Quantifying Isolation-by-Resistance and Connectivity in Dendritic Ecological Networks

Tyler K. Chafin[1,2], Steven M. Mussmann[1,3], Marlis R. Douglas[1], and Michael E. Douglas[1]

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701

[2]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309

[3]Southwestern Native Aquatic Resources and Recovery Center, U.S. Fish & Wildlife Service, PO Box 219, Dexter, NM 88230

Corresponding author and person to whom reprint requests should be addressed:

Tyler K. Chafin

Department of Ecology and Evolutionary Biology

University of Colorado

Boulder, CO 80309

e-mail: tyler.chafin@colorado.edu

Keywords: Gene flow; isolation-by-distance; stream hierarchy; riverscape genetics; landscape resistance

ABSTRACT(250)

Statistical methods that quantify individual movements have limited applicability within dendritic ecological networks (DENs), to include riverscapes. We promote an approach herein by deriving an analytical framework (=DISTNET and RESISTNET) that more appropriately quantifies differentiation by meshing individual DEN segments with fitted distances and annotated environmental features. We first explored the spatial arrangement of diversity in a river network by fitting pairwise distances against graph edges. We then derived effective resistance models (e.g., as a composite of an array of possible environmental covariates) by using a genetic algorithm for heuristic model selection, invoking circuit theory (CIRCUITSCAPE) to form a graph-based complement to RESISTANCEGA (a landscape ecology package). Although numerous pairwise distance metrics can be employed, we utilized genetic distances encompassing 13,218 ddRAD loci from N=762 Speckled Dace (Cyprinidae: *Rhinichthys osculus*) sampled at 78 Colorado River localities (μ=9.77/site). When distances for each locus were fitted into the network, we found intraspecific divergences due to stream reaches were greater than expected when distances alone. In fitting model-averaged 'effective resistance' models to the same network, we showed the manner by which a host of anthropogenic and environmental variables contribute to patterns of connectivity underlying riverscape genetic differentiation. Our framework represents a contemporary approach to deriving metapopulation/ metacommunity structure within DENs, in that it allows: (a) The extension of projections into unsampled temporal/spatial components; (b) Comparisons to be made among species and/or drainages, both of which benefit multi-species management; and (c) the quantification of locus-specific patterns, from which adaptive hypotheses could then be derived.

24 "*The intimate relation of hydrological conditions to fish life is proverbial; to be hopelessly or*
25 *dangerously maladjusted to one's environment is to be a 'fish out of water'*" (Hubbs 1941).
26
27 **1. INTRODUCTION**(1352)
28 The physical structure of watersheds not only defines the form and function of freshwater fishes
29 (Douglas and Matthews 1992, Jackson et al. 2001), but also their intraspecific evolutionary
30 trajectories and population dynamics (Mussmann et al. 2020b). Despite similar approaches, the
31 resulting patterns of diversification differ markedly from those found in terrestrial landscapes
32 and 'seascapes' (Martinez et al. 2018, Manel et al. 2020, Rabosky 2020). This is because the
33 unique constraints found in riverine networks not only serve to define dispersal pathways but
34 also as a template that dictates genetic relatedness within/ among aquatic biota (Meffe and
35 Vrijenhoek 1988). Network 'connectivity' (=potential for movement) and 'complexity'
36 (=branching rates/ nonlinear topology) thus represent constraints that control numerous
37 population processes, such as: Genetic divergence (Hopken et al. 2013), colonization
38 probabilities (Labonne et al. 2008, Falke et al. 2012), and ephemerality of metapopulation
39 components (Fagan 2002, Campbell Grant 2011).
40 In general, complex dendritic networks may promote intraspecific genetic diversities
41 (Thomaz et al. 2016), and divergences among distal (e.g., headwater) populations (Chiu et al.
42 2020a). Yet, the unidirectional flows in riverscapes differ from other such dendritic ecological
43 networks (DENs), and thus acts to bias the patterns of dispersal found therein (i.e., up- *versus*
44 downstream). This in turn promotes asymmetric gene flow and a source-sink metapopulation
45 structure (Campbell Grant et al. 2007). Consequently, a prevailing trend is for a downstream
46 increase in allelic diversity (Paz-Vinas and Blanchet 2015, Paz-Vinas et al. 2015, Blanchet et al.

2020), whereas unique genetic variation is most often retained within more peripheral headwater populations (Morrissey and De Kerckhove 2009, Chiu et al. 2020b).

Conversely, metapopulations reflect low N_e values (Lowe and Bolger 2002, Altermatt and Fronhofer 2018, Shen et al. 2018) and reduced genetic variability (Thomaz et al. 2016) within unimpeded linear systems that, in turn, impacts both resilience and stability of lineages (Mari et al. 2014, Seymour et al. 2015, Terui et al. 2019), as well as their capacity to proliferate (Harvey et al. 2019). However, numerous labile conditions can also serve to alter these expectations, such as: Life histories [e.g., directionality of dispersal bias (Chiu et al. 2020a)], nonequilibrium population processes (Pilger et al. 2017), geomorphic and climatic regimes (Douglas et al. 1999, Osborne et al. 2014), and abiotic species-level interactions (Paz-Vinas et al. 2015, Rodríguez-González et al. 2019) .

It is thus not surprising that a singular evaluation of stream distances rarely captures intraspecific genetic patterns. For example, in a comparison of 12 salmonid species, the R^2-value for isolation-by-distance (IBD) varied from 0.06–0.87 (μ=0.51; σ=0.26; Bradbury and Bentzen 2007), and from 0.0001–0.44 among three species of darter (Camak and Piller 2018). Although such deviations are partly driven by stochastic or nonequilibrium processes (Hutchison and Templeton 1999), their residuals should also reflect those spatial features that impede individual movements (Keis et al. 2013, Tang et al. 2020). Thus, divergence can alternatively be considered as a function of the movement-cost along stream segments, i.e., 'isolation-by-resistance' (IBR: McRae, 2006; McRae & Beier, 2007).

The IBR model applied within dendritic networks emphasizes specific attributes of individual edges or stream segments, and an array of localized factors have been identified as contributing to resistance (Hughes et al. 2009, Fullerton et al. 2010). These include: Rate of

elevational change (Lowe et al. 2006), flow regime (Rolls et al. 2016), channel width/ catchment area (Ma et al. 2020), water composition (Fourtune et al. 2016), and seasonal hydroclimatic fluctuations (Brauer et al. 2016, 2018). In addition, organisms that never (or but rarely) occupy upstream or headwater conditions can still be directly affected by those habitats (Frissell et al. 1986, Nadeau and Rains 2007, Fullerton et al. 2010).

Most linear streams are also inherently predisposed to fragmentation (Fagan 2002, Fagan et al. 2002, Brauer and Beheregaray 2020), a situation particularly exacerbated in the Anthropocene such that relatively few now remain 'free-flowing' (Grill et al. 2019). Water storage also impacts habitat suitability (Chafin et al. 2019, Fraser et al. 2019), and also shifts population trajectories (Dibble et al. 2020). Dams and diversions either extirpate populations (Brauer and Beheregaray 2020), or disrupt homing behaviors so as to promote homogenization (Baggio et al. 2018), with species of limited mobility particularly impacted (Coleman et al. 2018). Again, vulnerability and the specific nature of impacts vary by context and species (Prunier et al. 2018).

1.1 | *Applying landscape genetic methods to riverscapes*

The mechanics of landscape genetics/genomics often translate poorly to dendritic networks, and this in turn creates a disparity with regard to the number of 'riverscapes' analyzed (Davis et al. 2018, Grummer et al. 2019). At issue is the explicit consideration of riverscape structure upon which genetic relationships can be mapped. Frequent *ad hoc* approaches often employ (a) methods that lack a spatial framework (such as STRUCTURE; Pritchard et al. 2000), or (b) procedures commonly applied in landscape genetics, such as matrix dissimilarities (e.g., Kanno, Vokoun, & Letcher, 2011; Cooke, Landguth, & Beheregaray, 2014; Galbraith, Zanatta,

& Wilson, 2015). The latter, while capable of testing spatio-genetic associations, does so by contrasting a genetic distance matrix against pairwise differences in local conditions (e.g., Mantel test and derivatives; Douglas and Endler, 1982; Smouse, Long, & Sokal, 1986). These are controversial methods, in that the space separating sampling points is generally not evaluated (Legendre and Fortin 2010; Legendre et al. 2015; Peterman and Pope 2020). For riverscapes, this equates to those attributes of stream segments through which migrants must travel (Davis et al. 2018).

On the other hand, path-based methods consider the cumulative costs of movement among points, under the premise that multi- or omni-directional movements are available (an assumption that may not apply within riverscapes). Those environmental features encountered during movements are subsequently converted into ‘resistance values’ (Zeller et al. 2012), with the resulting ‘surface’ allowing for habitat-related metrics to be derived, such as: Least-cost paths, net flux between points, and movement corridors (Shah and McRae 2008, McRae et al. 2016). Operationally, this involves summarizing spatial variables as a raster image of n x m cells (analogous to pixels), with numerical resistance values assigned so as to form an 'effective resistance network' (Davis et al., 2018). Here, riverscapes transition into a graph structure, with nodes (i.e., populations or junctions) connected by edges annotated with environmental attributes (Eros et al. 2011, Peterson et al. 2013).

Few approaches have considered resistance within a spatially explicit graph-theoretic framework, and even fewer have done so by involving multiple covariates in an hypothesis-testing framework [although see White, Hanks, & Wagner (2020)]. Here, it becomes necessary to employ an heuristic approach that explores complex parameter space, particularly given the sensitivity of resistance models to parameterization and the difficulties with the *a priori*

weighting of specific variables within a larger pool of candidates (Spear et al. 2010). Such approaches have been developed for landscapes, using model exploration and selection algorithms (RESISTANCEGA: Peterman, 2018) successful in both empirical and *in silico* scenarios (Peterman et al. 2019, Kimmig et al. 2020, Winiarski et al. 2020).

Here, we employ resistance model perspectives as a mechanism to quantify riverscape genetics/genomics, and do so by exploiting recently developed, large-scale efforts that quantify/classify stream variables (McManamay et al. 2018, Grill et al. 2019, Linke et al. 2019, Troia and McManamay 2020). Our approach modifies those methods that fit distances to network edges with others that subsequently optimize graph-based resistance models (Kalinowski et al. 2008).

We utilized our package to evaluate a case study involving Speckled Dace (SPD; Cyprinidae, *Rhinichthys osculus*), the most widely-distributed species within the most endangered and regulated river system of North America, the Colorado River (Minckley and Deacon 1968, Nilsson et al. 2005, Grill et al. 2019). To accomplish this, we analyzed a large ddRAD dataset as a means to test the following hypotheses: 1) SPD populations are structured according to contemporary riverine connectivity; 2) Their genetic diversity is driven by the topology of the dendritic network; 3) Genetic differentiation is modulated by environmental attributes, represented as edges within the river network; and 4) Spatio-genetic patterns are not explained by distance alone but rather by an isolation-by-resistance model.

2. METHODS (3098)

2.1 | *The automation of population genetic analyses in dendritic networks*

Two complementary software tools, DISTNET and RESISTNET, facilitate high-throughput analyses by employing two complementary software tools that automate the exploration of

environmental correlates as well as the analyses of genetic differentiation in dendritic networks (i.e., DISTNET, RESISTNET; Fig. 1). Both are available as open-source which also includes instructions for virtual deployment as well as installation from scratch (github.com/tkchafin/DENdriscape).

2.1.1 | *The fitting of edge-wise genetic distances*

The DISTNET workflow (Fig. 1) parses geospatial input files (as *.shp*) representing large stream networks in which streams=edges, and endpoints or junctions=nodes. It does so by: 1) 'snapping' sample points to nodes in the river network; 2) Identifying shortest paths between each set of points using Dijkstra's algorithm (Dijkstra 1959), (as implemented in NETWORKX; Hagberg et al. 2008); and 3) Constructing a minimized network representation, preserving original geometries, forming contiguous edges by joining individual stream reaches/ segments, and retaining only those junctions necessary to fulfil shortest Dijkstra paths between points. DISTNET will annotate final results as part of the original geospatial dataset. Given this, we recommend employing the RIVERATLAS dataset (hydrosheds.org/page/hydroatlas) that contains pre-calculated variables representing various aspects of local-scale hydrology (such as physiography, climate, land-cover, geology, and anthropogenic effects) (Linke et al. 2019). This not only simplifies downstream analysis (as herein) but also facilitates visualization of results using graphical geospatial software such as ARCGIS (Esri, Inc.) or QGIS (www.qgis.org).

