1-22-2014

Fish Species Introductions Provide Novel Insights into the Patterns and Drivers of Phylogenetic Structure in Freshwaters

Angela L. Strecker
Portland State University, strecker@pdx.edu

Julian D. Olden
University of Washington - Seattle Campus

Citation Details

Strecker AL, Olden JD. 2014 Fish species introductions provide novel insights into the patterns and drivers of phylogenetic structure in freshwaters. Proc. R. Soc. B 281: 20133003. http://dx.doi.org/10.1098/rspb.2013.3003

RH: Freshwater fish community phylogenetics

Fish species introductions provide novel insights into the patterns and drivers of phylogenetic structure in freshwaters

Angela L. Strecker[1] and Julian D. Olden[2]

[1] Portland State University, Department of Environmental Science and Management, Portland, OR, 97207, tel. 503-725-2427, fax. 503-725-9040, email. strecker@pdx.edu

[2] University of Washington, School of Aquatic and Fishery Sciences, Seattle, WA, 98105

Despite longstanding interest from terrestrial ecologists, freshwater ecosystems are a fertile, yet unappreciated, testing ground for applying community phylogenetics to uncover mechanisms of species assembly. We quantify phylogenetic clustering and overdispersion for native and non-native fishes of a large river basin in the American Southwest to test for the mechanisms (environmental filtering vs. competitive exclusion) and spatial scales influencing community structure. Contrary to expectations, non-native species were phylogenetically clustered and related to natural environmental conditions, whereas native species were not phylogenetically structured, likely reflecting human-related changes to the basin. The species that are most invasive (in terms of ecological impacts) tended to be most phylogenetically divergent from natives across watersheds, but not within watersheds, supporting the hypothesis that Darwin's naturalization conundrum is driven by spatial scale. Phylogenetic distinctiveness may facilitate non-native establishment at regional scales, but environmental filtering restricts local membership to closely-related species with physiological tolerances for current environments. By contrast, native species may have been phylogenetically clustered in historical times, but species loss from contemporary populations by anthropogenic activities has likely shaped the phylogenetic signal. Our study implies that fundamental mechanisms of community assembly have changed, with fundamental consequences for the biogeography of both native and non-native species.

Keywords: community phylogenetics, desert ecology, Lower Colorado River Basin, biogeography

Introduction

Understanding patterns of community assembly and the factors that determine the biogeography of species remain central themes in ecology. Although empirical tests and derivation of assembly rules have yielded great insight [1,2], landscape scale studies are hindered by poor understanding of the historical factors that influence biogeography and ultimately, community structure. Introductions of non-native species present novel opportunities to uncover the mechanisms that structure communities [3], enabling broad scale experimental study of the ecological and evolutionary processes that determine community assembly.

Community phylogenetics has recently emerged as a promising tool in the field [4,5]. It has been hypothesized that competitive exclusion is the primary mechanism driving assembly when communities are composed of distantly related members [4-6], but that this so-called phylogenetic overdispersion may also result from environmental filtering on convergent traits [4,7]. By contrast, communities composed of closely related members (i.e., phylogenetic clustering) are hypothesized to be structured by environmental filtering on shared physiological tolerances when traits are conserved [4,5]. Competition could also lead to character displacement, however, where close relatives diverge ecologically [8], generating a clustering pattern [9]. Adding to this complexity is the influence of spatial scale, which can alter the signal of phylogenetic relatedness [10]. Thus, interpretations of phylogenetic community structure are likely complicated by incomplete knowledge of the mechanisms and spatial scales that influence particular communities.

More recently, the use of phylogenetic beta diversity has been proposed to elucidate patterns of change in phylogenetic community structure across space. Phylogenetic beta diversity measures divergence across pairs of communities in different locations and is a complementary

63 approach to local community phylogenetic analyses by implicitly considering issues of spatial
64 scaling through incorporation of environmental filters and barriers to dispersal [11]. This
65 combined approach demonstrated that phylogenetic beta diversity for hummingbirds was greater
66 along steep environmental gradients in the Andes Mountains, resulting in phylogenetic clustering
67 in the harsher high elevation sites, but a tendency to overdispersion in less harsh lower elevations
68 [11]. The presence of strong environmental gradients can thereby generate distinct patterns of
69 phylogenetic structure with unique mechanistic explanations.

70 Although there is mounting evidence of both phylogenetic clustering and overdispersion
71 in plant, animal, and bacterial communities from a range of ecozones [6,12], the majority of past
72 studies were conducted on primary producers in terrestrial ecosystems [7], limiting geographic
73 and taxonomic generality. By contrast, freshwater ecosystems, and in particular, freshwater
74 fishes, present a fertile testing ground for community phylogenetic hypotheses, stemming from
75 the unique physiographic and biogeographical constraints imposed by the aquatic landscape [13].
76 These constraints have led to a vast diversity of fishes in freshwater habitats worldwide. A prime
77 example of this diversification occurred in the arid American Southwest, where fish communities
78 were shaped by a long geologic history (e.g., volcanism, isolation, marine intrusions) [14], and
79 harsh environmental conditions, including droughts, floods, and extreme temperatures, leading to
80 the evolution of a highly endemic fauna [15,16]. Dam construction, water diversions, and flow
81 regulation have significantly altered the environmental conditions in the region, creating
82 conditions that have enabled non-native species that are not adapted to harsh conditions to
83 survive and thrive, displacing native species in many regions [17,18]. The Lower Colorado River
84 Basin has been a flashpoint for the predicament of native species, where the highly endemic
85 ichthyofauna has precipitously declined over the 20th century [19,20], while over one hundred

86 non-native fish species from both neighbouring and distant waters have been introduced (with
87 greater than half established), often to create recreational fishing opportunities in newly
88 developed reservoir habitats [20,21]. Thus, the unique combination of species from diverse
89 geographic locations and broad environmental gradients that range from highly altered to more
90 extreme natural conditions will enhance our scientific understanding of community assembly for
91 freshwater fishes.

92 In our study, we embrace the highly variable phylogenetic contrast between native and
93 non-native fish species in the Lower Colorado River Basin (draining >360,000 km^2 of the
94 American Southwest), and their accompanying adaptive histories (or lack thereof), to test the
95 following three hypotheses.

96 *Hypothesis 1*: Native species in fish assemblages are phylogenetically clustered,
97 reflecting the strong influence of natural environmental conditions in structuring the evolution of
98 these species; non-native species in fish assemblages are overdispersed, reflecting the
99 competitive influences generated by anthropogenic alterations to systems. Non-native fishes in
100 the Southwest often outcompete native fishes under more stable, human-altered flow regimes
101 [15]. Additionally, diet studies suggest that non-natives compete intensely with each other [22],
102 thus it is reasonable to expect that competition is the dominant structuring force in non-native
103 communities. Correspondingly, the phylogenetic structure of native fishes will be highly
104 influenced by environmental drivers representing natural conditions, with functional traits that
105 represent adaptations to these environmental conditions; conversely, phylogenetic structure of
106 non-native fishes will be weakly related to variables representing contemporary human-related
107 conditions (and unrelated to natural conditions), as competition is the primary mechanism
108 determining community structure. This hypothesis is supported by the recent evolutionary

history of fish in the Lower Colorado River Basin, which has been generally constrained to relatively few families (Electronic Supplementary Material S1) [23]. By contrast, non-native fishes in the basin come from a much larger array of families (Electronic Supplementary Material S1) [24]. Conversely, it is possible that native species will be overdispersed, reflecting competitive interactions, whereas non-native species will be underdispersed as a result of the shared biological attributes that allow them in establish in new habitats. This may reflect the long history of sport fish stocking within the basin, including many closely related species from eastern North America [20].

Hypothesis 2: Phylogenetic beta diversity of native taxa is highly correlated with environmental differences between sites representing natural drivers; non-natives are less structured by natural environmental variation. Conversely, non-native phylogenetic beta diversity will be highly correlated with spatial variables and variables that reflect the anthropogenic component of species introduction and spread [25]; native fishes will be less spatially structured as a result of their long evolutionary history in the basin.

Hypothesis 3: Non-native species that are the most 'invasive' (in terms of ecological impacts) will show greater phylogenetic divergence from native species compared to non-native species that are not 'invasive' at both regional basin and local watershed scales. This provides direct insight into the so-called Darwin's naturalization conundrum: phylogenetic relatedness of non-native species to native communities is predicted to promote establishment because they share similar pre-adaptations to local environmental conditions with allied species, but at the same time may hamper establishment because of niche overlap with native species [26-28]. The latter is known as Darwin's naturalization hypothesis [29,30]. As the spatial scale of

131 consideration for Darwin's hypothesis influences observed patterns [27], we contrasted
132 phylogenetic divergence across the entire region, as well as within localized watersheds.
133
134 **Methods**
135 *Data collection*
136 We test the preceding three interconnected hypotheses on a unique large database of fish
137 species occurrences from the Lower Colorado River Basin [31]. The database contains >1.8
138 million records from museum, university, and government collections dating from 1840 to 2009
139 [24,31,32]. Our study focuses on fish species records collected after 1980 (>1.66 million
140 records), as this is considered representative of contemporary assemblages [20,33]. Further, this
141 time-frame broadly corresponds with the collection period of contemporary molecular sequence
142 data. Geographic data were reviewed for accuracy, as were regional species lists [31]. Fish were
143 collected using a variety of gears and techniques by different entities, and different studies had
144 different objectives (e.g., population- vs. community-level study). Thus, in order to control for
145 these biases, species presence was determined at the local reach scale (i.e., section of river
146 between two confluences), and only records that indicated community-level sampling were
147 retained [24]. Fish species records were then summarized at the aquatic ecological system (AES)
148 scale, which delineates regions by changes in landform, gradient, and stream size, and then
149 further divided into 387 AES, which we henceforth refer to as watersheds. Watersheds ranged
150 from 200 – 1600 km^2, and are a useful intermediate scale for our analyses. As a result of
151 geographical biases in sample collections, we excluded all watersheds with little or no sampling
152 effort, such that our final dataset was comprised of n = 159 total watersheds. There were n = 134
153 and n = 147 watersheds for native and non-native species only, respectively, with n = 122 total

154 watersheds for paired native – non-native comparisons (Hypothesis 1: differences in
155 phylogenetic structure of natives and non-natives within same watershed).
156
157 *Phylogenetic data*
158 Despite the recent explosion of molecular data available to infer phylogenetic
159 relationships among taxa, the diversity of freshwater fishes in North America represents a unique
160 challenge to scientists. This is particularly true of native fishes of the American Southwest,
161 which continue to be taxonomically revised [23], despite the species pool being relatively
162 depauperate. For example, in a large sequence database on freshwater fishes of North America (*n*
163 = 685 species) [34], native species from the Lower Colorado River Basin were largely under-
164 represented, with <50% of the species pool present in the database, whereas 88% of the non-
165 native species in our study were represented (A. Strecker, *unpublished*). Though studies
166 examining evolutionary history of southwestern endemics have yielded great insight [14,16,35],
167 we are aware of no phylogeny that encompasses all of the fish in this region, which may in part
168 reflect the absence of common molecular markers used across taxa in previous studies. Utilizing
169 sequence divergence data has been recommended for phylogenetic analysis of understudied taxa
170 [36], thus we have chosen the conservative approach of assessing sequence divergence for the
171 mitochondrial cytochrome b, which was the most represented DNA sequence for freshwater
172 fishes in the region (Electronic Supplementary Material S1).
173 We downloaded sequence data from PhyloTA [37], which searches GenBank for similar
174 regions, called phylogenetically informative clusters. Sequence data were obtained for 54 of 66
175 species (82%); of the 12 species for which there was no sequence data, 11 were native fish
176 species (Electronic Supplementary Material S1). Therefore, we used mitochondrial DNA

