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

Angela L. Strecker¹ and Julian D. Olden²

¹Department of Environmental Science and Management, Portland State University, Portland, OR 97207, USA
²School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98105, USA

Despite long-standing interest of 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 of native and non-native fishes of a large river basin in the American Southwest to test for the mechanisms (environmental filtering versus 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 the most phylogenetically divergent from natives across watersheds, but not within watersheds, supporting the hypothesis that Darwin's naturalization conundrum is driven by the 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.

1. Introduction

Understanding patterns of community assembly and the factors that determine the biogeography of species remain central themes in ecology. Although empirical tests and the 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 probably 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 approach to local community phylogenetic analyses by implicitly considering issues of spatial scaling through the incorporation of environmental filters and barriers to dispersal [11]. This combined approach demonstrated that phylogenetic beta diversity for hummingbirds was greater along steep environmental gradients in the Andes Mountains, resulting in phylogenetic clustering in the harsher high-elevation sites, but a tendency to overdispense in less harsh lower elevations [11]. The presence of strong environmental gradients can thereby generate distinct patterns of phylogenetic structure with unique mechanistic explanations.

Although there is mounting evidence of both phylogenetic clustering and overdispersion in plant, animal and bacterial communities from a range of ecozones [6,12], the majority of past studies were conducted on primary producers in terrestrial ecosystems [7], limiting geographical and taxonomic generality. By contrast, freshwater ecosystems, and in particular, freshwater fishes, present a fertile testing ground for community phylogenetic hypotheses, stemming from the unique physiographic and biogeographic constraints imposed by the aquatic landscape [13]. These constraints have led to a vast diversity of fishes in freshwater habitats worldwide. A prime example of this diversification occurred in the arid American Southwest, where fish communities were shaped by a long geological history (e.g. volcanism, isolation and marine intrusions) [14], and harsh environmental conditions, including droughts, floods and extreme temperatures, leading to the evolution of a highly endemic fauna [15,16]. Dam construction, water diversions and flow regulation have significantly altered the environmental conditions in the region, creating conditions that have enabled non-native species that are not adapted to harsh conditions to survive and thrive, displacing native species in many regions [17,18]. The Lower Colorado River Basin has been a flashpoint for the predicament of native species, where the highly endemic ichthyofauna has precipitously declined over the twentieth century [19,20], while over 100 non-native fish species from both neighbouring and distant waters have been introduced (with greater than half established), often to create recreational fishing opportunities in newly developed reservoir habitats [20,21]. Thus, the unique combination of species from diverse geographical locations and broad environmental gradients that range from highly altered to more extreme natural conditions will enhance our scientific understanding of community assembly for freshwater fishes.

In our study, we embrace the highly variable phylogenetic contrast between native and non-native fish species in the Lower Colorado River Basin (draining more than 360,000 km² of the American Southwest) and their accompanying adaptive histories (or lack thereof), to test the following three hypotheses.

(a) Hypothesis 1
Native species in fish assemblages are phylogenetically clustered, reflecting the strong influence of natural environmental conditions in structuring the evolution of these species; non-native species in fish assemblages are overdispersed, reflecting the competitive influences generated by anthropogenic alterations to systems. Non-native fishes in the Southwest often outcompete native fishes under more stable, human-altered flow regimes [15]. Additionally, diet studies suggest that non-natives compete intensely with each other [22], thus it is reasonable to expect that competition is the dominant structuring force in non-native communities. Correspondingly, the phylogenetic structure of native fishes will be highly influenced by environmental drivers representing natural conditions, with functional traits that represent adaptations to these environmental conditions; conversely, phylogenetic structure of non-native fishes will be weakly related to variables representing contemporary human-related conditions (and unrelated to natural conditions), as competition is the primary mechanism 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 (see electronic supplementary material, S1) [23]. By contrast, non-native fishes in the basin come from a much larger array of families (see 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 to establish populations 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].

(b) 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 the 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.

(c) Hypothesis 3
Non-native species that are the most ‘invasive’ (in terms of ecological impacts) will show greater phylogenetic divergence from native species compared with 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 consideration for Darwin’s hypothesis influences observed patterns [27], we contrasted phylogenetic divergence across the entire region as well as within localized watersheds.