DISTNET will then calculate pairwise measures of genetic differentiation and stream distances among individuals, sites, or populations. For individual-based metrics, we implemented both uncorrected *p*-distances (i.e., proportion of nucleotide differences over the alignment length), as well as corrected measures [e.g., TN93 (Tamura and Nei 1993)]. Several frequency-

162 based metrics can also be implemented for population- or site-wise calculations, as well as
163 whole-locus single alignments, multi-locus alignments, or SNP datasets. These are: θ_{ST} (Weir
164 and Cockerham 1984); G'_{ST} (Hedrick 2005); G''_{ST} (Meirmans and Hedrick 2011); Linearized
165 F_{ST} (Rousset 1997); Jost's *D* (Jost 2008); Chord distance (Cavalli-Sforza and Edwards 1967), as
166 well as others (github.com/tkchafin/DENdriscape). For more efficient computation when very
167 large numbers of SNPs are employed (as herein), we also recommend that locus-wise distances
168 be pre-calculating using ADEGENET (Jombart 2008) Scripts again available at:
169 github.com/tkchafin/DENdriscape/scripts.

170 Although isolation-by-stream-distance methods are available in DISTNET, the primary
171 approach is the Kalinowski et al. (2008) model in which a least-squares optimization procedure
172 (Felsenstein 2004) is utilized to fit branch lengths. Here, the procedure begins with a matrix of
173 pairwise genetic distances, then seeks to map those stream distances as explained by *k* segments
174 in a network, such that the fitted distance values (*r*) for each segment separating two populations
175 will sum to the observed value from the pairwise matrix. The result is a distance (r_k) for each of *k*
176 segments that represents the genetic distance 'explained' by that segment (Fig. 1; see also
177 Kalinowski et al. 2008).

178 The underlying mathematical implementations include an iterative, constrained
179 optimization approach that corrects for negative-fitted distances (per Kalinowski et al., 2008).
180 However, we provide an additional opportunity for alternative weighting within the least-squares
181 optimization (e.g., Fitch & Margoliash, 1967), that will alter the weight of distances based upon
182 their similarity, which may in turn be beneficial when rates of change vary among branches (or
183 in this case, stream edges; Felsenstein 2004). In addition, the original method (Cavalli-Sforza
184 and Piazza 1975) also remains available for subsequent implementation.

185 Fitted distances can either be computed either for a single locus, as an aggregate (e.g.,
186 arithmetic mean) across loci. The result is a table of fitted distances per stream segment (and per-
187 locus, if so requested), as well as an annotated version of the original geospatial data. Each can
188 be employed for downstream analysis, as demonstrated in our case study. Interestingly, DISTNET
189 can be re-calculated using graphs previously constructed as input (i.e., as a programmatic
190 option).
191
192 2.1.2 | *Composite resistance models optimized within networks*
193 RESISTNET combines a genetic algorithm with multiple environmental variables so as to explore
194 the parameterization of resistance models (Fig. 1), similar to a process found in RESISTANCEGA
195 (Peterman 2018), but with several important differences. For example, genetic algorithms are
196 useful in modelling optimization and selection of features in that they are less impacted by local
197 optima when compared to other 'hill climbing' techniques (Prügel-Bennett 2004). They are
198 employed to simultaneously evaluate a 'population' of models by using log-likelihood or AIC to
199 gauge 'fitness.' Here, 'individuals' (as randomly sampled model configurations) consist of
200 'chromosomes,' with 'genes' (as parameter choices) being represented by: 1) Variable
201 inclusion/exclusion; 2) Variable transformations (using transformation functions available in
202 RESISTANCEGA; Fig. S1); 3) Shape parameters for transformation; or 4) Variable weights (with
203 user-defined ranges).
204 For an optimization problem involving *n*-environmental parameters, an individual
205 chromosome is represented as length=4*n*. Fitness values (e.g., AIC) then determine the survival
206 of individual models and their contribution to successive generations, with stochasticity
207 represented by mutation (random changes to parameter values) and recombination (parameter

choices from multiple individuals combined as offspring). Optimization continues until a threshold number of generations have passed without the fitness of the population increasing (Scrucca, 2013).

The optimization process is summarized as follows (with the DEAP evolutionary algorithm package enabling implementation; Fortin et al. 2012):

1. Network structure and environmental data are imported, using the output formats of DISTNET. Environmental data for contiguous stream reaches (=edges) are summarized and rescaled from 0–10, following Peterman (2018).
2. The initial population is randomly seeded, with a user-defined population size. Individuals are composed chromosomes of length n*4, where n=the number of environmental variables (as above). Transformation, shape, and weight parameters ensure that an optimal scale can be 'found' with respect to a potential non-linear relationship between gene flow and each variable, while also allowing the latter to vary in a composite resistance modelwith regards to relative influence and direction (e.g., positive or negative).
3. Individual fitness is calculated using a custom function, wherein all selected variables are transformed according to the chosen parameters, then re-scaled and combined using a weighted additive function to form a single composite set of resistance values that annotate the original network. Cumulative effective resistance between nodes is then computed using CIRCUITSCAPE. Final calculations employ a maximum likelihood population effects (MLPE) linear mixed effects model to compare pairwise genetic distances with scaled and centered resistance distances, fitted using the MLPE function from RESISTANCEGA (via the RPY2 interface; Gautier 2008). Fitness is then returned via

options available in RESISTANCEGA as either log-likelihood, AIC, or marginal R^2. Fitness can also be computed using ΔAIC_{null}, or the change in AIC values via the MLPE-fitted model, as compared to a 'null' model composed of only random effects.

4. Individuals are 'selected' for successive generations using a tournament selection scheme, a deterministic method whereby random sets of individuals of size *n* specified by the user are compared (=tournaments), with 'winners' being retained.
5. Mutation (=changing parameter values) and recombination (=exchanging parameter combinations among models) occur at generational intervals, with the respective probabilities of both processes as user-specified.
6. Steps 2-5 are repeated for a user-specified number of generations, or until a number of user-specified generations fail to increase the maximum fitness beyond a given threshold.
7. After stopping criteria are met, Akaike weights are computed for each of the 'top' models observed in the population, with models retained so as to meet a cumulative user-specified Akaike weight threshold (default value=0.95). Model-averaged resistance and multi-model, relative variable importance values are subsequently computed using the chosen model set (below).

Multi-model inference has several benefits: It avoids the inherent subjectivity of model selection, in that the estimates from every candidate model are integrated proportional to their approximation of the observed genetic distances (using the Akaike weight; w_i for *i* models), while also more concisely incorporating both uncertainty and information from closely-ranked models. Additionally, it also allows relative variable importance (RVI) to be quantified across models composed of different numbers of explanatory variables, and does so by deriving the sum

254 of Akaike weights that correlate well with *in silico* evaluations of r^2 variable rankings (Giam and
255 Olden 2016).
256
257 2.2 | Rhinichthys osculus *as an empirical case study*
258 As a study species, we focus on the small-bodied SPD, broadly endemic to western North
259 America and represented within all major drainage basins west of the Rocky Mountains. Its
260 distribution serves as a biodiversity bookmark for hydrography and geomorphology of western
261 North America (Hubbs and Miller 1948, Minckley et al. 1986, Oakey et al. 2004, Smith and
262 Dowling 2008). As such, it is remarkable adaptive, extending from high-altitude/ cold-water
263 streams of the Pacific Northwest (Pfrender et al. 2004, Hoekzema and Sidlauskas 2014), to
264 endorheic basins in the Mojave/ Sonoran deserts (Mussmann et al. 2020b). Extreme heat, aridity,
265 and salinity have not only extended its physiological limits but also diversified its phenotypic
266 evolution (Kaya et al. 1992, Lema 2014, Lema et al. 2019).
267 This extraordinary phylogenetic diversity also encompasses highly restricted habitats that
268 are vulnerable to stochastic and anthropogenic influences (Mussmann et al. 2020b).
269 Consequently, and as validated by history, it remains vulnerable to extinction, with two lineages
270 being contemporaneously eliminated in the 20th century, and four subspecies listed under the
271 United States Endangered Species Act of 1973, as amended (16 U.S.C. 1531-1544, 87 Stat. 884).
272 This contemporary decline and extirpation underscores a vulnerability to ongoing climate change
273 (Minckley and Deacon 1968), an hypothesis we test herein by examining several summary
274 indices of anthropogenic modification, in addition to an array of environmental and hydrological
275 predictors (Table 1, and descriptions below).
276

277 2.2.1 | *Sampling and data generation*

278 Fin tissues were collected for *N*=762 individuals distributed throughout 78 Colorado River Basin
279 localities (μ=9.8/site) (Fig. 2; Table S1; Mussmann, 2018). Whole genomic DNA extractions
280 involved several methods (i.e., Gentra Puregene DNA Purification Tissue kit, QIAGEN DNeasy
281 Blood and Tissue kit, QIAGEN QIAamp Fast DNA Tissue Kit, or CsCl-gradient extraction).
282 These were subsequently quantified via Qubit fluorometer (Thermo Fisher Scientific, Inc.) and
283 assessed via gel electrophoresis (2.0% agarose). Sequencing followed the ddRAD protocol
284 (Peterson et al. 2012), with restriction digest of 1µg DNA per sample using *Pst*I and *Msp*I
285 enzymes (New England Biolabs) at 37ºC for 24 hours, followed by bead-purification (Agencourt
286 AMPure XP; Beckman Coulter, Inc.). Samples were standardized (100µg each), ligated with
287 uniquely-barcoded adapters, then pooled into sets of *N*=48/ library. These were size selected
288 (375-425bp; Sage Science Pippin Prep), then extendedn by PCR-based adapter (x12 cycles)
289 using Phusion high-fidelity DNA polymerase (New England Biolabs). Three uniquely-indexed
290 libraries (=144 samples total) were pooled per 1x100 base pair (bp) lane for sequencing on the
291 Illumina HiSeq 2000 (University of Wisconsin Biotechnology Center) or HiSeq 4000
292 (University of Oregon Genomics & Cell Characterization Core Facility).

293

294 2.2.2 | *Data assembly and filtering*

295 Reads were demultiplexed and assembled using the *process_radtags* module (STACKS V2.41;
296 Rochette et al. 2019), followed by elimination of low-quality reads (Phred quality score <20),
297 uncalled bases, and missing restriction cut sites, while allowing one barcode mismatch.
298 Assembly was performed using *de novo* clustering in STACKS, with samples hierarchically
299 grouped into *N*=5 subregions: Upper Colorado River Basin (*N*=221), Virgin River (*N*=223),

300 Grand Canyon and tributaries (*N*=149), Gila River Basin (*N*=105), and Little Colorado River
301 (*N*=64). This ensured recovery of informative loci within each region. Trial runs using STACKS
302 yielded optimal assembly parameters (Paris et al. 2017) for the minimum number of raw reads
303 necessary to form a stack (m=4), as well as the number of mismatches between stacks for
304 identification of putative loci (M=2), and catalog formation (n=2). The *populations* module was
305 utilized to identify SNPs shared by individuals located within a subregion, as outlined above.
306 Loci appearing in <50% of individuals from each subregion were removed, as were loci with an
307 observed heterozygosity >0.6, and a minor allele frequency (MAF) <0.01. These measures
308 removed putative paralogs and minimally informative loci (Eaton 2014; McKinney et al. 2017).
309
310 2.2.3 | *SPD population structure*
311 To provide context for our analyses of riverscape resistance and genetic differentiation, we first
312 gauged population structure by employing an aspatial clustering method (ADMIXTURE;
313 Alexander et al. 2009). Data were filtered to one SNP/ locus, with samples partitioned by region.
314 To do so, we evaluated models containing *K*=1-20 possible sub-populations across 20 replicates,
315 executed in parallel (ADMIXPIPE pipeline; Mussmann et al. 2020a). Cross-validation (CV) values
316 were calculated in ADMIXTURE and summarized across replicates.
317 Replicates were clustered and evaluated in CLUMPAK (Kopelman et al. 2015) so as to
318 identify potentially discordant modes within a single *K* value. Major clusters were identified
319 using a threshold of 90%. Optimal *K* represented the lowest CV value.
320
321 2.2.4 | *Environmental and anthropogenic variables*

322 To test environmental correlates of genetic differentiation, we examined 26 continuous
323 environmental and anthropogenic variables (see Table 1) per stream reach, as parsed from global
324 databases (Grill et al. 2019, Linke et al. 2019). Additional variables were removed from the
325 referenced databases if they exhibited a significant pairwise correlation coefficient (Pearson's *r*)
326 >0.7 (per Holm's multiple test correction; Holm 1979).
327
328 2.2.5 | *Environmental associations and riverscape resistance*
329 We first calculated pairwise linearized F_{ST} (=F_{ST}/1- F_{ST}) for each sampled SNP locus
330 (ADEGENET; Jombart 2008). These were input into DISTNET using the *--locmatdir* and *-r*
331 *RUNLOCI* options. Samples were grouped by locality and, to maximize comparability with
332 STREAMTREE estimates (Kalinowski et al. 2008), the unweighted least-squares approach
333 (Cavalli-Sforza and Edwards 1967, Kalinowski et al. 2008) was employed to fit our stream trees.
334 We used the *--iterative* argument to correct for negative distances. For the input stream matrix,
335 we used a stream layer representing all of North America, as clipped from the RIVERATLAS
336 dataset (Linke et al. 2019). This was subsequently reduced to a minimal contiguous sub-network
337 by an automated procedure performed within DISTNET.
338 Output included a geodatabase and shapefile that mapped contiguous-edge IDs with the
339 original segment identifiers from RIVERATLAS (=HYRIV_ID), as well as fitted F_{ST} values for
340 each locus assigned to each edge, and a mean/ standard deviation across all loci. We performed
341 an automated generalized linear mixed model selection procedure (GLMULTI; Calcagno and de
342 Mazancourt 2010) so as to first examine the relationship between the mean fitted distances and
343 the underlying environmental RIVERATLAS features (Table 1). A genetic algorithm was used to
344 explore all possible sets of explanatory variables, including pairwise interactions among

345 variables. RVI values were computed using the sum of Akaike weights. Prior to analysis, all
346 variables were scaled, centered, and transformed to approximate normality (BESTNORMALIZE;
347 Peterson and Cavanaugh 2019). To consider the effect of variable dispersion, a second GLMULTI
348 run examined the coefficients of variation (c_v) for raw values, calculated per network edge as the
349 ratio of the standard deviation among all contiguous stream reaches composing an edge by the
350 absolute value of the mean.