177 (mtDNA) sequences from a congener (10 species) or the closest relative in the dataset (2 species)
178 for unrepresented species. An analysis of the sensitivity of our results to this taxon substitution
179 was performed (Electronic Supplementary Material S2). For species that had multiple sequenced
180 individuals, a consensus sequence was constructed [38]. Sequences were aligned [39] and the
181 amount of sequence divergence between all native and non-native species was determined using
182 a Kimura 2-parameter model [40].
183
184 *Phylogenetic analyses*
185 To test our first hypothesis, we calculated mean phylogenetic distance (MPD) and mean
186 nearest neighbour phylogenetic distance (MNND) using sequence divergence data. MPD is an
187 intra-community or local measure that takes the average distance between all pairs of species
188 present in a watershed, whereas MNND is the average distance between each taxon and its most
189 closely related neighbour [7,41]. As these metrics are biased by species richness, we calculated
190 the standardized effect size (SES) by comparing the observed pattern to a null model using an
191 independent swap algorithm [42], which performs well (i.e., has low Type I error rates) for MPD
192 and MNND [43]. The algorithm holds the number of species per watershed constant, as well as
193 the frequency of occurrence of species across samples, and randomizes the occurrence matrix
194 [42]. There were 2000 matrix iterations and 5000 runs of the null model for each watershed. A
195 positive SES value indicates that species are overdispersed or evenly distributed throughout the
196 phylogeny, whereas negative SES indicates phylogenetic clustering. Only watersheds with ≥2
197 species were included, a constraint of the phylogenetic analyses. Analyses were performed
198 jointly, as well as separately on native and non-native sub-communities in watersheds; hereafter
199 we refer to these as native and non-native communities.

To test the relationship between phylogenetic divergence and functional divergence, we used five continuous biological traits for native fishes of the Colorado River Basin [17]: shape factor (the ratio of total body length to maximum body depth), swim factor (ratio of minimum depth of the caudal peduncle to the maximum depth of the caudal fin), maximum body length (mm), length at maturation (mm), and fecundity (total number of eggs or offspring per breeding season). These continuous traits describe some of the key dimensions of morphological and life history strategies exhibited by native fishes in this region [17]. As this test requires continuous variables, categorical traits could not be analyzed (e.g., trophic guilds).

To test our second hypothesis, we calculated phylogenetic beta diversity, which is an inter-community metric that assesses the MPD across watersheds considering the species that are present across all pairs of watersheds [7,41]. Larger values of phylogenetic beta diversity represent greater phylogenetic dissimilarity and smaller values represent less phylogenetic dissimilarity (i.e., greater similarity). A null model that shuffled the names of the taxa across the divergence matrix was used to evaluate results (n = 999 permutations), comparing the randomized results to observed results using the SES metric [44]. This null model is useful in that it holds constant species alpha and beta diversity, species occupancy, and spatial patterns, allowing for dispersal limitation of species to be controlled for [44] (see Electronic Supplementary Material S3). As with MPD, analyses were done on native and non-native communities in watersheds with ≥2 species.

To test our third hypothesis, we conducted a survey of 20 professional biologists with knowledge of regional fish communities to identify the non-native species that are considered most harmful to native fish species [45]. Following established methodology [46], we asked each survey respondent to classify non-native species as either being invasive (i.e., associated

223 ecological impact in their introduced range) or not. Non-native fishes selected by ≥75% of
224 experts as invasive were included in the analysis (Electronic Supplementary Material S1). These
225 invasive species also have spread at the greatest rate since introduction [20]. Phylogenetic
226 divergence was calculated across the entire region and in each watershed between: i) all pairs of
227 invasive and native species, ii) all pairs of remaining non-native (i.e., non-invasive) and native
228 species, and iii) all pairs of native species [30]. At the basin-scale, all recorded species were
229 compared. However, in order to test our hypothesis at the local watershed scale, we could only
230 include catchments that contained ≥2 species from each category (invasive, non-invasive, native)
231 (n = 85). We used an ANOVA followed by a Tukey HSD test to distinguish differences between
232 multiple comparisons. These pairwise comparisons are not independent of each other, therefore,
233 we used permutation tests (n = 199) to evaluate the significance of phylogenetic divergence
234 across species groups. This analysis was also repeated at the basin scale for non-native species
235 that were failed introductions [47], comparing all pairs of: i) successfully introduced non-natives
236 and natives, ii) all pairs of unsuccessfully introduced non-natives and natives, and iii) all pairs of
237 native species.
238
239 *Statistical analysis*
240 We assessed the influence of environmental and spatial factors on our intra- and inter-
241 community phylogenetic metrics by compiling data for 14 environmental variables known to be
242 important in structuring fish communities in this region [32]. These variables reflected both
243 natural features (e.g., seasonal precipitation, temperature, watershed area) and anthropogenic
244 influences (e.g., agriculture, canals, dams) (Electronic Supplementary Material S4).

245 At the local watershed scale, we assessed the effects of environmental variation on
246 phylogenetic structure using linear models. Preliminary tests indicated that errors were normally
247 distributed and that there was no significant spatial autocorrelation in the residuals, thus, general
248 linear models were sufficient for our purposes. We used a comparative model selection approach
249 to test our hypothesis that native and non-native community phylogenetic structure (i.e., SES)
250 would be better predicted by natural and anthropogenic descriptors of the environment,
251 respectively. Models of the full set of environmental variables were tested against subsets of
252 natural and anthropogenic environmental variables, and compared with Akaike's Information
253 Criterion (AIC), which penalizes models with larger numbers of variables [48]. We measured the
254 phylogenetic signal in functional traits of native species by constructing a phylogeny from
255 mtDNA and estimating Blomberg's *K*, which assumes a Brownian motion model of trait
256 evolution [49]. The phylogeny was constructed using maximum likelihood on a Tamura-Nei
257 model [40]. The phylogenetic tree is available in TreeBASE
258 (http://purl.org/phylo/treebase/phylows/study/TB2:S14973). Observed values of *K* were
259 compared to a null model that was generated by shuffling taxa labels across the phylogeny tips.
260 Lower values of *K* correspond to random or convergent evolutionary patterns, while higher
261 values indicate increasing trait conservatism.
262 We used multiple regression on distance matrices (MRM; $n = 4999$ permutations) to test
263 if environmental or spatial dissimilarity was related to phylogenetic patterns across watersheds
264 [50]. We used variation partitioning to examine the independent and joint effects of
265 anthropogenic environmental variables, natural environmental variables, and space on
266 phylogenetic beta diversity SES. We created separate Euclidean distance matrices for natural and
267 anthropogenic variables. All environmental variables were standardized to *z*-scores prior to

analysis. Spatial dissimilarity was calculated as the Euclidean distance between the centroids of all watersheds. While this approach has been criticized for underestimating explained variance [51], it is useful as a comparative tool for our purposes. A *t*-test with randomization was used to test for differences between native and non-native community phylogenetic beta diversity (n = 4999 permutations). All analyses were performed in R v2.12.1 [52]; phylogenetic metrics were calculated using the library *picante* [53] and MRMs using the library *ecodist* [54].

Results

Hypothesis 1: Local phylogenetic structure

On average, there were almost twice as many non-native fish species in watersheds (mean = 8.0 ± 4.1SD, range = 2-22) as there were native fish species (mean = 4.5 ± 1.6SD, range = 2-10). Pairwise sequence divergence between species ranged from 0.003 – 0.262 (mean = 0.163 ± 0.033SD), with 87% of values falling within the range of values considered optimal for mtDNA to uncover relationships [55]. Sensitivity analyses indicated that mean phylogenetic distance (MPD) results were relatively robust to taxon substitutions, but mean nearest neighbor phylogenetic distance (MNND) results were sensitive to taxon substitutions (Electronic Supplementary Material S2). This is not a surprising result given that MNND is evaluating the nearest neighbor and is therefore more focused on the terminal phylogenetic structure of the assemblage. Thus, MNND results will not be considered further. MPD was higher in native communities compared to non-native communities; however, native communities were not significantly phylogenetically structured ($t_{133} = 1.29$, $p = 0.20$) compared to the non-native communities, which exhibited significant phylogenetic clustering (i.e., negative MPD; $t_{146} = -3.32$, $p < 0.01$). When all species in a watershed were considered, the entire basin and most sub-

basins were significantly phylogenetically clustered (t_{158} = -5.71, $p < 0.01$). At the level of the individual watershed, 15% of non-native communities exhibited lower MPD than the null model expectation (95% confidence interval).

There was some evidence for geographic structure to the phylogenetic patterns, particularly for native fishes (Figure 1). When watersheds were grouped by historical biogeographic sub-basins, native fishes were significantly phylogenetically clustered in the Colorado sub-basin, but were significantly overdispersed in the Lower Gila, whereas non-native fishes were significantly overdispersed in the Colorado, but clustered in the Lower Colorado and Lower Gila (Electronic Supplementary Material S5). There were significant differences between native and non-native assemblages in some of the basins that had large contributing watersheds (Colorado and Lower Gila sub-basins) compared to the basins with relatively smaller watersheds (Electronic Supplementary Material S5).