2. Material and methods
(a) Data collection
We test the preceding three interconnected hypotheses on a unique large database of fish species occurrences from the Lower Colorado River Basin [31]. The database contains more
than 1.8 million records from museum, university and government collections dating from 1840 to 2009 [24,31,32]. Our study focuses on fish species records collected after 1980 (more than 1.66 million records), as this is considered representative of contemporary assemblages [20,33]. Furthermore, this time frame broadly corresponds to the collection period of contemporary molecular sequence data. Geographical data were reviewed for accuracy, as were regional species lists [31]. Fish were collected using a variety of gears and techniques by different entities, and different studies had different objectives (e.g. population-versus community-level study). Thus, in order to control for these biases, species presence was determined at the local reach scale (i.e. section of river between two confluences), and only records that indicated community-level sampling were retained [24]. Fish species records were then summarized at the aquatic ecological system (AES) scale, which delineates regions by changes in landform, gradient and stream size, and then further divided into 387 AES, which we henceforth refer to as watersheds. Watersheds ranged from 200 to 1600 km$^2$ and are a useful intermediate scale for our analyses. As a result of geographical biases in sample collections, we excluded all watersheds with little or no sampling effort, such that our final dataset comprised $n = 159$ total watersheds. There were $n = 134$ and $n = 147$ watersheds for native and non-native species only, respectively, with $n = 122$ total watersheds for paired native–non-native comparisons (Hypothesis 1: differences in phylogenetic structure of natives and non-natives within same watershed).

(b) Phylogenetic data

Despite the recent explosion of molecular data available to infer phylogenetic relationships among taxa, the diversity of freshwater fishes in North America represents a unique challenge to scientists. This is particularly true of native fishes of the American Southwest, which continue to be taxonomically revised [23], despite the species pool being relatively depauperate. For example, in a large sequence database on freshwater fishes of North America ($n = 685$ species) [34], native species from the Lower Colorado River Basin were largely under-represented, with less than 50% of the species pool present in the database, whereas 88% of the non-native species in our study were represented (A. Streeker 2013, unpublished data). Though studies examining evolutionary history of southwestern endemics have yielded great insight [14,16,35], we are aware of no phylogeny that encompasses all of the fish in this region, which may in part reflect the absence of common molecular markers used across taxa in previous studies. Using sequence divergence data have been recommended for phylogenetic analysis of understudied taxa [36], thus we have chosen the conservative approach of assessing sequence divergence for the mitochondrial cytochrome b, which was the most represented DNA sequence for freshwater fishes in the region (see electronic supplementary material, S1).

We downloaded sequence data from PhyloTA [37], which searches GenBank for similar regions, called phylogenetically informative clusters. Sequence data were obtained for 54 of 66 species (82%); of the 12 species for which there were no sequence data, 11 were native fish species (see electronic supplementary material, S1). Therefore, we used mitochondrial DNA (mtDNA) sequences from a congener (10 species) or the closest relative in the dataset (two species) for unrepresented species. An analysis of the sensitivity of our results to this taxon substitution was performed (see electronic supplementary material, S2). For species that had multiple sequenced individuals, a consensus sequence was constructed [38]. Sequences were aligned [39] and the amount of sequence divergence between all native and non-native species was determined using a Kimura 2-parameter model [40].

(c) Phylogenetic analyses

To test our first hypothesis, we calculated mean phylogenetic distance (MPD) and mean nearest neighbour phylogenetic distance (MNND) using sequence divergence data. MPD is an intra-community or local measure that takes the average distance between all pairs of species present in a watershed, whereas MNND is the average distance between each taxon and its most closely related neighbour [7,41]. As these metrics are biased by species richness, we calculated the standardized effect size (SES) by comparing the observed pattern to a null model using an independent swap algorithm [42], which performs well (i.e. has low type I error rates) for MPD and MNND [43]. The algorithm holds the number of species per watershed constant, as well as the frequency of occurrence of species across samples and randomizes the occurrence matrix [42]. There were 2000 matrix iterations and 5000 runs of the null model for each watershed. A positive SES value indicates that species are overdispersed or evenly distributed throughout the phylogeny, whereas negative SES indicates phylogenetic clustering. Only watersheds with greater than or equal to two species were included, a constraint of the phylogenetic analyses. Analyses were performed jointly, as well as separately on native and non-native subcommunities in watersheds; hereafter, 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 (the 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 analysed (e.g. trophic guilds).