351 We then performed redundancy analysis (RDA), with the number of variables reduced
352 using a forward selection procedure with 10,000 permutations [ADESPATIAL; (Dray et al. 2018)].
353 The iterative selection procedure was halted using an adjusted R^2 threshold computed from a
354 global RDA using all variables (VEGAN; Oksanen et al 2007). A partial RDA (pRDA)
355 conditioned on edge lengths, was employed to 'partial-out' the IBD effects on F_{STs} fitted edge-
356 wise. Significance was evaluated using the *anova.cca()* function in VEGAN, with outliers
357 identified as stream edges or SNPs having loadings +/- 3 standard deviations from the mean, and
358 labelled according to the predictor variable having the greatest correlation.

359 The above methods examined the correlation of fitted distances *versus* edgewise
360 environmental variables in the stream network. As contrast, we also applied the RESISTNET
361 approach using raw values to compensate for the internal optimization, transformation and
362 scaling of variables. Four replicates were employed, with population sizes equating to either 288
363 or 512. Exploratory analyses were used to tune mutation (*mutpb* and *indpb*) and crossover (*cxpb*)
364 parameters, and to benchmark runtime. The maximum number of generations for full runs was
365 set at 500, with a stop-criterion (=50 generations) engaged if indeed the best model (*nfail*) failed
366 to improve. All transformation types were allowed, and variables weights ranged from 0-1. Input
367 represented the optimized graph and outputs from DISTNET.

368 **3. RESULTS** (1414)

369 3.1 | *Data assembly and population structure*

370 The global assembly represented 13,218 loci at 25.94x mean depth of coverage, with 27.6%
371 missing data after those loci present in <50% of individuals were removed. ADMIXTURE
372 identified *K*=30 populations summed across the five *a priori* defined sub-basins (Fig. 3). The
373 Upper Colorado River Basin was divided into *K*=6 groups composed of major tributaries: Chinle
374 Wash, Green River, Paria River, San Juan River, Smiths Fork River, and Vermillion Creek, with
375 the San Juan River encompassing the Fremont and San Rafael rivers. The Virgin River Basin
376 was partitioned into *K*=7 sub-populations: Pluvial White River, Pahranagat River, Moapa River,
377 Meadow Valley Wash, Beaver Dam Wash headwaters, and Virgin River headwater/ mainstem
378 (Fig. 3). Of note, two localities contain unique SPD subspecies: Pahranagat (*R. o. velifer*), and
379 Moapa (*R. o. moapae*).

380 The remaining *K*=17 sub-populations were distributed throughout the Lower Colorado
381 River Basin, with *K*=5 being distributed linearly through the Grand Canyon (Fig. 3). The
382 remainder were divided among the Bill Williams River (*K*=1), Little Colorado River (*K*=4), and
383 Gila River (*K*=7). The latter was also represented by major tributaries and stream hierarchy:
384 Agua Fria, Eagle Creek, San Francisco River, San Pedro River, San Simon River, Verde River,
385 and Upper Gila tributaries (Fig. 3).

386

387 3.2 | *Stream hierarchy via DISTNET*

388 The minimal graph from DISTNET contained 114 edges representing contiguous combinations of
389 stream reaches, with nodes either as sampling localities or the minimum number of necessary
390 junctions required for shortest Dijkstra paths (Fig. S2). Fitted, linearized F_{ST} for many mainstem

391 reaches of the Colorado and Green rivers (globally over all loci; Fig. 4A) trended very low
392 (F_{ST}=0.0-0.1). Elevated values (F_{ST}>0.2) were more common in the Lower Colorado River
393 Basin. For example, the Agua Fria and Bill Williams rivers both had values >0.35, with higher
394 values also found in multiple smaller tributaries (e.g., Salt River >0.25; Fig. 4A). Several inter-
395 drainage connections also displayed elevated values: Grand Canyon–Virgin River (>0.15);
396 Grand Canyon–Little Colorado (>0.15); Lower Colorado–Gila River (~=0.10). Few in the upper
397 Colorado River Basin tributaries had F_{ST}>0.1, save for the North Fork of Vermillion Creek in
398 Wyoming (>0.25).

399 Linearized F_{ST} for all SNP loci averaged across DISTNET (via *-r RUNLOCI*) generally
400 mirrored the global estimates (Fig. 4B). The connections between Little Colorado and Gila River
401 basins were exceptions, with the Colorado River mainstem being slightly higher relative to the
402 median. The same was observed when examining standard deviation of per-locus fitted-F_{ST} (Fig.
403 4C), with slightly higher relative values in the segment connecting the Virgin-Muddy Basin with
404 the Colorado River in the Grand Canyon.

405 The majority of sampling localities were within 0-2km of nodes in the input stream
406 network, suggesting scant effect due to 'snapping' error or missing stream segments (Fig. 4D).
407 The fit between predicted distances *versus* input F_{ST} was similar to that calculated in the original
408 STREAMTREE program (adjusted R^2=0.77 versus 0.787), again suggesting that slight differences
409 in implementation produced minimal error (Fig. 4E). Stream distances alone were also a
410 relatively poor predictor of F_{ST}, particularly when compared with those predicted using the
411 STREAMTREE model (Mantel r=0.47, p<0.001; Linear model R^2=0.22, p<0.01; Fig. 4F).
412

413 3.3 | *Environmental covariates and riverscape resistance*

414 3.3.1 | *Testing correlates of fitted distances*

415 Generalized linear model (GLM) selection evaluated edgewise variable means (μ) and six
416 variables [via variation coefficients (c_v) via GLMULTI (Fig. 5B, Table 2)] so as to identify eight
417 variables at a model-averaged RVI threshold of 0.8 (Fig. 5A, Table 2). Two variables (i.e., USE
418 and TMP) exceeded >0.8 when both μ and c_v were considered. Analysis of variable μ_s also
419 implicated three anthropogenic (POP, HDI, CSI), two hydrologic (RES, DOF), and one land-
420 cover variable (PAC). Additionally, model-averaged RVI for c_v implicated ARI, SOC, and RIV
421 as well (Fig. 5A, Table 2).

422 Global RDA of stream-wise means were statistically significant (p<0.001, adjusted
423 R^2=0.127). Forward selection evaluated a minimum number of maximally discriminatory
424 variables and chose 10 that met the adjusted R-threshold. Implementing a correction for false-
425 discovery rate (FDR) via the Benjamini-Hochberg procedure resulted in four variables being
426 selected: AET, PET, USE, and DOF (Fig. 5B-E, Table 2). VIFs ranged from 1.19-1.88, and both
427 RDA and partial RDA were significant for the selected variables (p<0.001) (with pRDA
428 conditioned on LENGTH). The pRDA designated 221 SNPs (Figs. B-C) and 9 edges (Figs. D-E)
429 as outliers, with the largest proportion having the strongest direct correlations with evapo-
430 transpiration (AET/ PET; Table S2).

431 Stream-wise variation coefficients from the global RDA were also significant (p=0.007,
432 R^2=0.12), with four variables chosen via forward selection: PET, SOC, SLP, and CSI (Figs. 5F-I,
433 Table 2). Both RDA and pRDA for the selected variables remained significant (p<0.001), with
434 VIFs ranging between 1.04-1.21. Based on pRDA loadings, 349 SNPs (Figs. 5F-G) and eight
435 edges were identified as outliers (Figs. 5H-I), with outlier associations most prevalent for SOC
436 and PET (Table S2).

437

438 3.3.2 | *The optimization of resistance models*

439 We ran four replicate RESISTNET analyses within two different computing environments:
440 Replicates 1 and 2 done virtually on the JETSTREAM cloud (Hancock et al. 2018), while 3 and 4
441 were on Linux-based high-performance compute (HPC) nodes in the TRESTLES cluster at
442 University of Arkansas. We also varied the population size from N=288 (for 1 and 3) to N=512
443 (2 and 4).

444 All replicates running the complete SPD dataset terminated in <24 hours, with an
445 increase in *mutpb* and *indpb* parameters that controlled the probability of individual models and
446 parameters being selected for random mutation. This is necessary step in that rates of mutation
447 (*mutpb* and *indpb*) and crossover (*cxpb*) will vary as a function of parameter space and
448 population size. Here, we recommend these parameters be tuned a priori so as to promote mixing
449 and prevent stationarity.

450 Parallelization was examined during exploratory runs and yielded non-linear
451 improvement with increasing numbers of dedicated processor cores. We saw ~3.5X speed-up
452 going from 1 to 4 processors, whereas increasing from 1 to 16 processors yielded ~10X
453 improvement (Fig. S3). Even so, we found this to provide a substantial performance benefit over
454 dedicating additional threads to parallel processing by CIRCUITSCAPE (Fig. S4), thus we advise
455 complete CPU utilization should be dedicated to distributing genetic algorithm evaluations by
456 RESISTNET (supplied using the *–procs* command-line argument).

457 Full analyses at N=288 or 512 met stopping criteria in ~60-70 generations and
458 demonstrated a fairly rapid plateau in fitness (Fig. S5). The 'best' models in each replicate run
459 varied in fitness, with a moderate predictive capacity as indicated in the MLPE mixed effects

models (marginal R^2=0.296-0.343). Model-averaged effective resistance networks were similar across all replicates, with streams in the Virgin, Gila, and Little Colorado River sub-basins generally having greatest resistance, with the lowest being found in upper Green River tributaries and the mainstem Colorado River above the Grand Canyon (Fig. 6A). Some smaller tributaries in the Upper Colorado River Basin also had high relative resistance values, such as tributaries of the San Juan (i.e., mainstem of Chinle Wash and the Mancos River), and tributaries of the Green River (i.e., Vermillion Creek, and the White River).

Individual replicates were also largely consistent with regard to environmental covariates that defined the composite resistance models, as selected for averaging: ARI, CRP, DOR, PAC, PET, SED, and USE in all four replicates (=100%), whereas SOC and URB were in 75% (Table 2). The remainder (i.e., AET, POP, SGR, SLP) were present in 50% of replicates (Table 2). However, those streams that failed to select catchment slope (SLP) as a covariate, did so instead with stream gradient (SGR), which we interpret as a functional equivalent. Three additional covariates (i.e., DOF, PST, RDD) were selected but once, with generally low model-averaged weights (0.08-0.42).

Model-averaged resistance values were highly correlated among replicates (Pearson's r=0.69-0.96; Fig. 6C), with Replicate 2 having the greatest discordance. Of note, this replicate converged on the highest likelihood model (Fig. S5) and also a replicate that that optimized the larger model population (N=512), potentially highlighting the necessity of a greater population size so as to identify global fitness maxima. Pairwise effective resistances among sites for each replicate were likewise highly consistent (Mantel r=0.94-0.99), with all being highly significant (p<0.0001; Fig. 6C).