Contrary to our hypothesis, the model with anthropogenic environmental variables was the most parsimonious for native fish phylogenetic structure, whereas the natural model received the most support for non-native fish community phylogenetic structure (Electronic Supplementary Material S4). Variability in summer precipitation was significant in models for both native and non-native fishes (full model: β_{native} = 0.33, p_{native} = 0.01; $\beta_{non\text{-}native}$ = -0.24, $p_{non\text{-}native}$ = 0.04), as was proximity to the nearest dam (anthropogenic model: β_{native} = -0.23, p_{native} = 0.01; $\beta_{non\text{-}native}$ = 0.24, $p_{non\text{-}native}$ = 0.01). Dam density (full model: β_{native} = 0.29, p_{native} = 0.02), watershed area (natural model: β_{native} = -0.268, p_{native} = 0.017), and reservoir surface area (full model: β_{native} = -0.38, p_{native} = 0.04) were also significant descriptors of native fish community phylogenetic structure. Overall model fit was poor, however, with the most parsimonious model

313 for native and non-native species explaining just 14 and 16% of variation, respectively ($p <$
314 0.001 for both models).
315 The phylogenetic signal of native fishes was significant for the functional traits shape
316 factor ($K = 1.10$, $p < 0.01$), maximum length ($K = 0.96$, $p < 0.01$), and length at maturation ($K =$
317 1.09, $p < 0.01$), indicating moderate trait conservatism. However, phylogenetic signal was not
318 significant for swim factor ($K = 0.21$, $p = 0.41$) or fecundity ($K = 0.33$, $p = 0.20$), indicating
319 convergence of traits.
320
321 *Hypothesis 2: Phylogenetic beta diversity*
322 In general, phylogenetic beta diversity was significantly greater for native communities
323 (mean = 0.44 ± 0.01SE) compared to non-native communities (mean = 0.08 ± 0.01SE)
324 (randomization $p < 0.001$). Contrary to our hypothesis, native phylogenetic beta diversity was
325 more strongly correlated with anthropogenic environmental variables ($\beta = 0.24$, $p < 0.01$)
326 compared to natural environmental variables ($\beta = 0.09$, $p = 0.02$) using MRM ($R^2 = 0.16$, $p <$
327 0.01); however, phylogenetic beta diversity of non-native communities was more correlated with
328 natural ($\beta = 0.25$, $p < 0.01$) compared to anthropogenic descriptors of the environment ($\beta = 0.14$,
329 $p < 0.01$) ($R^2 = 0.13$, $p < 0.01$). Phylogenetic beta diversity in non-native communities was
330 weakly correlated with spatial distance ($\beta = 0.06$, $p = 0.05$), but phylogenetic beta diversity in
331 native communities was strongly correlated with distance ($\beta = 0.20$, $p < 0.01$). These results were
332 supported by variation partitioning analyses (Figure 2): space had minimal independent effects
333 on non-native communities, but was more influential for native communities. Additionally, there
334 was evidence for spatially-structured environmental gradients playing a substantial role in

335 structuring phylogenetic beta diversity for both native and non-native fishes (shared variation
336 between natural environmental variables and space; Figure 2).
337
338 *Hypothesis 3: Biotic interactions*
339 To test Darwin's naturalization hypothesis, we compared pairwise phylogenetic
340 divergence between native, invasive, and non-invasive fish species across the entire basin and
341 within each watershed. At the basin scale, both invasive and non-invasive fishes were, on
342 average, significantly more phylogenetically divergent from native species, compared to the
343 amount of divergence between all pairs of native fishes ($F_{2,1223} = 28.85$, $p < 0.01$; permutation
344 Tukey HSD $p = 0.01$) (Figure 3a). Additionally, invasive fish species were also significantly
345 more divergent from native taxa compared to non-invasive fishes (permutation Tukey HSD $p =$
346 0.02). However, at the local watershed scale, patterns were less resolved: in 28% of watersheds
347 invasive species were significantly divergent from native species, whereas non-invasive species
348 were significantly diverged from native species in 18% of watersheds (Figure 3b). Invasive
349 species were significantly divergent from non-invasive species in 1% of watersheds. Non-native
350 species that were successfully introduced were significantly more phylogenetically divergent
351 from native species (mean = 0.16 ±0.001SE; $F_{2,764} = 17.84$, $p < 0.01$; permutation Tukey HSD p
352 = 0.02) compared to unsuccessfully introduced non-natives at the basin scale (mean = 0.15
353 ±0.003SE; permutation Tukey HSD $p = 0.03$).
354
355 **Discussion**
356 Phylogenetic structure provides a powerful template for understanding the mechanisms of
357 community assembly and biogeography. The fishes of the American Southwest are a particularly

358 valuable faunal assemblage with which to test general hypotheses about phylogenetic patterns
359 and processes in aquatic environments as a result of the unique geological and evolutionary
360 history of the region. Using a comprehensive fish database for the Lower Colorado River Basin,
361 we were able to test hypotheses about: 1) *within* watershed patterns and drivers of phylogenetic
362 structure, 2) *between* watershed patterns and drivers of phylogenetic structure, and 3)
363 phylogenetic determinants of invasiveness.

364 We observed differences between native and non-native community phylogenetic
365 structure; however, the pattern did not match our expectation that native communities would be
366 significantly more phylogenetically clustered compared to non-native communities. Rather, non-
367 native assemblages (and entire assemblages) were phylogenetically clustered, whereas native
368 communities showed no significant phylogenetic structure. Our results concur with those for
369 exotic plant communities in California [56]; however the authors suggested that environmental
370 filters were not controlling the distribution of introduced plants due to the broad range size of
371 non-native species, combined with low phylogenetic beta diversity. On the contrary, we propose
372 that phylogenetic clustering of non-native fishes is the result of environmental filtering on shared
373 physiological tolerances (i.e., trait conservatism [5]). This conjecture is supported by the strong
374 responses of non-native phylogenetic structure to natural environmental variables compared to
375 native assemblages (Electronic Supplementary Material S4). Additionally, the significantly
376 higher correlation of phylogenetic beta diversity of the non-native fishes with environmental
377 variables compared to spatial distance suggests that the distribution of non-native fishes may be
378 more limited by environmental filters than by dispersal; the latter is likely to be unconstrained
379 due to human-mediated vectors of introduction. Patterns of significant phylogenetic clustering of
380 non-natives in the watersheds with the greatest upstream contributing area (e.g., Lower

Colorado, Lower Gila) suggest that relatively harsher environmental conditions in parts of the basin, such as variability in stream flow driven by summer precipitation, may influence phylogenetic structure. Indeed, variability in summer precipitation was significantly greater in the Lower Colorado and Lower Gila sub-basins compared to the Colorado (*t*-test: $t_{35} = -8.58$, $p < 0.01$; $t_{73} = -7.86$, $p < 0.01$; respectively). Differential effects of flow conditions on native and non-native fishes have previously been observed [18].

An intriguing alternative possibility is that non-native fish phylogenetic clustering represents a history of introduction within the basin, whereby closely related species from eastern North America were widely introduced as sport fish into western waterways (e.g., centrarchids, such as *Micropterus* spp. and *Lepomis* spp.)[20]. This may also apply to aquarium trade species introduced into the wild from relatively few families (e.g., Cichlidae). Thus, the pattern of clustering may represent the history of introduction rather than the establishment success of non-native fishes. The relatively brief evolutionary history of introduced fish species in the basin likely precludes ecological divergence of closely related species as a mechanistic explanation for phylogenetic clustering of non-native fishes.

Native fish communities were not significantly phylogenetically structured in most of the studied watersheds; several factors may have influenced these results. First, many native species have been locally extirpated from watersheds, including species from highly diverged groups. For example, there were 16 cyprinid species in our study; cyprinids show great evolutionary diversification in the Lower Colorado River Basin [57]. Seven cyprinids in our study are endangered, four are threatened, one is of special concern, and two are candidates for listing by the Endangered Species Act. On average, the range size of cyprinids has declined by >30% since the 1950s [range: -14 to 100%; 20], such that average occupancy for cyprinid species is just 15.4

404 km^2 in the basin [58]. Second, environmental conditions were already dramatically changed prior
405 to the contemporary time period (post-1980) used to characterize the fish communities, such that
406 closely related species that may once have been locally adapted are no longer at an advantage.
407 Despite our hypothesis that native fishes would be more influenced by the environmental
408 variables that they have evolved in response to historically, native phylogenetic structure both
409 within and among watersheds was more strongly related to anthropogenic variables (Figure 2;
410 Electronic Supplementary Material S4). It is striking that the only region where we observed
411 significant phylogenetic clustering in native fishes is the Colorado sub-basin, which contains the
412 Grand Canyon, and is therefore one of the most protected (i.e., a national park) and least
413 degraded regions of the entire basin [21], with the notable exception of downstream mainstem
414 impacts from Glen Canyon Dam. For some species, such as *Gila cypha* and *Catostomus*
415 *discobolus*, the Grand Canyon is the last remaining fraction of their historical range in the lower
416 basin [59]. Conversely, the only region where native species were significantly overdispersed
417 was in the Lower Gila; this sub-basin has some of the highest levels of anthropogenic threats
418 [21] and invasive species [32] in the basin. This suggests that in this region, human activities
419 may result in non-random extinctions [45] that can shift native communities along a
420 phylogenetic gradient from clustering to overdispersion. We found evidence for trait
421 conservatism in native fishes for some morphological characters (shape factor, maximum length)
422 and life history traits (length at maturation), but not for other traits (swim factor, fecundity). This
423 suggests that closely related native fishes had similar adaptations for the local environmental
424 conditions. Others have demonstrated the tendency of closely related native species to adopt
425 intermediate life history strategies (i.e., evolutionary "bet-hedging") [17], which is considered
426 adaptive in highly unpredictable environments. Thus, although native fishes may have been

427 closely related historically, with morphological and life history adaptations to local conditions,
428 contemporary assemblages no longer reflect this pattern.

429 Phylogenetic beta diversity demonstrates how phylogenetic structure changes across
430 space, adding a necessary landscape element to studies of community assembly [11]. Here, we
431 observed similar patterns as in the local watershed phylogenetic structure: native communities
432 were influenced by anthropogenic environmental factors and space, whereas non-native
433 communities were structured by natural environmental factors describing patterns of successful
434 establishment. These results are indicative that dispersal limitation was historically a significant
435 factor for fish communities; the lack of an independent spatial signal in the beta diversity of non-
436 natives suggests that these fishes are not dispersal limited, likely reflecting the role of human-
437 mediated spread. Further, the greater beta diversity of native communities compared to non-
438 native communities reinforces previous research that introductions of closely related fish taxa are
439 homogenizing fish community composition across the landscape [60] at different levels of
440 organization (i.e., taxonomic, functional, phylogenetic).

441 Darwin's naturalization conundrum has long been of interest to ecologists; it is only
442 recently that advances in molecular biology have enabled tests of the hypothesis using
443 phylogenetic distances, without the artificial constraints of taxonomy [30]. We found evidence to
444 support Darwin's naturalization hypothesis at the basin scale, where the most invasive species
445 were more phylogenetically divergent from native species compared to non-invasive species
446 (Figure 3). However, at the watershed scale, support for the hypothesis was weaker: invasive and
447 non-invasive fish communities in the majority of watersheds were not phylogenetically divergent
448 from native fishes. These results concur with those from Hypothesis 1, where the mean
449 phylogenetic divergence of native and non-native communities at the watershed scale was

largely insignificant (Table 1). This suggests that at local scales phylogenetic relatedness of non-native (both invasive and non-invasive) species to native communities reflects higher establishment potential because closely related species share similar pre-adaptations to local environmental conditions. Thus, both facets of Darwin's naturalization conundrum may be valid, but ultimately determined by spatial scale [27]. These results run counter to the hypothesis that environmental filters determine community composition at larger spatial scales and biotic interactions are more important at smaller spatial scales [2]. It may be that environmental filtering can only happen at small spatial scales in these desert ecosystems, where high variability and extreme conditions are the norm. Thus, at large spatial scales the ability of an introduced species to survive in this basin is predicated on its uniqueness compared to the species pool. Prior to human intervention, there was only one piscivorous fishes in the Lower Colorado Basin [33]. The introduction of vast numbers of non-native species into a relatively depauperate species pool guarantees that most introductions are of phylogenetically divergent species. This is supported by our finding that non-natives species that did not successfully establish were less phylogenetically divergent compared to non-natives that did successfully establish populations at the basin scale.