To test our second hypothesis, we calculated phylogenetic beta diversity, which is an intercommunity 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 the electronic supplementary material, S3). As with MPD, analyses were done on native and non-native communities in watersheds with greater than or equal to two 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 ecological impact in their introduced range) or not. Non-native fishes selected by greater than or equal to 75% of experts as invasive were included in the analysis (see electronic supplementary material, S1). These invasive species also have spread at the greatest rate since introduction [20]. Phylogenetic divergence was calculated across the entire region and in each watershed between: (i) all pairs of invasive and native species, (ii) all pairs of remaining non-native (i.e. non-invasive) and native species and (iii) all pairs of native species [30]. At the basin scale, all recorded species were compared. However, in order
to test our hypothesis at the local-watershed scale, we could only include catchments that contained greater than or equal to two species from each category (invasive, non-invasive, native) (n = 85). We used an ANOVA followed by Tukey’s HSD test to distinguish differences between multiple comparisons. These pairwise comparisons are not independent of each other, therefore we used permutation tests (n = 199) to evaluate the significance of phylogenetic divergence across species groups. This analysis was also repeated at the basin scale for non-native species that had failed introductions [47], comparing all pairs of: (i) successfully introduced non-natives and natives, (ii) all pairs of unsuccessfully introduced non-natives and natives and (iii) all pairs of native species.

(d) Statistical analysis
We assessed the influence of environmental and spatial factors on our intra- and intercommunity phylogenetic metrics by compiling data for 14 environmental variables known to be important in structuring fish communities in this region [32]. These variables reflected both natural features (e.g. seasonal precipitation, temperature, watershed area, etc.) and anthropogenic influences (e.g. agriculture, canals, dams, etc.) (see electronic supplementary material, S4).

At the local-watershed scale, we assessed the effects of environmental variation on phylogenetic structure using linear models. Preliminary tests indicated that errors were normally distributed and that there was no significant spatial autocorrelation in the residuals, thus, general linear models were sufficient for our purposes. We used a comparative model selection approach to test our hypothesis that native and non-native community phylogenetic structure (i.e. SES) would be better predicted by natural and anthropogenic descriptors of the environment, respectively. Models of the full set of environmental variables were tested against subsets of natural and anthropogenic environmental variables, and compared with Akaike’s information criterion (AIC), which penalizes models with larger numbers of variables [48]. We measured the phylogenetic signal in functional traits of native species by constructing a phylogeny from mtDNA and estimating Blomberg’s K, which assumes a Brownian motion model of trait evolution [49]. The phylogeny was constructed using maximum likelihood on a Tamura-Nei model [40]. The phylogenetic tree is available in TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S14973). Observed values of K were compared to a null model that was generated by shuffling taxa labels across the phylogeny tips. Lower values of K correspond to random or convergent evolutionary patterns, while higher values indicate increasing trait conservatism.

We used multiple regression on distance matrices (MRM; n = 4999 permutations) to test whether environmental or spatial dissimilarity was related to phylogenetic patterns across watersheds [50]. We used variation partitioning to examine the independent and joint effects of anthropogenic environmental variables, natural environmental variables and space on phylogenetic beta diversity SES. We created separate Euclidean distance matrices for natural and 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 v. 2.12.1 [52]; phylogenetic metrics were calculated using the library picante [53] and MRMs using the library ecodist [54].

3. Results
(a) Hypothesis 1: local phylogenetic structure
On average, there were almost twice as many non-native fish species in watersheds (mean = 8.0 ± 4.1 s.d., range = 2–22) as there were native fish species (mean = 4.5 ± 1.6 s.d., range = 2–10). Pairwise sequence divergence between species ranged from 0.003 to 0.262 (mean = 0.163 ± 0.033 s.d.), with 87% of values falling within the range of values considered optimal for mtDNA to uncover relationships [55]. Sensitivity analyses indicated that MPD results were relatively robust to taxon substitutions, but MNND results were sensitive to taxon substitutions (see electronic supplementary material, S2). This is not a surprising result, given that MNND is evaluating the nearest neighbour 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 with non-native communities; however, native communities were not significantly phylogenetically structured (t133 = 1.29, p = 0.20) compared with the non-native communities, which exhibited significant phylogenetic clustering (i.e. negative MPD; t133 = −3.32, p < 0.01). When all species in a watershed were considered, the entire basin and most sub-basins were significantly phylogenetically structured (t133 = −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 geographical 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 (see 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 with the basins with relatively smaller watersheds (see 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 (see electronic supplementary material, S4). Variability in summer precipitation was significant in models for both native and non-native fishes (full model: βnativate = 0.33, pnativate = 0.01; βnon-native = −0.24, pnon-native = 0.04), as was proximity to the nearest dam (anthropogenic model: βnativate = −0.23, pnativate = 0.01; βnon-native = 0.24, pnon-native = 0.01). Dam density (full model: βnativate = 0.29, pnativate = 0.02), watershed area (natural model: βnativate = −0.268, pnativate = 0.017) and reservoir surface area (full model: βnativate = −0.38, pnon-native = 0.04) were also significant descriptors of native fish community phylogenetic structure. Overall model fit was poor, however, with the most parsimonious model for native and non-native species explaining just 14% and 16% of variation, respectively (p < 0.001 for both models).