483 **4. DISCUSSION** (1669)

484 Spatio-genetic patterns in landscape and riverscape genetics must first be defined before
485 hypotheses can be tested regarding habitat availability, inherent barriers, and potential corridors.
486 Yet, most methods common to riverscape genetics, for example, fail to consider network
487 architecture (Davis et al. 2018), or those attributes that define paths through which individuals
488 must travel. The analytical framework we provide serves to characterize the manner by which
489 genetic diversity and network topology juxtapose, and does so by mapping multi-locus distances
490 onto habitat edges, with causation then explored in a heuristic but rigorous manner. In so doing,
491 we examine an arbitrary number of covariates by employing an explicit graph structure,
492 accomplished without designating variable weighting or *a priori* candidate models.

493 The utility of our approach is underscored by its application in a case study involving
494 SPD in the Colorado River Basin. For some tributaries, highly-fitted F_{ST} (Fig. 3) and high
495 resistance in multiple sub-basins (Fig. 6) indicate the elevated divergence of isolated lineages,
496 even within relatively well-connected riverscapes. Interestingly, segment length was often non-
497 explanatory with regards to basin-wide genetic diversity (e.g., Table 2), implicating a loss of
498 connectivity as more important with regard to intraspecific divergences. This, in turn, juxtaposes
499 well with the prevalence of unique and relictual lineages within isolated watersheds (Oakey et al.
500 2004, Mussmann et al. 2020b).

501 Connectivity was likewise implicated as an important factor for consistent recovery of
502 covariates linked to drought tolerance and seasonal persistence of streams (e.g., AET, ARI,
503 PET). Additionally, the dominant association of these variables with outlier loci (Table S2, Fig.
504 5), suggests the presence of local adaptation with regard to ephemeral flow regimes. Hydrology
505 has similarly been implicated as a driver of morphological diversification in numerous

506 freshwater systems (Langerhans 2008, Meyers and Belk 2014, Bruckerhoff and Magoulick
507 2017), and its importance in sustaining native fish assemblages has been documented across
508 southwestern streams (Propst and Gido 2004, Probst et al. 2008, Kominoski et al. 2018). The
509 large weights attributed to those variables that define stream persistence clearly reflect the
510 impacts of aridity and persistent drought in the region, both of which are omnipresent across
511 contemporary desert riverscapes (Kingsford 2006, Ruhí et al. 2015).

512 Covariates relating to contemporary (e.g., anthropogenic) water manipulation and stream
513 fragmentation were also repeatedly chosen across analyses. Here, factors such as the prevalence
514 of anthropogenic land-use categories [e.g., cropland (CRP); urban extent (UBR)], as well as
515 multiple indices of hydrologic manipulation [e.g., water usage (USE); degree of regulation
516 (DOR); sedimentation rate (SED)], consistently contributed to composite resistance models (Fig.
517 6; Table 2). Several of these are functionally-related, such as: Agricultural/ urban areas as major
518 drivers of water usage (USE); flow regulation/ flood dampening (DOR) as controlling sediment
519 supply (SED) as well as channel width and morphology (Grams et al. 2020, Walker et al. 2020).

520 Anthropogenic factors that promote habitat instability in the Colorado River also
521 contribute significantly to the permeability of species boundaries in endemic cyprinids, with
522 population demographic and genetic trends being impacted (Chafin et al. 2019). These altered
523 conditions may also change the distribution and suitability of low-flow refugia, serving in turn to
524 extirpate peripheral populations (Schlosser 1995, Bezzerides and Bestgen 2002, Magoulick and
525 Kobza 2003). Altered conditions also promote the establishment of non-native species (Gido and
526 Propst 1999, Whitney et al. 2017, Dibble et al. 2020), which may in turn elevate predation rates
527 and promote competition (Marsh and Douglas 1997, Seegert et al. 2014). Predation pressure by

528 resident fishes also constrains the invasion success of introduced SPD (Harvey et al. 2004), thus
529 suggesting a potential role for non-natives in promoting 'biotic' resistance.
530
531 4.1 | *Stream hierarchy and riverscape resistance models*
532 Our case study demonstrates both the utility and limitations of the DISTNET and RESISTNET
533 approaches. Before each can be applied, a necessary understanding is required of how historical
534 factors such as shifting distributions may also impact results. For example, the optimization of a
535 resistance model will be poor when 'true' correlation among genetic and resistance distances is
536 low (Winiarski et al. 2020). Another implicit assumption is that genetic divergences are
537 attributable to the edges within the network (Kalinowski et al. 2008). A potential example of
538 where this might fail could be admixture involving unsampled lineages. For example, Minckley
539 (1973) suggested that intergradation between morphological forms of SPD in the Little Colorado
540 and Gila river drainages was driven by historic stream capture along the Mogollon Rim, with
541 historically-transient connections possible as well between some Upper Colorado River Basin
542 tributaries and those of the Bear or Snake rivers (Smith and Dowling 2008, Loxterman and
543 Keeley 2012). Historic reticulation violates an assumption of our resistance model, in that
544 connectivity does not involve edges external to the network.
545 As with landscape genetic studies, individual movements via extremely rare natural
546 connection or human-mediated translocations will also bias result by altering spatio-genetic
547 associations (Chafin et al., 2021). However, these occurrences could potentially be adjusted for
548 by adding additional edges as historic connections, under the assumption that *a priori* hypotheses
549 can be formulated. Inflated genetic divergences may also emerge as vestiges of historic barriers,
550 such as vicariant Pleistocene lava flows that separate the upstream Little Colorado River from

551 the Grand Canyon (Duffield et al. 2006). If indeed unrecognized, then F_{ST}-values fitted along
552 that stretch would be incorrectly inflated. Historic population dynamics could also lead to a
553 situation where genetic divergences fail to juxtapose with extant conditions, which may explain
554 the high divergence displayed by Vermillion Creek, which may have inadvertently escaped
555 large-scale, indiscriminate rotenone poisoning in the 1950s-60s as a mechanism to exterminate
556 non-native species in the Green River (Holden 1991). Finally, it is also possible that
557 environmental and/ or anthropogenic components of riverscape resistance may vary at smaller
558 spatial scales, especially in those species widely-distributed. Here, a mechanism of test is
559 provided in RESISTNET as an analysis partitioned by sub-basin. The relationship between
560 particular covariates and resistance may often be non-linear, or instead drive resistance as a
561 function of variance (e.g., depth or flow within a reach), rather than by per-edge means (as were
562 examined in resistance models herein). A contrast between variation coefficients and the GLM
563 results of edge means could lends weight to the latter. Non-linear interactions are evaluated in
564 RESISTNET via the simultaneous optimization of variable transformation and shape parameters,
565 yet this remains a limitation in the *post hoc* analysis of DISTNET results (as herein) and may
566 explain why numerous covariates were implicated in the resistance models. Alternate
567 aggregation methods (e.g., variation coefficients) could also be employed with RESISTNET,
568 although they are not currently integrated at this point.
569 These limitations serve as a necessary context from which to develop additional
570 applications, or as avenues for future research, as do emerging developments in riverscape
571 genetics (i.e., this study and White et al., 2020). One extension for locus-wise analysis (as
572 demonstrated herein with RDA; Fig. 5), is the identification of functional genomic associations
573 that can promote isolation-by-adaptation rather than environmental resistance or limitation via

574 dispersal (Orsini et al. 2013). Adaptive genetic variation is a relatively underexplored theme in
575 riverscape genetics (Davis et al. 2018), due largely to a dearth of methods that detect 'outlier'
576 loci within the autocorrelative structure of dendritic networks. Thus, to this end, we hope that a
577 broader exploration of adaptive patterns in riverscapes can be facilitated multi-locus fitting in
578 DISTNET.

579 Another advantage of the RESISTNET is the potential to extrapolate and predict genetic
580 responses in unsampled space. Here, optimized resistance models may be applied to any
581 dendritic ecological network (to include riverscapes, as herein), so as to predict effective
582 resistance in regions within which genetic data has yet to be obtained. This could promote *a*
583 *priori* hypotheses of population structure and phylogeography that serve to predict transecting
584 regions with elevated divergences, or inform analyses requiring assignment priors (Hubisz et al.
585 2009). The resulting contrasts between hypothesized and observed divergences may also target
586 sub-drainages within which discordance has been driven by historical processes.

587 Optimized resistance models might also be used to predict temporal change, with models
588 forged by contemporary data then used to extrapolate metapopulation dynamics under conditions
589 predicted due to climate change or other disturbance (Comte and Grenouillet 2013, Comte and
590 Olden 2017). Importantly, the approach could also be applied to any type of pairwise distance
591 matrix. For example, DISTNET could employ community data to examine edge-wise distances or
592 topological importance in generating differences (Eros et al. 2011), whereas RESISTNET could be
593 used to examine how metacommunity structure is defined by environmental resistance. Again,
594 any form of spatially structured ecological network can be generalized to accommodate either.

595 The potential of DISTNET and RESISTNET for increased automation also serves to
596 encourage large-scale comparative riverscape studies involve variation among species in spatial

or temporal trends, the results of which could vary substantially as a function of trait synergism or life histories (Comte et al. 2014, McManamay and Frimpong 2015, Troia et al. 2019). While rare, such intraspecific studies have implicated ongoing evolution within dendritic networks, while testing specific ecological questions (Fourtune et al. 2016, Pilger et al. 2017, Paz-Vinas et al. 2018, Blanchet et al. 2020). For example: Why do some species exhibit greater intraspecific variation in functional/ morphological attributes than do co-occurring species? (Raffard et al. 2020). How might traits or life histories shape this contrast? How do these patterns relate to long-term macroevolutionary trends (Singhal et al. 2018, Harvey et al. 2019)? Further, a comparative framework also promotes an examination of those commonalities driving community-level structure or shared biogeographic breaks, which may in turn yield trends generalizable to conservation planning at contemporary spatial or taxonomic scales (Linke et al. 2011, Paz-Vinas et al. 2018), and which can effectively promote multi-species management objectives (Nielsen et al. 2017; Douglas et al. 2020).

ACKNOWLEDGEMENTS

We thank Z.D. Zbinden and B.T. Martin for helpful discussions, including invaluable advice with regards to statistical and analytical procedures. We also recognize the Arkansas High Performance Computing Center for computational resources that allowed this package to be tested on the Trestles cluster at University of Arkansas, as funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission. Additional computation support was provided by the NSF XSEDE Jetstream cloud via an allocation awarded to TKC (TG-BIO200074). This work was also indirectly supported by University of Arkansas doctoral fellowships (TKC, SMM, ZDZ), and endowments (MED: 21st Century Global

Change Biology Chair; MRD: Bruker Life Sciences Professorship). TKC is supported by an NSF Postdoctoral Fellowship in Biology under Grant No. DBI: 2010774. The use of trade, product, industry or firm names is for informative purposes only and does not constitute an endorsement by the U.S. Government or the U.S. Fish and Wildlife Service (USFWS). Links to non-Service Web sites do not imply official USFWS endorsement of opinions or ideas expressed therein or guarantee the validity of the information provided. The findings, conclusions, and opinions expressed in this article represent those of the authors, and do not necessarily represent the views of the USFWS, NSF, nor any other funding or affiliated organizations.

Data Accessibility Statement

Raw FASTQ files for each individual have been uploaded to the NCBI Sequence Read Archive (BioProject ID PRJNA656098) and will be released to the public upon publication. All code is available open-source at: github.com/tkchafin/DENdriscape. Curated alignment files and inputs will be made available via the Open Science Framework (DOI: XXXXX) following acceptance.

REFERENCES

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655–1664.

Altermatt, F., and E. A. Fronhofer. 2018. Dispersal in dendritic networks: Ecological consequences on the spatial distribution of population densities. Freshwater Biology 63:22–32.

Baggio, R. A., S. B. L. Araujo, D. Ayllón, and W. A. Boeger. 2018. Dams cause genetic homogenization in populations of fish that present homing behavior: Evidence from a demogenetic individual-based model. Ecological Modelling 384:209-220.

Bezzerides, N., and K. Bestgen. 2002. Status review of Roundtail chub (*Gila robusta*), Flannelmouth sucker (*Catostomus latipinnis*), and Bluehead sucker (*Catostomus discobolus*) in the Colorado River Basin. Final report to the U.S. Bureau of Reclamation.

Blanchet, S., J. G. Prunier, I. Paz-Vinas, K. Saint-Pé, O. Rey, A. Raffard, E. Mathieu-Bégné, G. Loot, L. Fourtune, and V. Dubut. 2020. A river runs through it: The causes, consequences, and management of intraspecific diversity in river networks. Evolutionary Applications 13:1195–1213.

Bradbury, I. R., and P. Bentzen. 2007. Non-linear genetic isolation by distance: Implications for dispersal estimation in anadromous and marine fish populations. Marine Ecology Progress Series 340:245–257.

Brauer, C. J., and L. B. Beheregaray. 2020. Recent and rapid anthropogenic habitat fragmentation increases extinction risk for freshwater biodiversity. Evolutionary Applications 13:2857–2869.