A caveat of our study was that native fish species were comparatively underrepresented in surveys of molecular sequence data. While we were able to use sequences from close relatives for all unrepresented species, this constitutes a potential bias in our data. Sensitivity analyses indicated that these substitutions had minimal effects on MPD, but increased the likelihood of detecting clustering with the MNND metric. Future studies should use caution in interpreting results of MNND analyses when taxon substitutions are used. Substitution of close relatives is common practice in phylogenetic studies [56], as not all taxonomic groups have adequate

representation, highlighting the importance of broad classification databases [34]. This study represents the first attempt at bringing together phylogenetic and biogeographic characters of the entire native fish fauna of the Lower Colorado River Basin into a single synthesis. Additional investigations are needed when more resolved data becomes available.

Introductions of non-native species provide unique opportunities to resolve mechanisms of community assembly by creating natural experiments across different spatial scales. Our study provides evidence that native and non-native fishes of the Lower Colorado River Basin have distinct phylogenetic structure, which is being driven by a combination of harsh natural environmental conditions such as flooding, but also by human-influenced variables, such as flow regulation by dams and reservoir creation in the basin. By utilizing the distinctive geological and physiographical limitations that structure freshwater fishes, our study demonstrates that while some patterns of phylogenetic structure may be generalizable across taxa (i.e., phylogenetic clustering of non-natives)[56], others may be less universal, underscoring the importance of testing mechanisms of community assembly more broadly across taxonomic groups.

Acknowledgements

We thank two anonymous reviewers, Marlis Douglas, and Michael Douglas for their constructive feedback, as well as Jodi Whittier, Craig Paukert, Ben Stewart-Koster, Thomas Pool, Jesse Klinger, and Jared Anderson for database construction and assistance. Funding was provided by the U.S. Geological Survey Status and Trends Program, National Gap Analysis Program, and the National Climate Change and Wildlife Science Center.

References

1. Violle C., Nemergut D.R., Pu Z., Jiang L. 2011 Phylogenetic limiting similarity and competitive exclusion. *Ecol Lett* **14**, 782-787.
2. Weiher E., Keddy P.A. 1995 Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos* **74**, 159-164.
3. Sax D.F., Stachowicz J.J., Brown J.H., Bruno J.F., Dawson M.N., Gaines S.D., Grosberg R.K., Hastings A., Holt R.D., Mayfield M.M., et al. 2007 Ecological and evolutionary insights from species invasions. *Trends Ecol Evol* **22**, 465-471.
4. Cavender-Bares J., Kozak K.H., Fine P.V.A., Kembel S.W. 2009 The merging of community ecology and phylogenetic biology. *Ecol Lett* **12**, 693-715.
5. Webb C.O., Ackerly D.D., McPeek M.A., Donoghue M.J. 2002 Phylogenies and community ecology. *Annu Rev Ecol Syst* **33**, 475-505.
6. Cavender-Bares J., Ackerly D.D., Baum D.A., Bazzaz F.A. 2004 Phylogenetic overdispersion in Floridian oak communities. *Am Nat* **163**, 823-843.
7. Vamosi S.M., Heard S.B., Vamosi J.C., Webb C.O. 2009 Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol Ecol* **18**, 572-592.
8. Grant P.R., Grant B.R. 2006 Evolution of character displacement in Darwin's finches. *Science* **313**, 224-226.
9. Mayfield M.M., Levine J.M. 2010 Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol Lett* **13**, 1085-1093.
10. Swenson N.G., Enquist B.J., Pither J., Thompson J., Zimmerman J.K. 2006 The problem and promise of scale dependency in community phylogenetics. *Ecology* **87**, 2418-2424.

11. Graham C.H., Parra J.L., Rahbek C., McGuire J.A. 2009 Phylogenetic structure in tropical hummingbird communities. *Proc Natl Acad Sci USA* **106**, 19673-19678.

12. Helmus M.R., Savage K., Diebel M.W., Maxted J.T., Ives A.R. 2007 Separating the determinants of phylogenetic community structure. *Ecol Lett* **10**, 917-925.

13. Olden J.D., Kennard M.J., Leprieur F., Tedesco P.A., Winemiller K.O., Garcia-Berthou E. 2010 Conservation biogeography of freshwater fishes: recent progress and future challenges. *Divers Distrib* **16**, 496-513.

14. Spencer J.E., Smith G.R., Dowling T.E. 2008 Middle to late Cenozoic geology, hydrography, and fish evolution in the American Southwest. In *Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives: Geological Society of America Special Paper 439* (eds. Reheis M.C., Hershler R., Miller D.M.). Boulder, CO, The Geological Society of America.

15. Meffe G.K. 1984 Effects of abiotic disturbance on coexistence of predator-prey fish species. *Ecology* **65**, 1525-1534.

16. Smith G.R., Dowling T.E., Gobalet K.W., Lugaski T., Shiozawa D.K., Evans R.P. 2002 Biogeography and timing of evolutionary events among Great Basin fishes. In *Great Basin Aquatic Systems History* (eds. Hershler R., Madsen D.B., Currey D.R.). Washington, Smithsonian Contributions to the Earth Sciences, Number 33.

17. Olden J.D., Poff N.L., Bestgen K.R. 2006 Life-history strategies predict fish invasions and extirpations in the Colorado River Basin. *Ecol Monogr* **76**, 25-40.

18. Propst D.L., Gido K.B., Stefferud J.A. 2008 Natural flow regimes, nonnative fishes, and native fish persistence in arid-land river systems. *Ecol Appl* **18**, 1236-1252.

19. Minckley W.L., Deacon J.E. 1968 Southwestern fishes and the enigma of "endangered species". *Science* **159**, 1424-1432.

20. Olden J.D., Poff N.L. 2005 Long-term trends of native and non-native fish faunas in the American Southwest. *Animal Biodiversity and Conservation* **28**, 75-89.

21. Paukert C.P., Pitts K.L., Whittier J.B., Olden J.D. 2011 Development and assessment of a landscape-scale ecological threat index for the Lower Colorado River Basin. *Ecol Indic* **11**, 304-310.

22. Pilger T.J., Gido K.B., Propst D.L. 2010 Diet and trophic niche overlap of native and nonnative fishes in the Gila River, USA: implications for native fish conservation. *Ecology of Freshwater Fish* **19**, 300-321.

23. Minckley W.L., Marsh P.C. 2009 *Inland Fishes of the Greater Southwest: Chronicle of a Vanishing Biota*. Tucson, USA, University of Arizona Press.

24. Strecker A.L., Olden J.D., Whittier J.B., Paukert C.P. 2011 Defining conservation priorities for freshwater fishes according to taxonomic, functional, and phylogenetic diversity. *Ecol Appl* **21**, 3002-3013.

25. Leprieur F., Olden J.D., Lek S., Brosse S. 2009 Contrasting patterns and mechanisms of spatial turnover for native and exotic freshwater fish in Europe. *J Biogeogr* **36**, 1899-1912.

26. Diez J.M., Sullivan J.J., Hulme P.E., Edwards G., Duncan R.P. 2008 Darwin's naturalization conundrum: dissecting taxonomic patterns of species invasions. *Ecol Lett* **11**, 674-681.

27. Thuiller W., Gallien L., Boulangeat I., de Bello F., Münkemüller T., Roquet C., Lavergne S. 2010 Resolving Darwin's naturalization conundrum: a quest for evidence. *Divers Distrib* **16**, 461-475.

562 28. Strauss S.Y., Webb C.O., Salamin N. 2006 Exotic taxa less related to native species are more
563 invasive. *Proc Natl Acad Sci USA* **103**, 5841-5845.
564 29. Darwin C. 1859 *The Origin of Species*. London, J. Murray.
565 30. Schaefer H., Hardy O.J., Silva L., Barraclough T.G., Savolainen V. 2011 Testing Darwin's
566 naturalization hypothesis in the Azores. *Ecol Lett* **14**, 389-396.
567 31. Whittier J.B., Paukert C.P., Olden J.D., Pitts K.L., Strecker A.L. 2011 *Lower Colorado River*
568 *Basin aquatic gap analysis project: final report*. Reston, USA, U.S. Geological Survey, Gap
569 Analysis Program.
570 32. Pool T.K., Olden J.D., Whittier J.B., Paukert C.P. 2010 Environmental drivers of fish
571 functional diversity and composition in the Lower Colorado River Basin. *Can J Fish Aquat*
572 *Sci* **67**, 1791-1807.
573 33. Fagan W.F., Unmack P.J., Burgess C., Minckley W.L. 2002 Rarity, fragmentation, and
574 extinction risk in desert fishes. *Ecology* **83**, 3250-3256.
575 34. Ratnasingham S., Hebert P.D.N. 2007 BOLD: The Barcode of Life Data System
576 (www.barcodinglife.org). *Mol Ecol Notes* **7**, 355-364.
577 35. Douglas M.R., Douglas M.E. 2010 Molecular approaches to stream fish ecology. In
578 *Community ecology of stream fishes: concepts, approaches, and techniques* (eds. Gido K.B.,
579 Jackson D.A.), pp. 157-195. Bethesda, MD, American Fisheries Society.
580 36. Makowsky R., Cox C.L., Roelke C.E., Chippindale P.T. 2013 The relative utility of sequence
581 divergence and phylogenetic informativeness profiling in phylogenetic study design. *Mol*
582 *Phylogenet Evol* **66**, 437.
583 37. Sanderson M.J., Boss D., Chen D., Cranston K.A., Wehe A. 2008 The PhyloTA browser:
584 Processing GenBank for molecular phylogenetics research. *Syst Biol* **57**, 335-346.

585 38. Hall T.A. 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis
586 program for Windows 95/98/NT. *Nucl Acid S* **41**, 95-98.
587 39. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H.,
588 Valentin F., Wallace I.M., Wilm A., Lopez R., et al. 2007 Clustal W and Clustal X version
589 2.0. *Bioinformatics* **23**, 2947-2948.
590 40. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. 2011 MEGA5:
591 molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance,
592 and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739.
593 41. Webb C.O., Ackerly D.D., Kembel S.W. 2008 Phylocom: software for the analysis of
594 phylogenetic community structure and trait evolution. *Bioinformatics* **24**, 2098-2100.
595 42. Gotelli N.J., Entsminger G.L. 2003 Swap algorithms in null model analysis. *Ecology* **84**, 532-
596 535.
597 43. Hardy O.J. 2008 Testing the spatial phylogenetic structure of local communities: statistical
598 performances of different null models and test statistics on a locally neutral community. *J*
599 *Ecol* **96**, 914-926.
600 44. Swenson N.G., Erickson D.L., Mi X.C., Bourg N.A., Forero-Montana J., Ge X.J., Howe R.,
601 Lake J.K., Liu X.J., Ma K.P., et al. 2012 Phylogenetic and functional alpha and beta diversity
602 in temperate and tropical tree communities. *Ecology* **93**, S112-S125.
603 45. Olden J.D., Poff N.L., Bestgen K.R. 2008 Trait synergisms and the rarity, extirpation, and
604 extinction risk of desert fishes. *Ecology* **89**, 847-856.
605 46. Kolar C.S., Lodge D.M. 2002 Ecological predictions and risk assessment for alien fishes in
606 North America. *Science* **298**, 1233-1236.