The phylogenetic signal of native fishes was significant for the functional traits shape factor (K = 1.10, p < 0.01),
maximum length ($K = 0.96, p < 0.01$) and length at maturation ($K = 1.09, p < 0.01$), indicating moderate trait conservatism. However, phylogenetic signal was not significant for swim factor ($K = 0.21, p = 0.41$) or fecundity ($K = 0.33, p = 0.20$), indicating convergence of traits.

(b) Hypothesis 2: phylogenetic beta diversity
In general, phylogenetic beta diversity was significantly greater for native communities ($\text{mean} = 0.44 \pm 0.01 \text{ s.e.}$) compared with non-native communities ($\text{mean} = 0.08 \pm 0.01 \text{ s.e.}$) (randomization $p < 0.001$). Contrary to our hypothesis, native phylogenetic beta diversity was more strongly correlated with anthropogenic environmental variables ($\beta = 0.24, p < 0.01$) compared with natural environmental variables ($\beta = 0.09, p = 0.02$) using MRM ($R^2 = 0.16, p < 0.01$); however, phylogenetic beta diversity of non-native communities was more correlated with natural ($\beta = 0.25, p < 0.01$) compared with anthropogenic descriptors of the environment ($\beta = 0.14, p < 0.01$) ($R^2 = 0.13, p < 0.01$). Phylogenetic beta diversity in non-native communities was weakly correlated with spatial distance ($\beta = 0.06, p = 0.05$), but phylogenetic beta diversity in native communities was strongly correlated with distance ($\beta = 0.20, p < 0.01$). These results were supported by variation partitioning analyses (figure 2): space had minimal independent effects on non-native communities, but was more influential for native communities. Additionally, there was evidence for spatially structured environmental gradients playing a substantial role in structuring phylogenetic beta diversity for both native and non-native fishes (shared variation between natural environmental variables and space; figure 2).

(c) Hypothesis 3: biotic interactions
To test Darwin’s naturalization hypothesis, we compared pairwise phylogenetic divergence between native, invasive and non-invasive fish species across the entire basin and within each watershed. At the basin scale, both invasive and non-invasive fishes were, on average, significantly more phylogenetically divergent from native species, compared with the amount of divergence between all pairs of native fishes ($F_{2,1223} = 28.85, p < 0.01$; permutation Tukey’s HSD $p = 0.01$) (figure 3a). Additionally, invasive fish species were also significantly more divergent from native taxa compared with non-invasive fishes (permutation Tukey’s HSD $p = 0.02$). However, at the local-watershed scale, patterns were less resolved: in 28% of watersheds, invasive species were significantly divergent from native species, whereas non-invasive species were significantly divergent from native species in 18% of watersheds (figure 3b). Invasive species were significantly divergent from non-invasive species in 1% of watersheds. Non-native species that were successfully introduced were significantly more phylogenetically divergent from native species ($\text{mean} = 0.16 \pm 0.001 \text{ s.e.}; F_{2,764} = 17.84, p < 0.01$; permutation Tukey’s HSD $p = 0.02$) compared with unsuccessfully introduced non-natives at the basin scale ($\text{mean} = 0.15 \pm 0.003 \text{ s.e.};$ permutation Tukey’s HSD $p = 0.03$).

4. Discussion
Phylogenetic structure provides a powerful template for understanding the mechanisms of community assembly and biogeography. The fishes of the American Southwest are a particularly valuable faunal assemblage with which to test general hypotheses about phylogenetic patterns and processes in aquatic environments as a result of the unique geological and evolutionary history of the region. Using a comprehensive fish database for the Lower Colorado River Basin, we were able to test hypotheses about: (i) within watershed patterns and drivers of phylogenetic structure, (ii) between watershed patterns and drivers of phylogenetic structure and (iii) phylogenetic determinants of invasiveness.
Figure 2. Variation partitioning of phylogenetic beta diversity between anthropogenic (anthro) and natural environmental variables, and space in (a) native and (b) non-native fish communities.

We observed differences between native and non-native community phylogenetic structure; however, the pattern did not match our expectation that native communities would be significantly more phylogenetically clustered compared with non-native communities. Rather, non-native assemblages (and entire assemblages) were phylogenetically clustered, whereas native communities showed no significant phylogenetic structure. Our results concur with those for exotic plant communities in California [56]; however, the authors suggested that environmental filters were not controlling the distribution of introduced plants owing to the broad range size of non-native species, combined with low phylogenetic beta diversity. On the contrary, we propose that phylogenetic clustering of non-native fishes is the result of environmental filtering on shared physiological tolerances (i.e. trait conservatism [5]). This conjecture is supported by the strong responses of non-native phylogenetic structure to natural environmental variables compared with native assemblages (see electronic supplementary material, S4).