Brauer, C. J., M. P. Hammer, and L. B. Beheregaray. 2016. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Molecular Ecology 61:5093–5113.

Brauer, C. J., P. J. Unmack, S. Smith, L. Bernatchez, and L. B. Beheregaray. 2018. On the roles of landscape heterogeneity and environmental variation in determining population genomic structure in a dendritic system. Molecular Ecology 27:3484–3497.

Bruckerhoff, L. A., and D. D. Magoulick. 2017. Hydrologic regimes as potential drivers of morphologic divergence in fish. Evolutionary Ecology 31:517–531.

Calcagno, V., and C. de Mazancourt. 2010. glmulti: An R package for easy automated model selection with (generalized) linear models. Journal of Statistical Software 34:29.

Camak, D. T., and K. R. Piller. 2018. Going with the flow: Testing the role of habitat isolation among three ecologically divergent darter species. Copeia 106:375–387.

Campbell Grant, E. H. 2011. Structural complexity, movement bias, and metapopulation extinction risk in dendritic ecological networks. Journal of the North American Benthological Society 30:252–258.

Campbell Grant, E. H., W. H. Lowe, and W. F. Fagan. 2007. Living in the branches: Population dynamics and ecological processes in dendritic networks. Ecology Letters 10:165–175.

Cavalli-Sforza, L. L., and A. W. Edwards. 1967. Phylogenetic analysis. Models and estimation procedures. American Journal of Human Genetics 19:550–570.

Cavalli-Sforza, L. L., and A. Piazza. 1975. Analysis of evolution: Evolutionary rates, independence and treeness. Theoretical Population Biology 8:127-165.

Chafin, T. K., Z. D. Zbinden, M. R. Douglas, B. T. Martin, C. M. Middaugh, M. C. Gray, J. R. Ballard, and M. E. Douglas. 2021. Spatial population genetics in heavily managed species: Separating patterns of historical translocation from contemporary gene flow in white-tailed deer. Evolutionary Applications In Press.

Chafin, T. K., M. R. Douglas, B. T. Martin, and M. E. Douglas. 2019. Hybridization drives genetic erosion in sympatric desert fishes of western North America. Heredity 123:759–773.

Chiu, M. C., B. Li, K. Nukazawa, V. H. Resh, T. Carvajal, and K. Watanabe. 2020a. Branching networks can have opposing influences on genetic variation in riverine metapopulations. Diversity and Distributions 26:1813–1824.

Chiu, M. C., K. Nukazawa, T. Carvajal, V. H. Resh, B. Li, and K. Watanabe. 2020b. Simulation modeling reveals the evolutionary role of landscape shape and species dispersal on genetic variation within a metapopulation. Ecography 43:1891–1901.

Coleman, R. A., B. Gauffre, A. Pavlova, L. B. Beheregaray, J. Kearns, J. Lyon, M. Sasaki, R. Leblois, C. Sgro, and P. Sunnucks. 2018. Artificial barriers prevent genetic recovery of small isolated populations of a low-mobility freshwater fish. Heredity 120:515–532.

Comte, L., and G. Grenouillet. 2013. Do stream fish track climate change? Assessing distribution shifts in recent decades. Ecography 36:1236–1246.

Comte, L., J. Murienne, and G. Grenouillet. 2014. Species traits and phylogenetic conservatism of climate-induced range shifts in stream fishes. Nature Communications 5:5053.

Comte, L., and J. D. Olden. 2017. Climatic vulnerability of the world's freshwater and marine fishes. Nature Climate Change 7:718–722.

Cooke, G. M., E. L. Landguth, and L. B. Beheregaray. 2014. Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. Evolution 68:1947–1960.

Davis, C. D., C. W. Epps, R. L. Flitcroft, and M. A. Banks. 2018. Refining and defining riverscape genetics: How rivers influence population genetic structure. Wiley Interdisciplinary Reviews: Water 5:e1269.

Dibble, K. L., C. B. Yackulic, J. C. Schmidt, T. A. Kennedy, and K. R. Bestgen. 2020. Water storage decisions will determine the distribution and persistence of imperiled river fishes. Ecological Applications.

Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1:269-271.

Douglas, M. E., and J. A. Endler. 1982. Quantitative matrix comparisons in ecological and evolutionary investigations. Journal of Theoretical Biology 99:777—795

Douglas, M. E., and W. J. Matthews. 1992. Does morphology predict ecology? Hypothesis testing within a freshwater stream fish assemblage. Oikos 65:213.

Douglas, M. E., W. L. Minckley, and B. D. DeMarais. 1999. Did vicariance mold phenotypes of western North American fishes? Evidence from Gila River cyprinids. Evolution 53:238–246.

Douglas, M. R., W. J. B. Anthonysamy, S. M. Mussmann, M. A. Davis, W. Louis, and M. E. Douglas. 2020. Multi-targeted management of upland game birds at the agroecosystem interface in midwestern North America. PLoS ONE 15(4): e0230735

Dray, S., G. Blanchet, D. Borcard, G. Guenard, T. Jombart, G. Larocque, P. Legendre, N. Madi, and H. H. Wagner. 2018. Package 'adespatial'. R package version, 3-8.

Duffield, W., N. Riggs, D. Kaufman, D. Champion, C. Fenton, S. Forman, W. McIntosh, R. Hereford, J. Plescia, and M. Ort. 2006. Multiple constraints on the age of a Pleistocene lava dam across the Little Colorado River at Grand Falls, Arizona. GSA Bulletin 118:421-429.

Eaton, D. A. 2014. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. Bioinformatics, 30:1844–1849.

Eros, T., D. Schmera, and R. S. Schick. 2011. Network thinking in riverscape conservation - A graph-based approach. Biological Conservation 144:184–192.

Fagan, W. F. 2002. Connectivity, fragmentation, and extinction risk in dendritic metapopulations. Ecology 83:3243–3249.

Fagan, W., P. Unmack, C. Burgess, and W. Minckley. 2002. Rarity, fragmentation, and extinction risk in desert fishes. Ecology 82:3250–3256.

Falke, J. A., L. L. Bailey, K. D. Fausch, and K. R. Bestgen. 2012. Colonization and extinction in dynamic habitats: An occupancy approach for a Great Plains stream fish assemblage. Ecology 93:858–867.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer associates Sunderland, MA.

Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees.

Fourtune, L., I. Paz-Vinas, G. Loot, J. G. Prunier, and S. Blanchet. 2016. Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. Freshwater Biology 61:1830–1845.

Fraser, G. S., K. R. Bestgen, D. L. Winkelman, and K. G. Thompson. 2019. Temperature-not flow-predicts native fish reproduction with implications for climate change. Transactions of the American Fisheries Society 148:509-527.

Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environmental Management 10:199-214.

Fullerton, A. H., K. M. Burnett, E. A. Steel, R. L. Flitcroft, G. R. Pess, B. E. Feist, C. E. Torgersen, D. J. Miller, and B. L. Sanderson. 2010. Hydrological connectivity for riverine fish: Measurement challenges and research opportunities. Freshwater Biology 55:2215–2237.

Galbraith, H. S., D. T. Zanatta, and C. C. Wilson. 2015. Comparative analysis of riverscape genetic structure in rare, threatened and common freshwater mussels. Conservation Genetics 16:845–857.

Gido, K. B., and D. L. Propst. 1999. Habitat use and association of native and nonnative fishes in the San Juan River, New Mexico and Utah. Copeia:321–332.

Grams, P. E., D. J. Dean, A. E. Walker, A. Kasprak, and J. C. Schmidt. 2020. The roles of flood magnitude and duration in controlling channel width and complexity on the Green River in Canyonlands, Utah, USA. Geomorphology 371:107438.

Grill, G., B. Lehner, M. Thieme, B. Geenen, D. Tickner, F. Antonelli, S. Babu, P. Borrelli, L. Cheng, H. Crochetiere, H. Ehalt Macedo, R. Filgueiras, M. Goichot, J. Higgins, Z. Hogan, B. Lip, M. E. McClain, J. Meng, M. Mulligan, C. Nilsson, J. D. Olden, J. J. Opperman, P. Petry, C. Reidy Liermann, L. Sáenz, S. Salinas-Rodríguez, P. Schelle, R. J. P. Schmitt, J. Snider, F. Tan, K. Tockner, P. H. Valdujo, A. van Soesbergen, and C. Zarfl. 2019. Mapping the world's free-flowing rivers. Nature 569:215–221.

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

Hagberg, A. A., D. A. Schult, and P. J. Swart. 2008. Exploring network structure, dynamics, and function using NetworkX. 7th Python in Science Conference (SciPy 2008).

Hancock, D. Y., C. A. Stewart, M. Vaughn, J. Fischer, J. M. Lowe, G. Turner, T. L. Swetnam, T. K. Chafin, E. Afgan, M. E. Pierce, and W. Snapp-Childs. 2018. Jetstream-Early operations performance, adoption, and impacts. Concurrency and Computation: Practice and Experience:e4683.

Harvey, B. C., J. L. White, and R. J. Nakamoto. 2004. An emergent multiple predator effect may enhance biotic resistance in a stream fish assemblage. Ecology 85:127–133.

Harvey, E., M. G. Harvey, S. Singhal, and D. L. Rabosky. 2019. Beyond reproductive isolation : demographic controls on the speciation process 50:75-95.

Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:1633–1638.

Hoekzema, K., and B. L. Sidlauskas. 2014. Molecular phylogenetics and microsatellite analysis reveal cryptic species of speckled dace (Cyprinidae: *Rhinichthys osculus*) in Oregon's Great Basin. Molecular Phylogenetics and Evolution 77:238–250.

Holden, P. B. 1991. Ghosts of the Green River: impacts of Green River poisoning on management of native fishes. Eds: Minckley, W. L., and J. E. Deacon. Battle against extinction. University of Arizona Press, Tuscon, AZ.

Hopken, M. W., M. R. Douglas, and M. E. Douglas. 2013. Stream hierarchy defines riverscape genetics of a North American desert fish. Molecular Ecology 22:956–971.

Hubbs, C. L. 1941. The relation of hydrological conditions to speciation in fishes. A Symposium in Hydrobiology 1941:182–195.

Hubbs, C. L., and R. R. Miller. 1948. The zoological evidence: correlation between fish distribution and hydrogeographic history in the desert basin of western North America. Bulletin of the University of Utah 10:17–166.

Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources

9:1322–1332.

Hughes, J. M., D. J. Schmidt, and D. S. Finn. 2009. Genes in streams: Using DNA to understand the movement of freshwater fauna and their riverine habitat. BioScience 59:573–583.

Hutchison, D. W., and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53:1898–1914.

Jackson, D. A., P. R. Peres-Neto, and J. D. Olden. 2001. What controls who is where in freshwater fish communities - The roles of biotic, abiotic, and spatial factors. Canadian Journal of Fisheries and Aquatic Sciences 58:157–170.

Jombart, T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jost, L. 2008. GST and its relatives do not measure differentiation. Molecular Ecology 17:4015–4026.

Kalinowski, S. T., M. H. Meeuwig, S. R. Narum, and M. L. Taper. 2008. Stream trees : a statistical method for mapping genetic differences between populations of freshwater organisms to the sections of streams that connect them. Canadian journal of fisheries and aquatic sciences 65:2752–2760.

Kanno, Y., J. C. Vokoun, and B. H. Letcher. 2011. Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. Molecular Ecology 20:3711–3729.

Kaya, C. M., P. F. Brussard, D. G. Cameron, and E. R. Vyse. 1992. Biochemical genetics and thermal tolerances of Kendall Warm Springs Dace (*Rhinichthys osculus thermalis*) and Green River Speckled Dace (*R. o. yarrowi*). Copeia 1992:528.

Keis, M., J. Remm, S. Y. W. Ho, J. Davison, E. Tammeleht, I. L. Tumanov, A. P. Saveljev, P. Männil, I. Kojola, A. V. Abramov, T. Margus, and U. Saarma. 2013. Complete mitochondrial genomes and a novel spatial genetic method reveal cryptic phylogeographical structure and migration patterns among brown bears in north-western Eurasia. Journal of Biogeography 40:915–927.

Kimmig, S. E., J. Beninde, M. Brandt, A. Schleimer, S. Kramer-Schadt, H. Hofer, K. Börner, C. Schulze, U. Wittstatt, M. Heddergott, T. Halczok, C. Staubach, and A. C. Frantz. 2020. Beyond the landscape: Resistance modelling infers physical and behavioural gene flow barriers to a mobile carnivore across a metropolitan area. Molecular Ecology 29:466–484.

Kingsford, R. 2006. Ecology of desert rivers. Cambridge University Press, Cambridge, UK.