607 47. United States Geological Survey. 2013 Nonindigenous Aquatic Species Database.
608 http://nas.er.usgs.gov/; accessed 26 July 2013.
609 48. Burnham K.P., Anderson D.R. 2002 *Model Selection and Multimodel Inference: A Practical*
610 *Information-Theoretic Approach*. 2nd ed. New York, USA, Springer.
611 49. Blomberg S.P., Garland T., Ives A.R. 2003 Testing for phylogenetic signal in comparative
612 data: Behavioral traits are more labile. *Evolution* **57**, 717-745.
613 50. Zhang J.-L., Swenson N.G., Chen S.-B., Liu X.-J., Li Z.-S., Huang J.-H. 2013 Phylogenetic
614 beta diversity in tropical forests: implications for the roles of geographical and environmental
615 distance. *J Syst Evol* **51**, 71-85.
616 51. Legendre P., Borcard D., Peres-Neto P.R. 2005 Analyzing beta diversity: partitioning the
617 spatial variation of community composition data. *Ecol Monogr* **75**, 435-450.
618 52. R Development Core Team. 2010 R: a language and environment for statistical computing.
619 (Vienna, Austria, R Foundation for Statistical Computing.
620 53. Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D.,
621 Blomberg S.P., Webb C.O. 2010 Picante: R tools for integrating phylogenies and ecology.
622 *Bioinformatics* **26**, 1463-1464.
623 54. Goslee S.C., Urban D.L. 2007 The ecodist package for dissimilarity-based analysis of
624 ecological data. *J Stat Softw* **22**, 1-19.
625 55. Makowsky R., Cox C.L., Roelke C., Chippindale P.T. 2010 Analyzing the relationship
626 between sequence divergence and nodal support using Bayesian phylogenetic analyses. *Mol*
627 *Phylogenet Evol* **57**, 485-494.

628 56. Cadotte M.W., Borer E.T., Seabloom E.W., Cavender-Bares J., Harpole W.S., Cleland E.,
629 Davies K.F. 2010 Phylogenetic patterns differ for native and exotic plant communities across
630 a richness gradient in Northern California. *Divers Distrib* **16**, 892-901.
631 57. Dowling T.E., Tibbets C.A., Minckley W.L., Smith G.R. 2002 Evolutionary relationships of
632 the plagopterins (Teleostei : Cyprinidae) from cytochrome b sequences. *Copeia*, 665-678.
633 58. Fagan W.F., Kennedy C.M., Unmack P.J. 2005 Quantifying rarity, losses, and risks for
634 native fishes of the lower Colorado River Basin: implications for conservation listing.
635 *Conserv Bio* **19**, 1872-1882.
636 59. Minckley W.L., Marsh P.C., Deacon J.E., Dowling T.E., Hedrick P.W., Matthews W.J.,
637 Mueller G. 2003 A conservation plan for native fishes of the lower Colorado River.
638 *BioScience* **53**, 219-234.
639 60. Pool T.K., Olden J.D. 2012 Taxonomic and functional homogenization of an endemic desert
640 fish fauna. *Divers Distrib* **18**, 366-376.
641

642 **Table and Figure Captions**

643 **Figure 1.** Standardized mean phylogenetic distance of a) native, b) non-native fish, and c) entire
644 fish communities by watershed (filled). Watersheds for which there was insufficient data are in
645 white. Positive values indicate phylogenetic overdispersion, negative values indicate
646 phylogenetic clustering. State and country boundaries are indicated with gray lines and italicized
647 font, and zoogeographical boundaries (see Electronic Supplementary Material S1) are the
648 thickened black lines and plain font. (Online version in colour)
649

650 **Figure 2.** Variation partitioning of phylogenetic beta diversity between anthropogenic (anthro)
651 and natural environmental variables, and space in (a) native and (b) non-native fish communities.
652
653 **Figure 3**. Pairwise phylogenetic divergence (mtDNA) between native species and invasive
654 species (black), non-invasive species (gray), and native species (white) for the entire region (a)
655 and within each watershed (b). See text for distinction of non-native species as invasive vs. non-
656 invasive. Boxplots show the 25^{th}, 50^{th}, and 75^{th} percentile, whiskers show the 10^{th} and 90^{th}
657 percentile, with circles representing outliers. Tukey HSD comparison indicated above the boxes
658 with lowercase letters ($p < 0.05$) in (a). In (b), percentages represent the number of watersheds
659 out of the total ($n = 85$) that were significantly different with Tukey HSD comparisons.

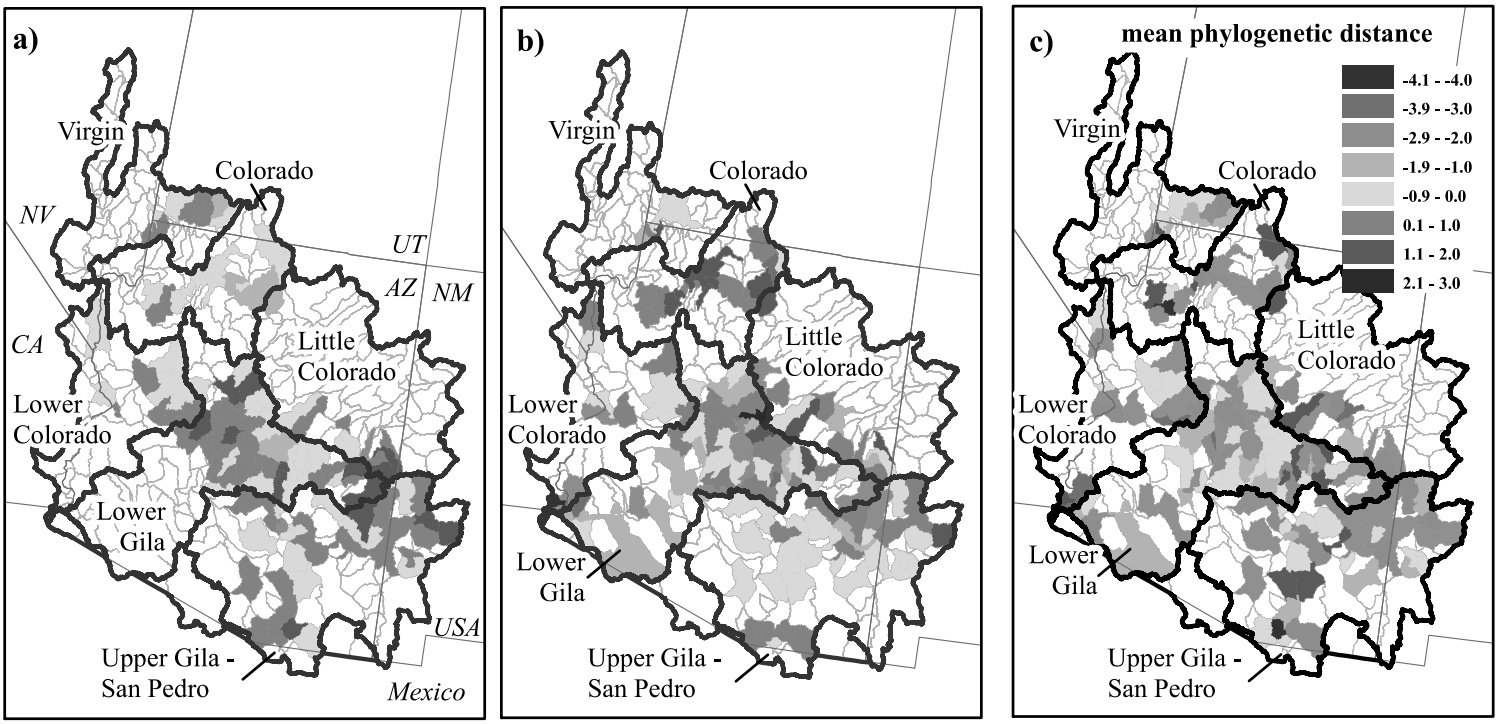

a)
Virgin
Colorado
NV
UT
AZ
NM
CA
Little
Colorado
Lower
Colorado
Lower
Gila
USA
Upper Gila -
San Pedro
Mexico
b)
Virgin
Colorado
Little
Colorado
Lower
Colorado
Lower
Gila
Upper Gila -
San Pedro
c)
mean phylogenetic distance
-4.1 - -4.0
-3.9 - -3.0
-2.9 - -2.0
-1.9 - -1.0
-0.9 - 0.0
0.1 - 1.0
1.1 - 2.0
2.1 - 3.0
Virgin
Colorado
Little
Colorado
Lower
Colorado
Lower
Gila
Upper Gila -
San Pedro

Figure 1

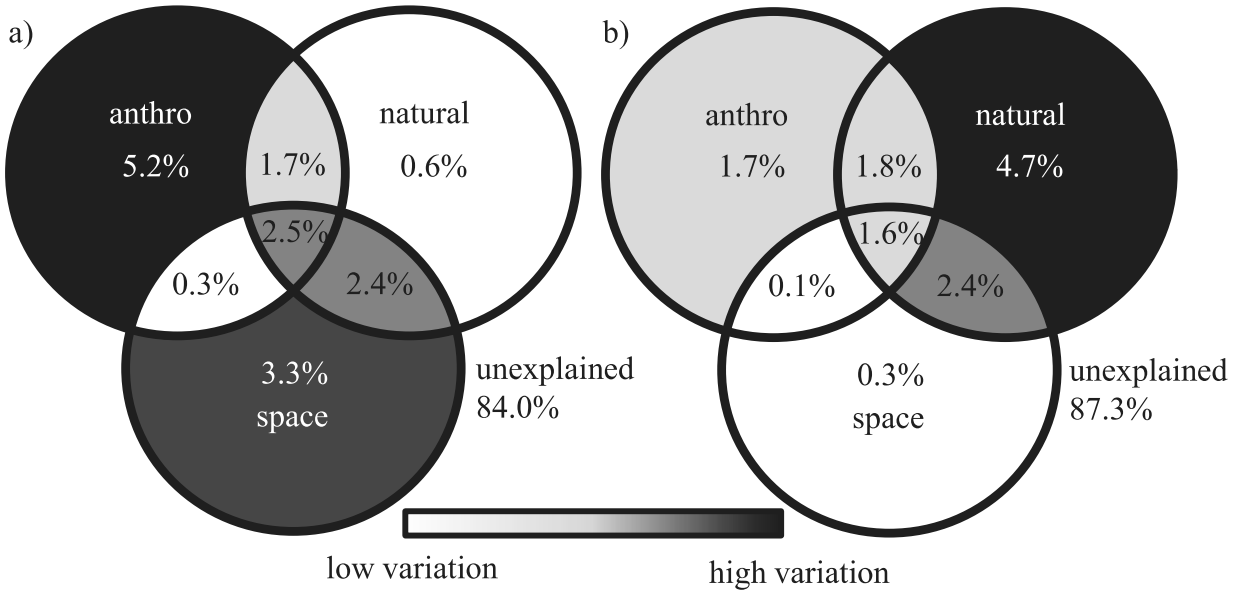

a)
anthro
5.2%
natural
0.6%
1.7%
2.5%
0.3%
2.4%
3.3%
space
unexplained
84.0%
b)
anthro
1.7%
natural
4.7%
1.8%
1.6%
0.1%
2.4%
0.3%
space
unexplained
87.3%
low variation
high variation

