Dissertation Summary

In this dissertation research, we employed a genes-to-community approach toward understanding the environmental cues, genetic mechanisms, and community dynamics of reproductive phenology of Rio Grande fishes. In chapter one, we demonstrated that while rank order of spawning of Rio Grande fish species was largely consistent across a four year dataset, absolute timing differed substantially among years. Dry years with earlier spring snowmelt runoff appeared to result in concomitant earlier spawning across the entire fish community. Interannual shifts in reproductive phenology across years were not uniform across species, but rather the magnitude of shifts appeared to be related to rank order of spawning: earlier spawning species advanced spawning less in dry years than did later spawning species. As a result of the different magnitudes of shifts, the overlap in spawning seasons increased in dry years with earlier spring snowmelt runoff. Conversely, the amount of temporal partitioning of resources among species decreased as a result of lower flows. It is not known how differential phenological responses to environmental variation among species will affect reproductive success of particular species. However, decreased temporal partitioning is of concern because larvae of many species of freshwater fishes exhibit substantial overlap in resource use (e.g., Turner et al. 2010). One possible outcome of such decreased temporal partitioning of resources is that competition dynamics will be exacerbated and reproductive success of some species could be altered. Many climate change scenarios have forecasted increasingly dry conditions in the southwestern United States, with earlier and diminished spring snowmelt runoff arising from warming temperatures (IPCC 2007). If these scenarios prove true, then community dynamics and biotic interactions of fishes in arid-land rivers are likely to be altered.

Future research is needed to assess the ecological implications and generality of altered phenological shifts arising from interannual environmental variation.

The dynamic interactions of environmental variation and reproductive phenology discussed in chapter 1 motivated research presented in the second chapter of this dissertation, looking at among-species allelic variation in a key circadian rhythm gene, *Clock*. In chapter 2, we presented DNA sequence level variation in *Clock* that is consistent with a combination of functional constraint of this gene (given conserved amino acid sequences across species), phylogenetic inertia (given the apparent non-independence of allelic variation in *Clock* among related species), and possibly being shaped by reproductive timing (given the correlation between *Clock* allele length and phenology). We conclude that the structure of the *Clock* gene was likely shaped by all three of these processes over evolutionary time to match the particular circadian phenotypic needs of these species. However, future research will be needed to assess whether these patterns hold as we examine a broader sampling of species. Unfortunately, however, the paucity of high resolution reproductive phenology data limit the number of comparisons that can be made between *Clock* allele length and reproductive timing.

DNA sequence and allele length analyses suggested that for most of the Rio Grande cyprinid species, a single allele is present at very high frequency, with one or a few additional alleles present at low or very low frequencies. From a conservation perspective, high amounts of genetic variation are often assumed (implicitly or explicitly) to facilitate long-term persistence of populations by allowing a genetic response to novel biotic or abiotic challenges (e.g., introduction of new pathogens). However, conservation genetics studies are often conducted on putatively selectively neutral markers (such as microsatellites), which more

accurately reflect demographic history. In chapter 2, we present evidence of functional genetic variation in a key circadian rhythms gene, *Clock*, which is likely to be under natural selection. I argue that the allelic variation we observe in *Clock* is precisely the kind of variation that could facilitate long-term persistence of these species in the face of changing environments (e.g., climate change). This underscores the importance of maintaining sufficiently large populations such that rare alleles in *Clock* and other functional loci are not lost to genetic drift.

While variation in *Clock* could underlie differences in reproductive timing among Rio Grande fishes, we also sought to characterize functional variation across a broad suite of genes potentially associated with among-species differences in reproductive biology, circadian rhythms, and phenology. In chapter 3 we utilized recently-developed next-generation DNA sequencing technology to partially characterize transcriptomes of three Rio Grande cyprinid fishes. We used the well annotated zebrafish (*Danio rerio*) genome as a reference and source of functional annotation and gene ontology information. Resulting gene ontology terms were then queried to identify and select candidate genes with functions putatively associated with reproduction and circadian rhythms. We identified a total of 86 candidate genes, of which 51 exhibited among-species amino acid substitutions. Such variation at the amino acid level could underlie some of the differences we see among these taxa. We also identified 15 genes will simple repeats in their amino acid sequences. Notably, many of these genes are circadian rhythm genes with poly-glutamine domains similar to that found in *Clock1a* in chapter 2. In particular, we observed a poly-glutamine domain in the gene *Cryptochrome2a* that differs in length between zebrafish and fathead minnow (*Pimephales promelas*). *Cryptochrome2a* sequences were not available for the other two species in this study. However, the functional

significance of the poly-glutamine region of this gene has not been studied to our knowledge. One possibility is that the high number of *Cryptochrome* paralogs in cyprinid fishes (n=6) has facilitated neo-functionalization in these genes relative to mammals and other non-teleosts (Tamai et al. 2007).

Research presented in the three chapters of this dissertation advance our understanding of the reproductive phenology in fishes of the Rio Grande, New Mexico. This research employed a combination of ecological (chapter 1) and genetics approaches (chapters 2 and 3), as well as community-level (chapter 1) and comparative (chapters 2 and 3) study. We found evidence for environmental drivers of variation in reproductive timing of these fishes, as well as candidates for genetic mechanisms and pathways by which natural selection could shape reproductive phenology in this fish community.

Literature Cited

Intergovernmental Panel on Climate Change (2007) Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K., & Reisinger, A. (Eds.)]. IPCC, Geneva, Switzerland, 104 pp.

Tamai TK, Young LC, Whitmore D (2007) Light signaling to the zebrafish circadian clock by *Cryptochrome 1a*. Proceedings of the National Academy of Sciences of the USA 104(37):14712-14717.

Turner TF, Krabbenhoft TJ, Burdett AS (2010) Reproductive phenology and fish community structure in an arid-land river system. *In: Community Ecology of Stream Fishes* (Gido, K, & Jackson, D, Eds.). *American Fisheries Society Symposium* 73:427-446.