Kominoski, J. S., A. Ruhí, M. M. Hagler, K. Petersen, J. L. Sabo, T. Sinha, A. Sankarasubramanian, and J. D. Olden. 2018. Patterns and drivers of fish extirpations in rivers of the American Southwest and Southeast. Global Change Biology 24:1175–1185.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources 15:1179–1191.

Labonne, J., V. Ravigné, B. Parisi, and C. Gaucherel. 2008. Linking dendritic network structures

to population demogenetics: The downside of connectivity. Oikos 117:1479–1490.

Langerhans, R. B. 2008. Predictability of phenotypic differentiation across flow regimes in fishes. Integrative and Comparative Biology 48:750–768.

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

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

Lema, S. C. 2014. Hormones and phenotypic plasticity in an ecological context: Linking physiological mechanisms to evolutionary processes. Integrative and Comparative Biology 54:850–863.

Lema, S. C., S. L. Bock, M. M. Malley, and E. A. Elkins. 2019. Warming waters beget smaller fish: Evidence for reduced size and altered morphology in a desert fish following anthropogenic temperature change. Biology Letters 15:20190518.

Linke, S., B. Lehner, C. Ouellet Dallaire, J. Ariwi, G. Grill, M. Anand, P. Beames, V. Burchard-Levine, S. Maxwell, H. Moidu, F. Tan, and M. Thieme. 2019. Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific data 6:283.

Linke, S., E. Turak, and J. Nel. 2011. Freshwater conservation planning: The case for systematic approaches. Freshwater Biology 56:6–20.

Lowe, W. H., and D. T. Bolger. 2002. Local and landscape-scale predictors of salamander abundance in New Hampshire headwater streams. Conservation Biology 16:183–193.

Lowe, W. H., G. E. Likens, M. A. McPeek, and D. C. Buso. 2006. Linking direct and indirect data on dispersal: Isolation by slope in a headwater stream salamander. Ecology 87:334–339.

Loxterman, J. L., and E. R. Keeley. 2012. Watershed boundaries and geographic isolation: Patterns of diversification in cutthroat trout from western North America. BMC Evolutionary Biology 12:38.

Ma, C., Y. Shen, D. Bearup, W. F. Fagan, and J. Liao. 2020. Spatial variation in branch size promotes metapopulation persistence in dendritic river networks. Freshwater Biology 65:426–434.

Magoulick, D. D., and R. M. Kobza. 2003. The role of refugia for fishes during drought: A review and synthesis. Freshwater Biology 48:1186–1198.

Manel, S., P. E. Guerin, D. Mouillot, S. Blanchet, L. Velez, C. Albouy, and L. Pellissier. 2020. Global determinants of freshwater and marine fish genetic diversity. Nature Communications 11:1–9.

Mari, L., R. Casagrandi, E. Bertuzzo, A. Rinaldo, and M. Gatto. 2014. Metapopulation persistence and species spread in river networks. Ecology Letters 17:426–434.

Marsh, P. C., and M. E. Douglas. 1997. Predation by introduced fishes on endangered humpback chub and other native species in the Little Colorado River, Arizona. Transactions of the American Fisheries Society 126:343–346.

Martinez, A. S., J. R. Willoughby, and M. R. Christie. 2018. Genetic diversity in fishes is influenced by habitat type and life-history variation. Ecology and Evolution 8:12022–12031.

McCluney, K. E., N. L. Poff, M. A. Palmer, J. H. Thorp, G. C. Poole, B. S. Williams, M. R. Williams, and J. S. Baron. 2014. Riverine macrosystems ecology: Sensitivity, resistance, and resilience of whole river basins with human alterations. Frontiers in Ecology and the Environment 12:48–58.

McKinney, G. J., R. K. Waples, L. W. Seeb, and J. E. Seeb. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Molecular Ecology Resources, 17:656–669

McManamay, R. A., and E. A. Frimpong. 2015. Hydrologic filtering of fish life history strategies across the United States: Implications for stream flow alteration. Ecological Applications 25:243–263.

McManamay, R. A., M. J. Troia, C. R. DeRolph, A. O. Sheldon, A. R. Barnett, S. C. Kao, and M. G. Anderson. 2018. A stream classification system to explore the physical habitat diversity and anthropogenic impacts in riverscapes of the eastern United States. Page PLoS ONE.

McRae, B. H. 2006. Isolation by resistance. Evolution 60:1551.

McRae, B. H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences of the United States of America 104:19885–19890.

Mcrae, B., V. Shah, and A. Edelman. 2016. Circuitscape: modeling landscape connectivity to promote conservation and human health. The Nature Conservancy.

Meffe, G. K., and R. C. Vrijenhoek. 1988. Conservation genetics in the management of desert fishes. Conservation Biology 2:157–169.

Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: FST and related measures. Molecular Ecology Resources 11:5–18.

Meyers, P. J., and M. C. Belk. 2014. Shape variation in a benthic stream fish across flow regimes. Hydrobiologia 738:147–154.

Miller, R. R., J. D. Williams, and J. E. Williams. 1989. Extinctions of North American fishes during the past century. Fisheries 14:22–38.

Minckley, W. L., D. Hendrickson, and C. Bond 1986. Geography of western North American freshwater fishes: description and relationships to intracontinental tectonism. The Zoogeography of North American Freshwater Fishes:519–613.

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

Minckley, W. L. 1973. Fishes of Arizona. Arizona Game and Fish Department, Phoenix, AZ.

Morrissey, M. B., and D. T. De Kerckhove. 2009. The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. American Naturalist 174:875–889.

Mussmann, S. M., M. R. Douglas, T. K. Chafin, and M. E. Douglas. 2020a. AdmixPipe:

Population analyses in ADMIXTURE for non-model organisms. BMC Bioinformatics 21:337.

Mussmann, S. M., M. R. Douglas, D. D. Oakey, and M. E. Douglas. 2020b. Defining relictual biodiversity: Conservation units in speckled dace (Leuciscidae: *Rhinichthys osculus*) of the Greater Death Valley ecosystem. Ecology and Evolution 10:10798–10817.

Nadeau, T. L., and M. C. Rains. 2007. Hydrological connectivity of headwaters to downstream waters: Introduction to the featured collection. Journal of the American Water Resources Association 43:1–4.

Nielsen, E. S., M. Beger, R. Henriques, K. A. Selkoe, and S. von der Heyden. 2017. Multispecies genetic objectives in spatial conservation planning. Conservation Biology 31:872–882.

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

Oakey, D. D., M. E. Douglas, and M. R. Douglas. 2004. Small fish in a large landscape: Diversification of *Rhinichthys osculus* (Cyprinidae) in Western North America. Copeia 2004:207–221.

Orsini, L., J. Vanoverbeke, I. Swillen, J. Mergeay, and L. De Meester. 2013. Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Molecular Ecology 22:5983–5999.

Osborne, M. J., J. S. Perkin, K. B. Gido, and T. F. Turner. 2014. Comparative riverscape genetics reveals reservoirs of genetic diversity for conservation and restoration of Great Plains fishes. Molecular Ecology 23:5663–5679.

Paris, J. R., J. R. Stevens, and J. M. Catchen. 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution, 8:1360–1373.

Paz-Vinas, I., and S. Blanchet. 2015. Dendritic connectivity shapes spatial patterns of genetic diversity: A simulation-based study. Journal of Evolutionary Biology 28:986–994.

Paz-Vinas, I., G. Loot, V. Hermoso, C. Veyssiere, N. Poulet, G. Grenouillet, and S. Blanchet. 2018. Systematic conservation planning for intraspecific genetic diversity. Proceedings of the Royal Society B: Biological Sciences 285:20172746.

Paz-Vinas, I., G. Loot, V. M. Stevens, and S. Blanchet. 2015. Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. Molecular Ecology 24:4586–4604.

Peterman, W. E. 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. Methods in Ecology and Evolution 9:1638–1647.

Peterman, W. E., and N. S. Pope. 2020. The use and misuse of regression models in landscape genetic analyses. Molecular Ecology:1–11.

Peterman, W. E., K. J. Winiarski, C. E. Moore, C. da S. Carvalho, A. L. Gilbert, and S. F. Spear. 2019. A comparison of popular approaches to optimize landscape resistance surfaces. Landscape Ecology 34:2197–2208.

Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. A. Falke, M. J. Fortin, C. E. Jordan, K. McNyset, P. Monestiez, A. S. Ruesch, A. Sengupta, N. Som, E. A. Steel, D. M. Theobald, C. E.

Torgersen, and S. J. Wenger. 2013. Modelling dendritic ecological networks in space: An integrated network perspective. Ecology Letters 16:707–719.

Pfrender, M. E., J. Hicks, and M. Lynch. 2004. Biogeographic patterns and current distribution of molecular-genetic variation among populations of speckled dace, *Rhinichthys osculus* (Girard). Molecular Phylogenetics and Evolution 30:490–502.

Pilger, T. J., K. B. Gido, D. L. Propst, J. E. Whitney, and T. F. Turner. 2017. River network architecture, genetic effective size and distributional patterns predict differences in genetic structure across species in a dryland stream fish community. Molecular Ecology.

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

Probst, D. L., K. B. Gido, and J. A. Stefferud. 2008. Natural flow regimes, nonnative fishes, and native fish persistence in arid-land river systems. Ecological Applications 18:1236–1252.

Propst, D. L., and K. B. Gido. 2004. Responses of native and nonnative fishes to natural flow regime mimicry in the San Juan River. Transactions of the American Fisheries Society 133:922–931.

Prügel-Bennett, A. 2004. When a genetic algorithm outperforms hill-climbing. Theoretical Computer Science 320:135–153.

Prunier, J. G., V. Dubut, G. Loot, L. Tudesque, and S. Blanchet. 2018. The relative contribution of river network structure and anthropogenic stressors to spatial patterns of genetic diversity in two freshwater fishes: A multiple-stressors approach. Freshwater Biology 63:6–21.

Rabosky, D. L. 2020. Speciation rate and the diversity of fishes in freshwaters and the oceans. Journal of Biogeography 47:1207–1217.

Raffard, A., F. Santoul, S. Blanchet, and J. Cucherousset. 2020. Linking intraspecific variability in trophic and functional niches along an environmental gradient. Freshwater Biology 65:1401–1411.

Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Molecular Ecology 28:4737–4754

Rodríguez-González, P. M., C. García, A. Albuquerque, T. Monteiro-Henriques, C. Faria, J. B. Guimarães, D. Mendonça, F. Simões, M. T. Ferreira, A. Mendes, J. Matos, and M. H. Almeida. 2019. A spatial stream-network approach assists in managing the remnant genetic diversity of riparian forests. Scientific Reports 9:1–10.

Rolls, R. J., J. Heino, and B. C. Chessman. 2016. Unravelling the joint effects of flow regime, climatic variability and dispersal mode on beta diversity of riverine communities. Freshwater Biology 61:1350–1364.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228.

Ruhí, A., E. E. Holmes, J. N. Rinne, and J. L. Sabo. 2015. Anomalous droughts, not invasion, decrease persistence of native fishes in a desert river. Global Change Biology 21:1482–1496.

Schlosser, I. J. 1995. Critical landscape attributes that influence fish population dynamics in headwater streams. Hydrobiologia 303:71–81.

Seegert, S. E. Z., E. J. Rosi-Marshall, C. V. Baxter, T. A. Kennedy, R. O. Hall, and W. F. Cross. 2014. High diet overlap between native small-bodied fishes and nonnative fathead minnow in the Colorado River, Grand Canyon, Arizona. Transactions of the American Fisheries Society 143:1072–1083.

Seymour, M., E. A. Fronhofer, and F. Altermatt. 2015. Dendritic network structure and dispersal affect temporal dynamics of diversity and species persistence. Oikos 124:908–916.

Shah, V. B., and B. H. McRae. 2008. Circuitscape: A tool for landscape ecology. Proceedings of the 7th Python in Science Conference.

Shen, Y., Z. Xu, I. Nijs, and J. Liao. 2018. Spatial arrangement of size-different patches determines population dynamics in linear riverine systems. Ecological Modelling 385:220–225.

Singhal, S., H. Huang, M. R. Grundler, M. R. Marchán-Rivadeneira, I. Holmes, P. O. Title, S. C. Donnellan, and D. L. Rabosky. 2018. Does population structure predict the rate of speciation? A comparative test across Australia's most diverse vertebrate radiation. The American Naturalist 192:432–447.

Smith, G. R., and T. E. Dowling. 2008. Correlating hydrographic events and divergence times of speckled dace (Rhinichthys: Teleostei: Cyprinidae) in the Colorado River drainage. Geological Society of America Special Papers 439:301–317.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology 35:627.