Figure 2

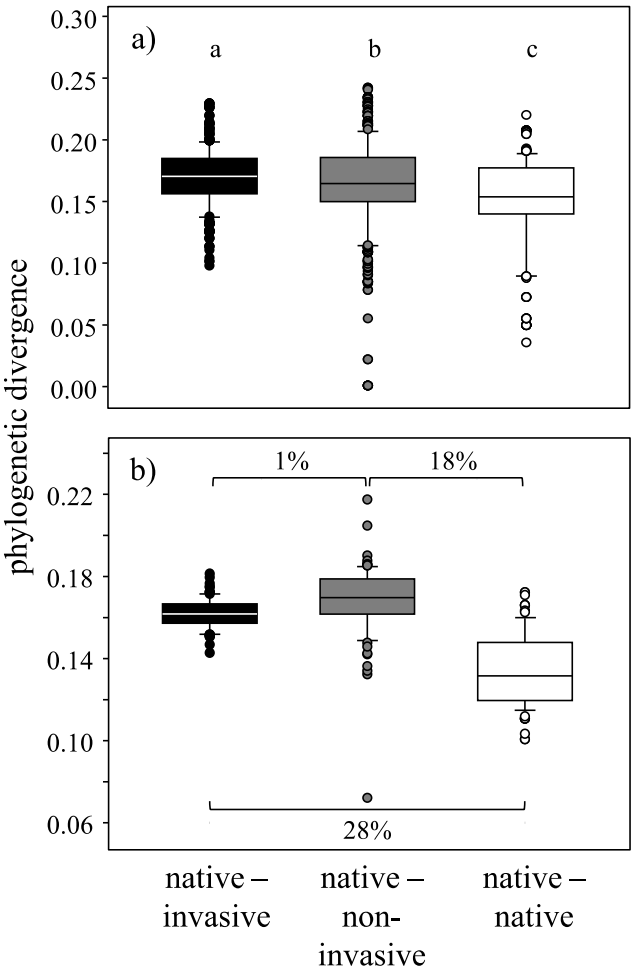

a)
a
b
c
0.30
0.25
0.20
0.15
0.10
0.05
0.00
b)
1%
18%
28%
0.22
0.18
0.14
0.10
0.06
phylogenetic divergence
native – invasive
native – non-invasive
native – native

Figure 3

Electronic Supplementary Material S1. *Molecular sequence data, sample location, and range coverage for freshwater fishes.*

Sequence data were obtained for 54 of 66 species (82%); of the 12 species for which there was no sequence data, 11 were native fish species (Table S1-1). Therefore, we used mitochondrial DNA (mtDNA) sequences from a congener (10 species) or the closest relative in the dataset (2 species) for unrepresented species. mtDNA has been used extensively in studies of phylogenetics and phylogeography. Although mtDNA can offer insight into the influence of historical patterns on populations [1], it has been criticized as biasing overall lineage history as a result of being maternally inherited and having relatively rapid substitution rates, reaching saturation and reducing phylogenetic signal [2]. A recommended solution is to compare phylogenetic patterns of mtDNA to nuclear DNA [3]; however, as this data is unavailable for our system, we have instead compared mtDNA sequence divergence to a qualitative phylogeny [4]. As the branch lengths of this phylogeny are not quantitative, we counted the number of nodes that separate each pair of species as a coarse estimate of divergence. This method has been used previously to represent phylogenetic diversity in the basin [5,6]. There was a highly significant correlation ($r = 0.73$, $p < 0.001$) between mtDNA divergence and the qualitative phylogeny (A. Strecker, unpublished), thereby validating the use of mtDNA sequence divergence for our analyses. Geographical localities were obtained from GenBank (when available; Figure S1-1). As expected, most sequences for non-native species are from outside the basin.

Table S1-1. Freshwater fish species used in the analysis, including information on taxonomic affiliation, number of cytochrome b mtDNA sequences in GenBank, and proportion of native species range in the Lower Colorado River Basin (defined as ecological drainage units: Electronic Supplementary Material S4) encompassed by molecular sequence data. Species used as a substitute indicated for species with no cytochrome b molecular data available.

species	common name	native/ non-native	family	number of cytochrome b sequences	range coverage
Agosia chrysogaster	longfin dace	native	Cyprinidae	4	0.5
Ambloplites rupestris	rock bass	non-native	Centrarchidae	26	
Ameiurus melas‡	black bullhead	non-native	Ictaluridae	4	
Ameiurus natalis‡	yellow bullhead	non-native	Ictaluridae	5	
Ameiurus nebulosus‡	brown bullhead	non-native	Ictaluridae	5	
Carassius auratus	goldfish	non-native	Cyprinidae	45	
Catostomus clarkii	desert sucker	native	Catostomidae	*Catostomus plebeius*	--
Catostomus discobolus	bluehead sucker	native	Catostomidae	*Catostomus plebeius*	--
Catostomus insignis	Sonora sucker	native	Catostomidae	*Catostomus plebeius*	--
Catostomus latipinnis	flannelmouth sucker	native	Catostomidae	*Catostomus plebeius*	--
Catostomus plebeius	Rio Grande sucker	non-native	Catostomidae	37	
Chaenobryttus gulosus‡	warmouth	non-native	Centrarchidae	2	
Ctenopharyngodon idella	grass carp	non-native	Cyprinidae	2	
Cyprinella lutrensis‡	red shiner	non-native	Cyprinidae	11	
Cyprinodon m. eremus	Sonoyta pupfish	native	Cyprinodontidae	*Cyprinodon macularius* (1)	0*
Cyprinodon m. macularius	desert pupfish	native	Cyprinodontidae	*Cyprinodon macularius* (1)	0*
Cyprinus carpio‡	common carp	non-native	Cyprinidae	1	
Dorosoma petenense	threadfin shad	non-native	Clupeidae	3	
Esox lucius‡	northern pike	non-native	Esocidae	217	
Fundulus zebrinus	plains killifish	non-native	Fundulidae	13	
Gambusia affinis‡	western mosquitofish	non-native	Poeciliidae	8	
Gila cypha	humpback chub	native	Cyprinidae	1	1.0
Gila elegans	bonytail	native	Cyprinidae	*Gila cypha*	--
Gila intermedia	Gila chub	native	Cyprinidae	*Gila robusta*	--
Gila nigra	headwater chub	native	Cyprinidae	*Gila robusta*	--
Gila robusta	roundtail chub	native	Cyprinidae	1	0*
Gila seminuda	Virgin chub	native	Cyprinidae	*Gila robusta*	--
Ictalurus punctatus‡	channel catfish	non-native	Ictaluridae	9	
Ictiobus bubalus	smallmouth buffalo	non-native	Catostomidae	40	
Ictiobus cyprinella	bigmouth buffalo	non-native	Catostomidae	31	
Lepidomeda mollispinis	Virgin R spinedace	native	Cyprinidae	2	1.0
Lepidomeda vittata	Little Colorado R spinedace	native	Cyprinidae	1	1.0

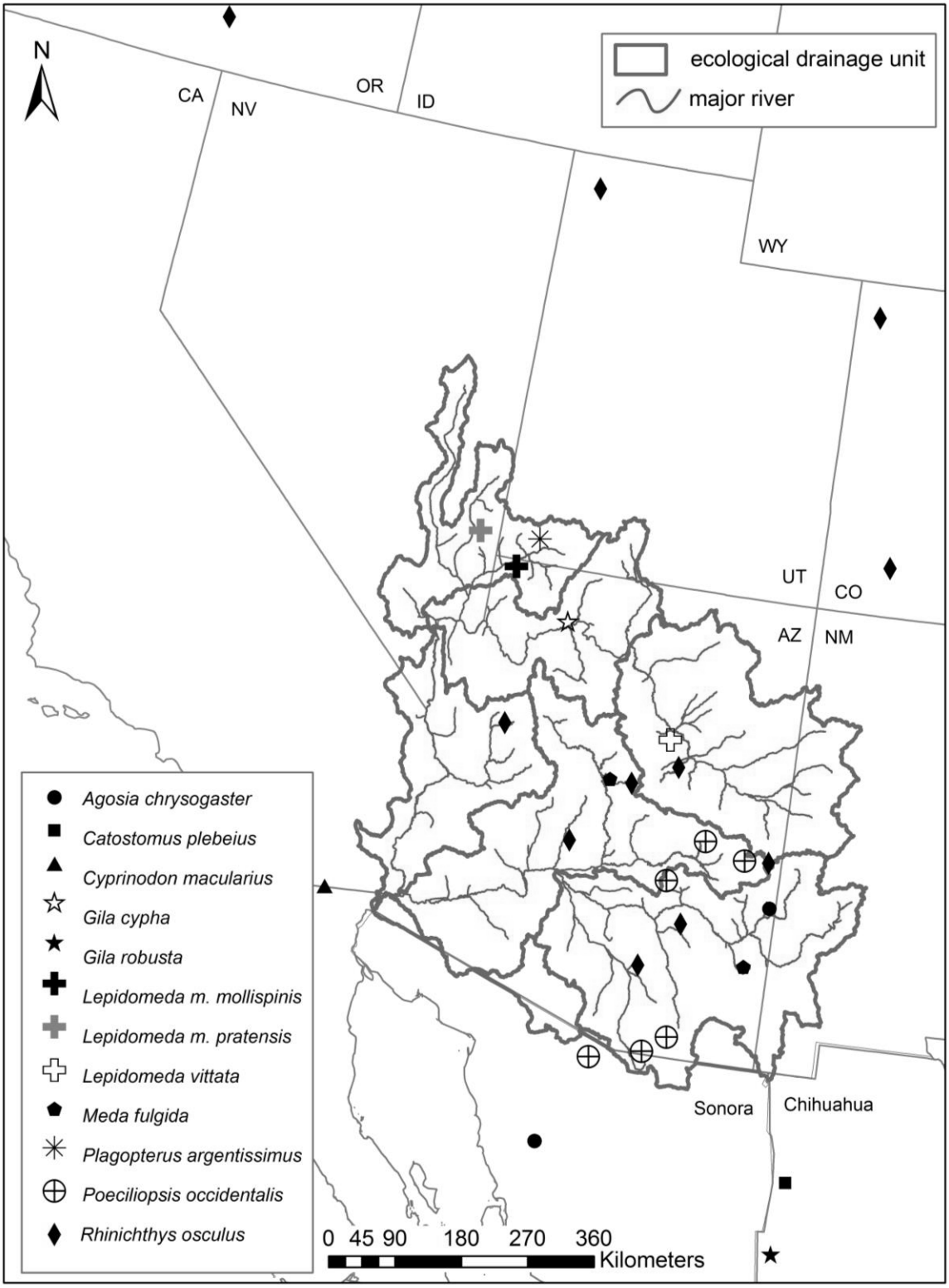
* = GenBank record indicates sample taken from outside the Lower Colorado River Basin
-- = molecular data from alternative species used
† = no location information available
‡ = classified as invasive in our analysis