Additionally, the significantly higher correlation of phylogenetic beta diversity of the non-native fishes with environmental variables compared with spatial distance suggests that the distribution of non-native fishes may be more limited by environmental filters than by dispersal; the latter is likely to be unconstrained owing to human-mediated vectors of introduction. Patterns of significant phylogenetic clustering of 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, for example, variability in stream flow driven by summer precipitation, may influence phylogenetic structure. Indeed, variability in summer precipitation was driven by summer precipitation, may influence phylogenetic in parts of the basin, for example, variability in stream flow upstream contributing area (e.g. Lower Colorado, Lower Gila). 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. Cichlidæ). 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 greater than 30% since the 1950s (range: −14 to 100%; [20]), such that average occupancy for cyprinid species is just 15.4 km² in the basin [58]. Second, environmental conditions were already dramatically changed prior to the contemporary time period (post-1980) used to characterize the fish communities, such that closely related species that may once have been locally adapted are no longer at an advantage.

Despite our hypothesis that native fishes would be more influenced by the environmental variables that they have evolved in response to historically, native phylogenetic structure both within and among watersheds was more strongly related to anthropogenic variables (figure 2; electronic supplementary material, S4). It is striking that the only region where we observed significant phylogenetic clustering in native fishes is the Colorado sub-basin, which contains the Grand Canyon and is therefore one of the most protected (i.e. a national park) and least degraded regions of the entire basin [21], with the notable exception of downstream mainstem impacts from Glen Canyon Dam. For some species, such as Gila cypha and Catostomus discobolus, the Grand Canyon is the last remaining fraction of their historical range in the lower basin [59]. Conversely, the only region
Figure 3. Pairwise phylogenetic divergence (mtDNA) between native species and invasive species (black), non-invasive species (grey), and native species (white) for the entire region (a) and within each watershed (b). See text for distinction of non-native species as invasive versus non-invasive. Boxplots show the 25th, 50th and 75th percentile, whiskers show the 10th and 90th percentile, with circles representing outliers. Tukey's HSD comparison is indicated above the boxes with lowercase letters (p < 0.05) in (a). In (b), percentages represent the number of watersheds out of the total (n = 85) that were significantly different from Tukey’s HSD comparisons.

where native species were significantly overdispersed was in the Lower Gila; this sub-basin has some of the highest levels of anthropogenic threats [21] and invasive species [32] in the basin. This suggests that in this region, human activities may result in non-random extinctions [45] that can shift native communities along a phylogenetic gradient from clustering to overdispersion. We found evidence for trait conservatism in native fishes for some morphological characters (shape factor, maximum length) and life-history traits (length at maturation) but not for other traits (swim factor, fecundity). This suggests that closely related native fishes had similar adaptations for the local environmental conditions. Others have demonstrated the tendency of closely related native species to adopt intermediate life-history strategies (i.e. evolutionary ‘bet-hedging’) [17], which is considered adaptive in highly unpredictable environments. Thus, although native fishes may have been closely related historically, with morphological and life-history adaptations to local conditions, contemporary assemblages no longer reflect this pattern.

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

Darwin’s naturalization conundrum has long been of interest to ecologists; it is only recently that advances in molecular biology have enabled tests of the hypothesis using phylogenetic distances, without the artificial constraints of taxonomy [30]. We found evidence to support Darwin’s naturalization hypothesis at the basin scale, where the most invasive species were more phylogenetically divergent from native species compared with non-invasive species (figure 3). However, at the watershed scale, support for the hypothesis was weaker: invasive and non-invasive fish communities in the majority of watersheds were not phylogenetically divergent from native fishes. These results concur with those from Hypothesis 1, where the mean phylogenetic divergence of native and non-native communities at the watershed scale was largely insignificant (see electronic supplementary material, table S1). 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 the 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 with the species pool. Prior to human intervention, there was only one piscivorous fishes in the Lower Colorado River 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 with non-natives that did successfully establish populations at the basin scale.

A caveat of our study was that native fish species were comparatively under-represented 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 in 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 become 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, for example flooding, but also by human-influenced variables, such as flow regulation by dams and reservoir creation in the basin. By using 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 statement. Funding was provided by the US Geological Survey Status and Trends Program, National Gap Analysis Program and the National Climate Change and Wildlife Science Center.

References