Spear, S. F., N. Balkenhol, M. J. Fortin, B. H. McRae, and K. Scribner. 2010. Use of resistance surfaces for landscape genetic studies: Considerations for parameterization and analysis. Molecular Ecology 19:3576–3591.

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

Tang, Q., T. Fung, and F. E. Rheindt. 2020. ResDisMapper: An R package for fine-scale mapping of resistance to dispersal. Molecular Ecology Resources 20:819–831.

Terui, A., N. Ishiyama, and H. Urabe. 2019. Metapopulation stability in branching river network. Proceedings of the National Academy of Sciences of the United States of America 115:E5963–E5969.

Thomaz, A. T., M. R. Christie, and L. L. Knowles. 2016. The architecture of river networks can drive the evolutionary dynamics of aquatic populations. Evolution 70:731–739.

Troia, M. J., A. L. Kaz, J. C. Niemeyer, and X. Giam. 2019. Species traits and reduced habitat suitability limit efficacy of climate change refugia in streams. Nature Ecology and Evolution 3:1321–1330.

Troia, M. J., and R. A. McManamay. 2020. Biogeographic classification of streams using fish community– and trait–environment relationships. Diversity and Distributions 26:108–125.

Walker, A. E., J. N. Moore, P. E. Grams, D. J. Dean, and J. C. Schmidt. 2020. Channel narrowing by inset floodplain formation of the lower Green River in the Canyonlands region, Utah. GSA Bulletin 132:2333–2352.

Weir, B., and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. Evolution 38:1358–1370.

White, S. L., E. M. Hanks, and T. Wagner. 2020. A novel quantitative framework for riverscape genetics. Ecological Applications 30:1–17.

Whitney, J. E., J. B. Whittier, and C. P. Paukert. 2017. Environmental niche models for riverine desert fishes and their similarity according to phylogeny and functionality. Ecosphere 8:e01658.

Winiarski, K. J., W. E. Peterman, and K. McGarigal. 2020. Evaluation of the R package 'ResistanceGA': A promising approach towards the accurate optimization of landscape resistance surfaces. Molecular Ecology Resources 20:1583–1596.

Zeller, K. A., K. McGarigal, and A. R. Whiteley. 2012. Estimating landscape resistance to movement: A review. Landscape Ecology 27:777–797.

Figures and Tables

Table 1: Description of environmental variables used as covariates in multiple tests of riverscape genetic differentiation in Speckled Dace (*Rhinichthys osculus*) in the Colorado River Basin, where "FFR" refers to the dataset of Grill et al. (2019)

Variable	Description	Source
AET	Actual evapotranspiration (from soil water deficit); equivalent to 'soil water stress coefficient'	RiverATLAS
ARI	Aridity index	RiverATLAS
CRP	Cropland extent as percentage of reach catchment	RiverATLAS
CSI	Connectivity status index	FFR
DOF	Degree/ index of fragmentation due to artificial and/ or natural barriers	FFR
DOR	Degree/ index of anthropogenic flow regulation	FFR
FLD	Floodplain extent, as percentage of reach catchment	FFR
FOR	Forest landcover percentage in reach catchment	RiverATLAS
HFT	Standardized human footprint index	RiverATLAS
HDI	Human development index	RiverATLAS
KAR	Percentage of catchment exposed karst terrain	RiverATLAS
PAC	Percentage of reach catchment in protected areas	RiverATLAS
PET	Potential evapotranspiration predicted from WorldClim	RiverATLAS
POP	Scaled human population size in 2010	RiverATLAS
PST	Pasture extent as percentage of reach catchment	RiverATLAS
RDD	Road crossing density	FFR
RIV	River volume along reach segment	RiverATLAS
REV	Reservoir volume in watershed upstream of reach	RiverATLAS
RUN	Annual average land surface discharge and runoff	RiverATLAS
SGR	Stream gradient; ratio of elevation differential in reach and reach length	RiverATLAS
SLP	Average slope degree in catchment	RiverATLAS
SOC	Soil organic carbon content	HydroATLAS
TMP	Mean annual temperature	RiverATLAS
URB	Percentage catchment area classified as urban	RiverATLAS
USE	Proportion of water removed for human use	FFR

Table 2: Relative importance of variables (=Variable) involved with multiple methods for testing environmental association versus genetic differentiation in Speckled Dace (*Rhinichthys osculus*) from the Colorado River Basin. Values for generalized linear modeling (=GLMULTI) represent the sum of Akaike weights for models fit, using edge-wise means (=RVI_{μ}), calculated from all reaches comprising a given stream segment, and of edge-wise variation coefficients (=RVI_{cv}). Results for partial redundancy analyses (=pRDA) represent R^2 values expressed as percentage (x100), from analyses of edge-wise means (=R^2_{μ}) and variation coefficients (=R^2_{cv}). Model-averaged weights (=w_{RepN}) for each variable are presented as a component of composite resistance models fitted in RESISTNET across four replicates, where Akaike weights are computed using a maximum-likelihood population effects mixed model. Variables not meeting a RVI threshold of 0.8 (GLMULTI and RESISTNET) or not chosen using a forward-selection procedure (pRDA) are represented with a hyphen (-).

	GLMULTI		pRDA		RESISTNET			
Variable	RVI_{μ}	RVI_{cv}	R^2_{μ} (x100)	R^2_{cv} (x100)	w_{Rep1}	w_{Rep2}	w_{Rep3}	w_{Rep4}
AET	-	-	3.9%	-	0.94	-	0.55	-
ARI	-	0.990	2.9%	-	0.38	0.63	0.30	0.56
CRP	-	-	-	-	0.60	0.65	0.72	0.45
CSI	0.992	-	-	2.2%	-	-	0.51	0.49
DOF	0.971	-	1.5%	-	0.32	-	-	-
DOR	-	0.998	1.7%	-	0.53	0.63	0.60	0.59
FOR	-	-	-	-	-	-	-	-
FLD	-	0.999	-	-	-	-	-	-
HDI	0.999	-	-	-	-	-	-	-
HFT	-	-	-	-	0.62	0.58	0.93	0.56
KAR	-	-	-	-	-	-	-	-
LENGTH	-	-	-	-	-	-	-	-
PAC	0.992	-	-	-	0.70	0.25	0.35	0.24
PET	-	-	-	4.3%	0.66	0.55	0.79	0.53
POP	0.993	-	-	-	0.35	-	-	0.15
PST	-	-	-	-	-	-	-	0.08
RDD	-	-	-	-	0.42	-	-	-
REV	0.992	0.930	-	-	-	-	-	-
RIV	-	-	-	-	-	-	-	-
RUN	-	-	-	-	-	-	-	-
SED	-	-	-	-	0.45	0.75	0.77	0.75
SGR	-	-	-	-	-	0.10	0.06	-
SLP	-	-	-	1.8%	0.36	-	-	0.29
SOC	-	0.999	-	3.3%	0.28	0.41	-	0.40
TMP	0.990	0.956	-	-	-	-	-	-
URB	-	-	-	-	-	0.16	0.08	0.24
USE	0.999	0.830	2.2%	-	0.72	0.51	0.43	0.52

Figure 1: Workflows summarized for DISTNET and RESISTNET. DISTNET reduces an input shapefile to a minimal graph, using Dijkstra shortest paths between points, and merging contiguous reaches to form edges. Input genetic distances are 'fitted' to the graph using a least-squares procedure, and results are annotated against the original geometries. RESISTNET optimizes resistance models using a genetic algorithm, where a population of random models is 'evolved' given a fitness function (e.g., AIC), which is computed by comparing the effective resistances computed on the graph for each model (using CIRCUITSCAPE) against either an input pairwise genetic distance matrix, or edge-wise fitted distances from DISTNET . For each generation, models are mutated (=random parameter changes), recombined (=swapping parameter sets), selected (=included/ excluded on the basis of a fitness function), and used to generate offspring for the succeeding generation. Final results are used to compute an effective resistance network and relative variable importance values using model-averaging.

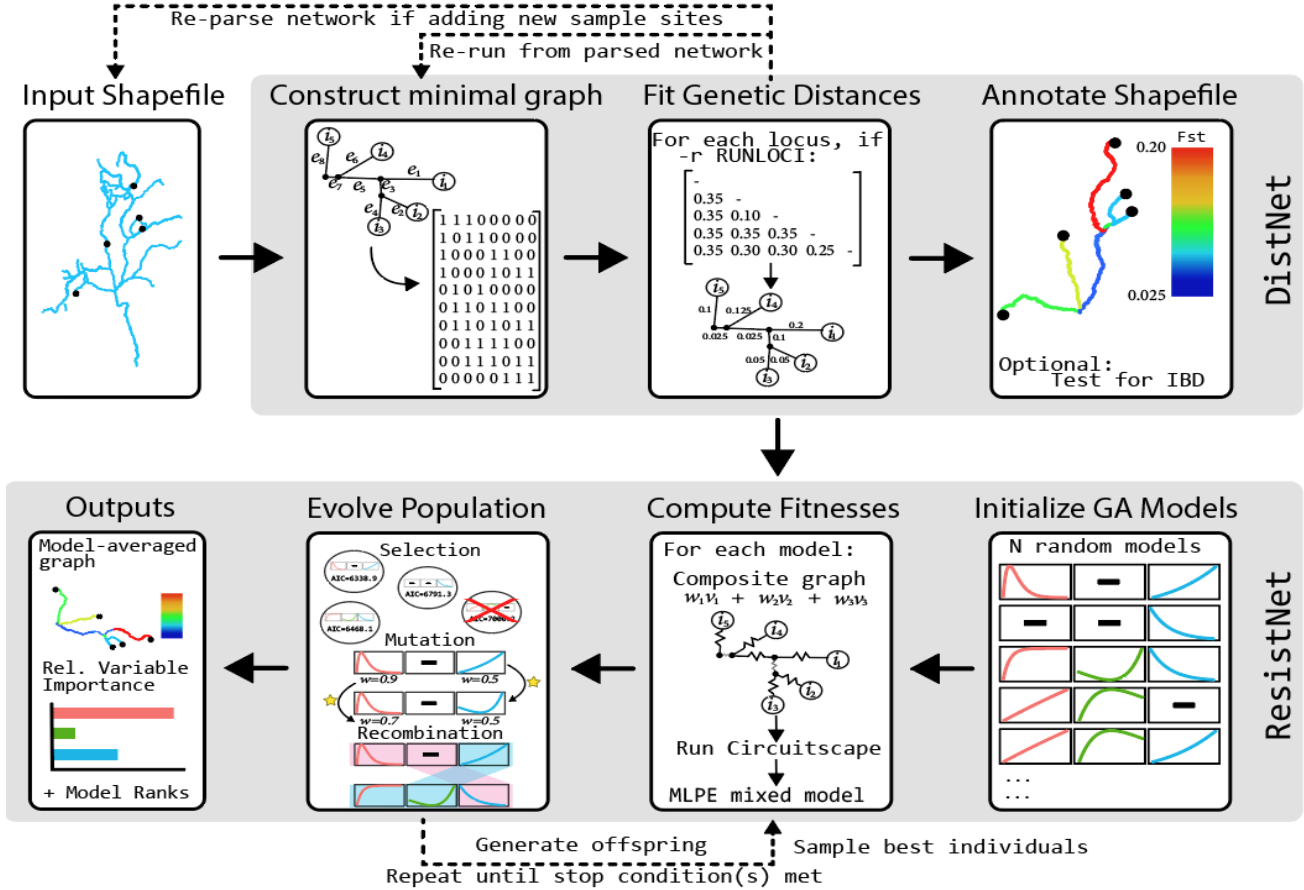


Figure 2: Distribution of *N*=78 Speckled Dace (*Rhinichthys osculus*) sample localities in the Colorado River Basin, as divided into five sub-basins: Virgin River; Grand Canyon; Little Colorado River (=Little CR); Upper Colorado River Basin (=Upper CR); and Gila+Bill Williams rivers (=Gila). Sample sites color-coded per resident sub-basin.