Table S1-1, *cont'd*

species	common name	native/ non-native	family	number of cytochrome b sequences	range coverage
Lepomis cyanellus‡	green sunfish	non-native	Centrarchidae	4	
Lepomis macrochirus	bluegill	non-native	Centrarchidae	7	
Lepomis microlophus	redear sunfish	non-native	Centrarchidae	5	
Meda fulgida	spikedace	native	Cyprinidae	2	1.0
Micropterus dolomieu‡	smallmouth bass	non-native	Centrarchidae	14	
Micropterus punctulatus	spotted bass	non-native	Centrarchidae	7	
Micropterus salmoides‡	largemouth bass	non-native	Centrarchidae	14	
Morone chrysops	white bass	non-native	Moronidae	3	
Morone mississippiensis	yellow bass	non-native	Moronidae	1	
Morone saxatilis‡	striped bass	non-native	Moronidae	1	
Notemigonus chrysoleucas	golden shiner	non-native	Cyprinidae	1	
Oncorhynchus clarkii	cutthroat trout	non-native	Salmonidae	2	
Oncorhynchus apache	Apache trout	native	Salmonidae	*Oncorhynchus mykiss*	--
Oncorhynchus gilae	Gila trout	native	Salmonidae	*Oncorhynchus mykiss*	--
Oncorhynchus mykiss	rainbow trout	non-native	Salmonidae	46	
Oreochromis aureus‡	blue tilapia	non-native	Cichlidae	7	
Oreochromis mossambicus‡	Mozambique tilapia	non-native	Cichlidae	2	
Perca flavescens	yellow perch	non-native	Percidae	10	
Pimephales promelas‡	fathead minnow	non-native	Cyprinidae	2	
Plagopterus argentissimus	woundfin	native	Cyprinidae	1	0.50
Poecilia latipinna	sailfin molly	non-native	Poeciliidae	1	
Poecilia reticulata	guppy	non-native	Poeciliidae	9	
Poeciliopsis occidentalis	Gila topminnow	native	Poeciliidae	6	0.67
Pomoxis nigromaculatus	black crappie	non-native	Centrarchidae	4	
Pylodictis olivaris‡	flathead catfish	non-native	Ictaluridae	7	
Rhinichthys osculus	speckled dace	native	Cyprinidae	>200	0.67
Richardsonius balteatus	redside shiner	non-native	Cyprinidae	128	
Salmo trutta	brown trout	non-native	Salmonidae	60	
Salvelinus fontinalis	brook trout	non-native	Salmonidae	16	
Sander vitreus	walleye	non-native	Percidae	3	
Thymallus arcticus	arctic grayling	non-native	Salmonidae	2	
Tiaroga cobitis	loach minnow	native	Cyprinidae	*Rhinichthys osculus*	--
Tilapia zilli‡	redbelly tilapia	non-native	Cichlidae	*Oreochromis aureus*	
Xyrauchen texanus	razorback sucker	native	Catostomidae	1	†

* = GenBank record indicates sample taken from outside the Lower Colorado River Basin
-- = molecular data from alternative species used
† = no location information available
‡ = classified as invasive in our analysis

Figure S1-1. Approximate location of sampling for molecular sequence data of native species (or their closest relative; GenBank). Ecological drainage units are clusters of 8-digit hydrologic unit codes (HUCs) developed by the US Geological Survey that represent zoogeographic regions [7]. The Lower Colorado River Basin is the outer border.



References

1. Douglas M.R., Douglas M.E. 2010 Molecular approaches to stream fish ecology. *In* Community ecology of stream fishes: concepts, approaches, and techniques (eds. Gido K.B., Jackson D.A.), pp. 157-195. Bethesda, MD, American Fisheries Society.
2. Ballard J.W.O., Whitlock M.C. 2004 The incomplete natural history of mitochondria. *Mol Ecol* **13**, 729-744.
3. Zink R.M., Barrowclough G.F. 2008 Mitochondrial DNA under siege in avian phylogeography. *Mol Ecol* **17**, 2107-2121.
4. Fagan W.F., Unmack P.J., Burgess C., Minckley W.L. 2002 Rarity, fragmentation, and extinction risk in desert fishes. *Ecology* **83**, 3250-3256.
5. Strecker A.L., Olden J.D., Whittier J.B., Paukert C.P. 2011 Defining conservation priorities for freshwater fishes according to taxonomic, functional, and phylogenetic diversity. *Ecol Appl* **21**, 3002-3013.
6. Olden J.D., Poff N.L., Bestgen K.R. 2008 Trait synergisms and the rarity, extirpation, and extinction risk of desert fishes. *Ecology* **89**, 847-856.
7. Whittier J.B., Paukert C.P., Olden J.D., Pitts K.L., Strecker A.L. 2011 Lower Colorado River Basin aquatic gap analysis project: final report. Reston, USA, U.S. Geological Survey, Gap Analysis Program.

Electronic Supplementary Material S2. *Sensitivity analysis of taxon substitutions.*

Given that mtDNA sequences were missing for several species, a number of species pairs had a default value of 0.0 sequence divergence as a result of substituting the nearest relative. To test how sensitive our phylogenetic analyses were to this assumption, we averaged the divergence values of all species within a genus as a proxy for the closest relative. This average value represents an estimate of divergence for the species in which we used a taxon substitute. Mean phylogenetic distance (MPD) and mean nearest neighbor phylogenetic distance (MNND) were re-analyzed with this proxy divergence value in the mtDNA divergence matrix (Table S2-1).

Table S2-1. Sensitivity analysis for A) native communities and B) non-native communities when divergence values for all species pairs in which we used a taxon substitute were substituted with average values (replacing the 0.0 pairwise divergence). The top values in each table are the statistics for standardized effect sizes (SES) of MPD and MNND, whereas the bottom panel of each table is the number of watersheds that were significant at $p < 0.05$ for clustering and overdispersion.

A) native	**SES MPD**		**SES MNND**	
statistics	**original**	**taxon substitute**	**original**	**taxon substitute**
average	0.085	0.094	-0.287	-0.134
median	0.010	0.188	-0.470	-0.093
minimum	-2.277	-1.696	-2.293	-2.894
maximum	1.442	1.707	2.438	2.226
	MPD		**MNND**	
significant results	**original**	**taxon substitute**	**original**	**taxon substitute**
cluster	0	0	18	11
overdispersed	1	2	1	3

B) non-native	**SES MPD**		**SES MNND**	
statistics	**original**	**taxon substitute**	**original**	**taxon substitute**
average	-0.320	-0.292	-0.336	-0.233
median	-0.251	-0.254	-0.252	-0.214
minimum	-4.125	-3.945	-3.587	-2.851
maximum	2.090	2.090	2.249	2.246
	MPD		**MNND**	
significant results	**original**	**taxon substitute**	**original**	**taxon substitute**
cluster	22	17	18	13
overdispersed	5	4	6	6

Electronic Supplementary Material S3. *Null model of phylogenetic beta diversity*.
Figure S3-1. Plot of phylogenetic beta diversity mean phylogenetic distance (MPD) of a) native and b) invasive communities along gradients of Euclidean distance (m), natural environmental distance (as Euclidean distance), and anthropogenic environmental distance (as Euclidean distance). Black symbols represent observed values, whereas gray lines represent the mean and standard deviation of null model randomizations. Lines indicate Lowess fit of observed (dashed blue) and null model (solid red) data.

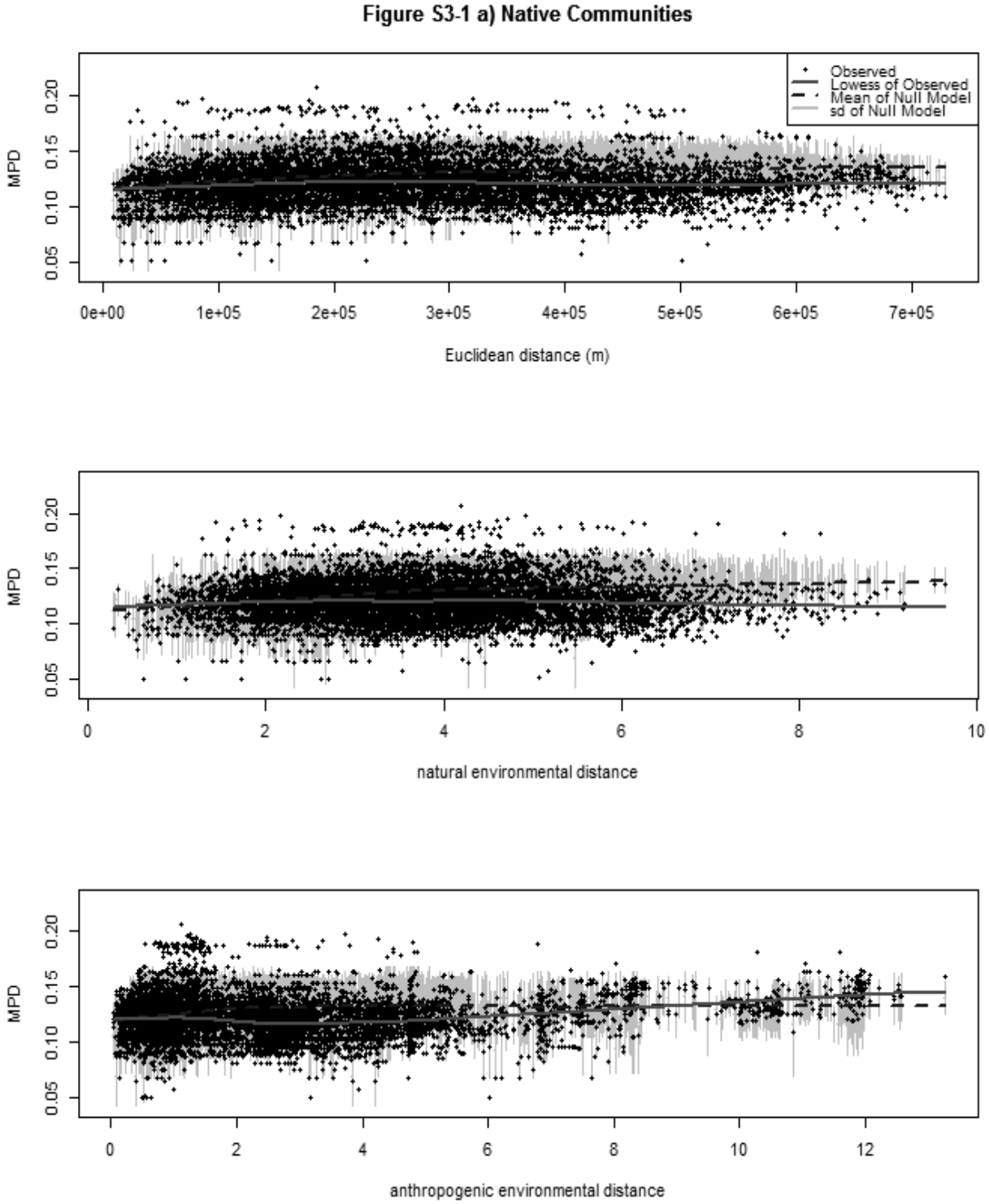
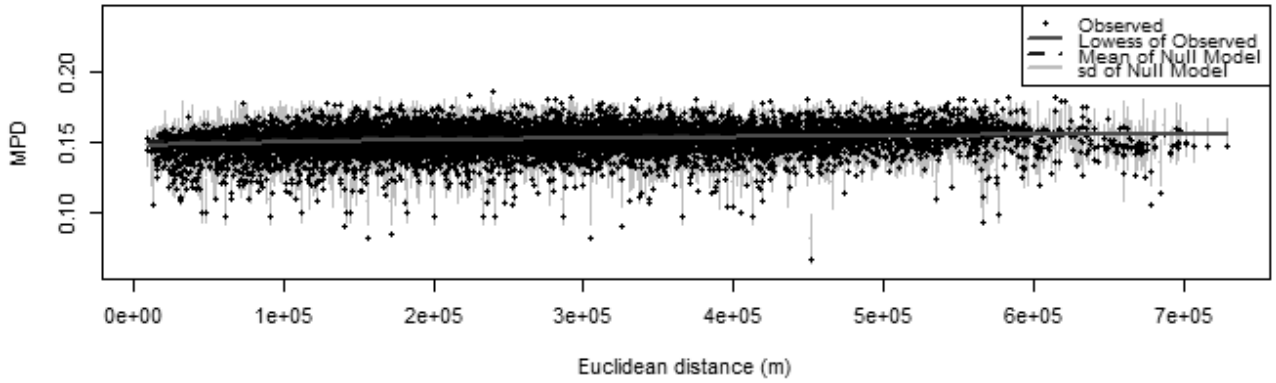
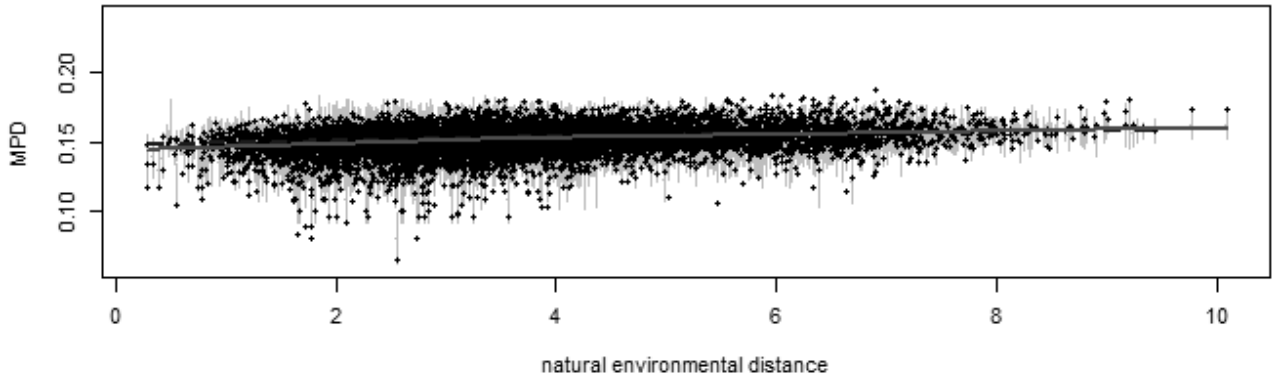
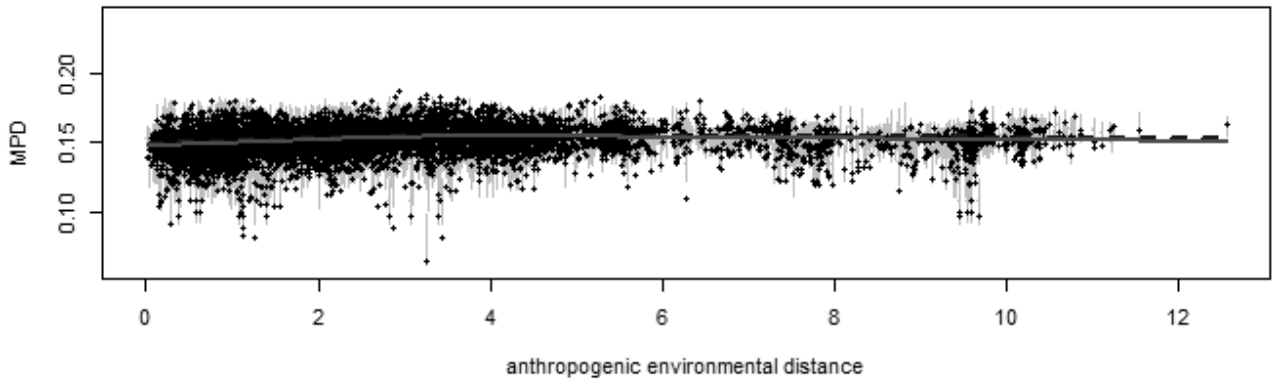


Figure S3-1 b) Invasive Communities







Electronic Supplementary Material S4. *General linear models and variables relating environmental factors to mean phylogenetic distance.*

Table S4-1. (A) Environmental variables used in model construction. Transformations were applied to variables that were not normally distributed. For some variables, a constant equal to the smallest value in the dataset was applied, as adding a constant of 1 can swamp the signal when there are very small values (i.e., <1). Although there are many other variables that may influence fish species, a number were excluded as a result of collinearity ($r > 0.8$). (B) General linear model comparisons for native and non-native fish community mean phylogenetic distance (MPD) in full (all environmental variables), natural (i.e., subset of natural environmental variables), and anthropogenic (i.e., subset of anthropogenic environmental variables) models. Significant variables in models and standardized coefficients (β) are indicated beside each model. Italics indicate most parsimonious model.

(A) Category	Metric	Definition	Unit	Trans-formation	Reference
natural	precipitation	average annual precipitation (1970-2000)	mm	sqrt	[1]
natural	temperature	average annual temperature (1970-2000)	°C		[1]
natural	winter ppt	coefficient of variation (CV) for winter precipitation (1970-2000; November – February)			[1]
natural	spring ppt	coefficient of variation (CV) for spring precipitation (1970-2000; March – April)			[1]
natural	summer ppt	coefficient of variation (CV) for summer precipitation (1970-2000; June – August)			[1]
natural	canyon	canyon length	m	sqrt	[2]
natural	watershed	upstream watershed area	km^2	ln+179	[3]
natural	protected lands	proportion of land that is protected		sqrt	[4]
anthropogenic	reservoir	total upstream surface area for reservoirs	km^2	ln+9	[5]
anthropogenic	canal	canal density	m/km^2	ln+0.0049	[5]
anthropogenic	agriculture	upstream agriculture	km^2/km^2	ln+0.0001	[4]
anthropogenic	dam distance	proximity to nearest downstream dam	m	sqrt	[3,5]
anthropogenic	development	proportion of land upstream that is developed		ln+0.001	[4]
anthropogenic	dam density	upstream dam density	$\#/km^2$	ln+0.0002	[5]

(B) MPD	model	AIC	Δ_i	w_i	variables	β	*p*
a) native	full	301.751	5.015	0.075	dam density	0.293	0.018
					reservoir	-0.382	0.040
					summer ppt	0.325	0.010
	natural	307.907	11.17	0.003	watershed	-0.268	0.017
					summer ppt	0.267	0.020
	anthro	*296.736*	*0.000*	*0.921*	*dam distance*	*-0.232*	*0.006*
					dam density	*0.251*	*0.020*
					reservoir	*-0.328*	*0.003*
b) non-native	full	456.340	8.274	0.016	summer ppt	-0.241	0.039
	natural	*448.065*	*0.000*	*0.983*	*summer ppt*	*-0.222*	*0.031*
	anthro	461.030	12.96	0.002	dam distance	0.235	0.005

AIC = Akaike's Information Criterion, Δ_i = deviation from model with lowest AIC, w_i = Akaike weight, anthro = anthropogenic variable subset

References

1. United States Department of Agriculture. 2007 PRISM climate mapping project. United States Department of Agriculture - Natural Resources Conservation Service, Corvallis, USA. http://www.prism.oregonstate.edu/.
2. Whittier J.B., Paukert C.P., Gido K.B. 2006 Development of an aquatic GAP for the Lower Colorado River Basin. Gap Anal. Bull. No. 14 USGS/BRD/Gap Analysis Program, Moscow, Idaho.
3. United States Geological Survey. 2004 The national hydrography dataset. United States Geological Survey. Washington, USA, http://nhd.usgs.gov/.
4. Multi-Resolution Land Characteristics. 2001 National land cover database. Multi-Resolution Land Characteristics Consortium, Washington, USA.
5. United States Army Corps of Engineers. 2007 National inventory of dams. United States Army Corps of Engineers, Washington, USA, http://www.nicar.org/data/dams/.

Electronic Supplementary Material S5. *Comparisons of mean phylogenetic distance geographically and by sub-community.*

Table S5-1. Statistical comparisons of standardized effect size (SES) of mean phylogenetic distance (MPD) across the entire Lower Colorado River Basin, and six historical biogeographic sub-basins using two-tailed *t*-tests. Basin and sub-basins ordered from largest (entire basin) to smallest (Virgin) upstream drainage area. In (a), (b), and (c), negative values indicate phylogenetic clustering, positive values indicate overdispersion. Four null hypotheses were tested: a) all fish MPD = 0; b) native fish MPD = 0; c) non-native fish MPD = 0; d) native MPD = non-native MPD within watersheds.

scale		a) all species	b) native	c) non-native	d) comparison
entire	*t* (df)	-5.71 (158)‡	1.29 (133)	-3.32 (146)‡	-1.54 (121)
basin	avg (SE)	-0.44 (0.08)	0.09 (0.07)	-0.32 (0.10)	
Lower	*t* (df)	-1.85 (16)	1.16 (11)	-2.23 (13)*	-1.52 (8)
Colorado	avg (SE)	-0.54 (0.29)	0.15 (0.13)	-0.82 (0.37)	
Colorado	*t* (df)	1.09 (19)	-4.39 (19)‡	2.42 (19)*	3.82 (19)†
	avg (SE)	0.20 (0.18)	-0.49 (0.11)	0.63 (0.26)	
Lower Gila	*t* (df)	-4.60 (54)‡	3.23 (42)†	-4.29 (51)‡	-3.50 (39)†
	avg (SE)	-0.58 (0.12)	0.35 (0.11)	-0.69 (0.16)	
Upper Gila –	*t* (df)	-2.65 (41)*	0.65 (37)	-1.90 (40)	-1.48 (36)
San Pedro	avg (SE)	-0.32 (0.12)	0.07 (0.13)	-0.23 (0.12)	
Little	*t* (df)	-1.02 (13)	0.07 (9)	-0.61 (13)	-0.01 (9)
Colorado	avg (SE)	-0.26 (0.26)	0.02 (0.36)	-0.20 (0.44)	
Virgin	*t* (df)	-5.13 (10)‡	0.45 (10)	-0.02 (5)	-0.61 (5)
	avg (SE)	-1.50 (0.29)	0.09 (0.20)	-0.01 (0.44)	

* $p < 0.05$, † $p < 0.01$, ‡ $p < 0.001$