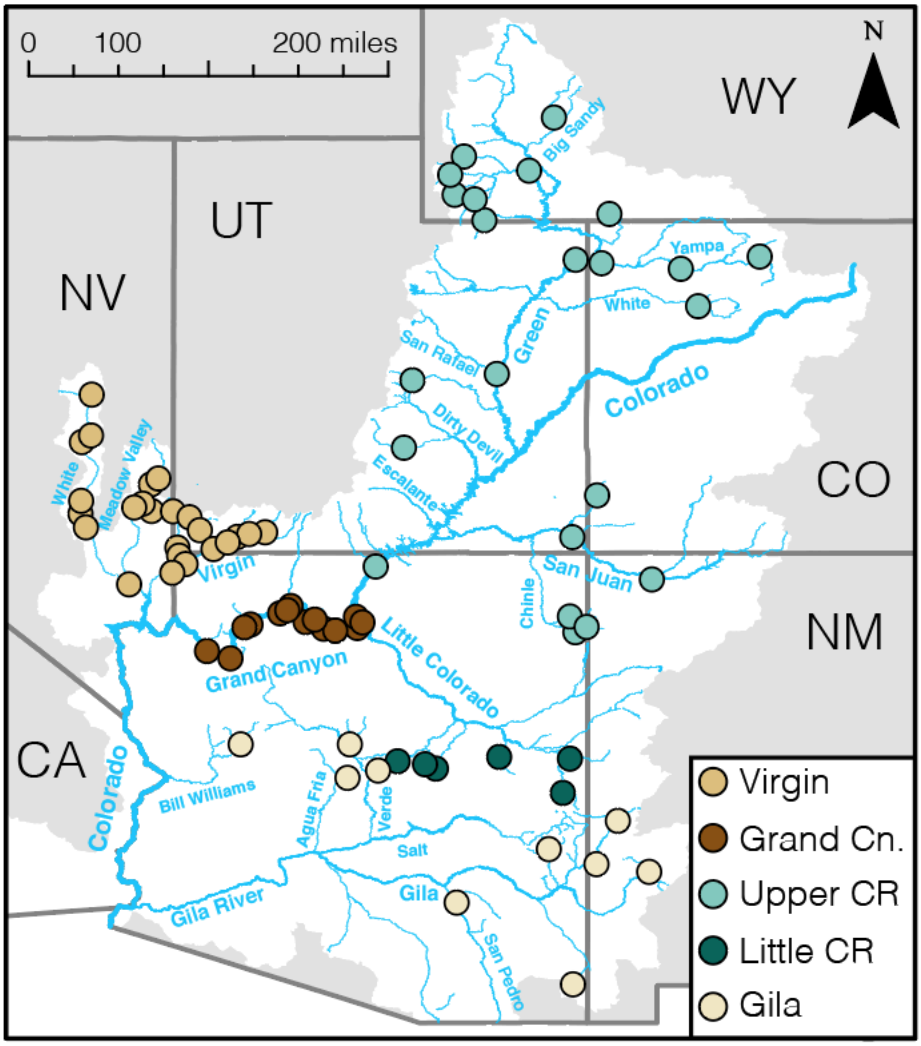


Figure 3: Population structure among Colorado River Basin (CRB) Speckled Dace populations. Upper Colorado/ Green River aggregates (*K*=6) represent: Paria River (PAR), Chinle Wash (CYC-WHY), San Juan+San Rafael+Fremont rivers (DOL-FER), Lower-Mid Green River (DES-BIG), Smiths Fork (SMF), and Vermillion Creek (NVC). The Lower CRB, represented by the Gila+Bill Williams River drainages (*K*=8) include: the Bill Williams River (FRA), Verde River (FOS-SPG), Agua Fria (SYC), San Pedro (ARU), San Simon (NCC), Eagle Creek (EAG), San Francisco River (GLN-TUL), and Upper Gila River (WFG). The Little Colorado River (*K*=4) encompasses East Clear Creek (WIL-ECC), Chevelon Creek (CHE), Silver Creek (SIC), and the East Fork (WEN-EFL). Grand Canyon (*K*=5): distributed linearly along the length of the canyon, with the exception of Havasu Creek (HAV). Virgin River (*K*=7): upper Virgin River (NTH-NFV), Mainstem Virgin River (LAV-WFA; BDW-VSC), Beaver Dam Wash (BDM-CCN), Meadow Valley Wash (BBL-MVW), Moapa River (MOA), Pahranagat (PAH-VEL), and White River (ISP-CAM).

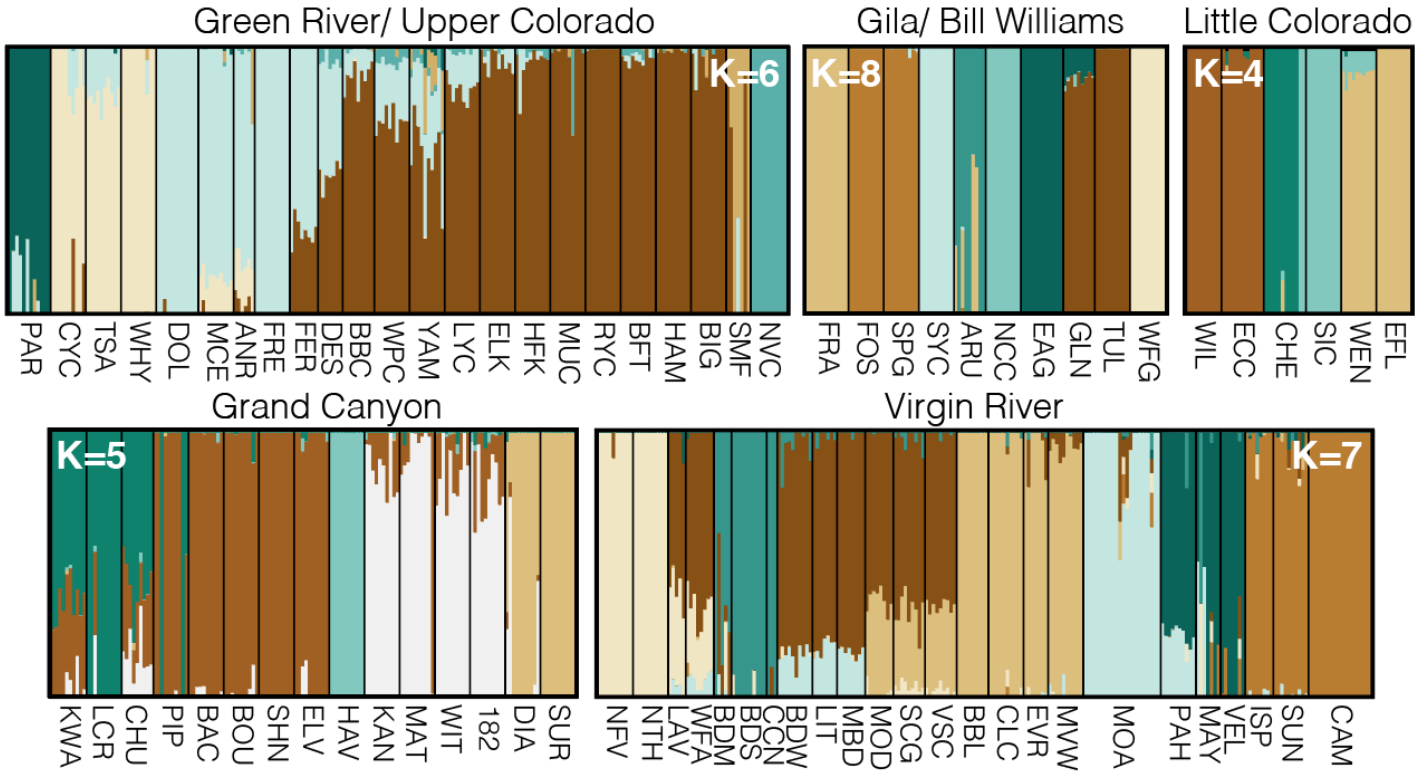


Figure 4: Composite plot of DISTNET results for N=78 Speckled Dace (*Rhinichthys osculus*) sampling localities distributed throughout the Colorado River Basin. (A) Linearized F_{ST} (F_{ST}/1-F_{ST}) computed globally across loci, fitted to stream edges; (B) Mean fitted F_{ST} estimates computed across 13,218 SNP loci; (C) Standard deviation (Std. Dev.) of per-locus fitted F_{ST} estimates; (D) Histogram of 'snap' distances between true sampling localities and nearest river network node; (E) Globally computed, linearized F_{ST} *versus* those predicted from DISTNET - fitted values (linear regression: adjusted R^2=0.77, p<1.0e^{-16}); (F) Pairwise stream distances among sites versus pairwise linearized F_{ST} (Mantel r=0.47, p<0.001).

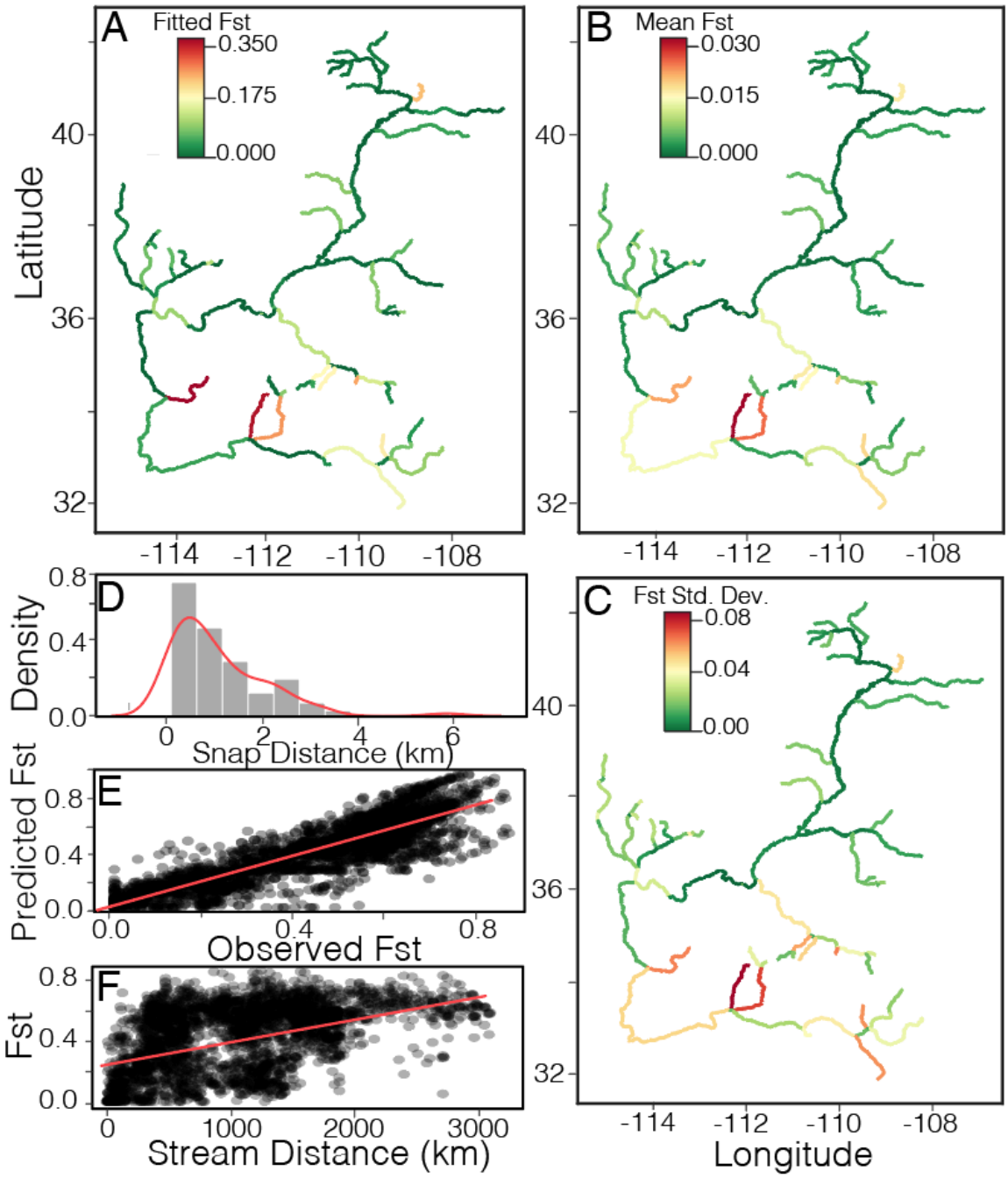


Figure 5: Modelling association of 27 environmental values and stream length with fitted-F_{ST} values for 113 contiguous stream segments (=network edges) in the Colorado River Basin, as inferred using N=78 sampling localities of Speckled Dace (*Rhinichthys osculus*) and 13,218 SNP loci and least-squares optimization in DISTNET. (A) model-averaged relative variable importance (RVI) for generalized linear models (GLMs) fit using mean F_{ST} across loci per edge (RVI_{μ}) and using variation coefficients (RVI_{cv}), with a RVI threshold of 0.8 (red line); (B-C) Partial redundancy analyses (pRDA controlled for stream length) of locus-wise F_{ST} means across edges, scaled by loci; (D-E) pRDA of locus-wise F_{ST} means, scaled by edge; and pRDA of locus-wise F_{ST} variation coefficients, scaled by loci (F-G), and by edge (H-I). pRDA variables chosen via forward selection. Variable abbreviations (e.g., AET) per Table 1.

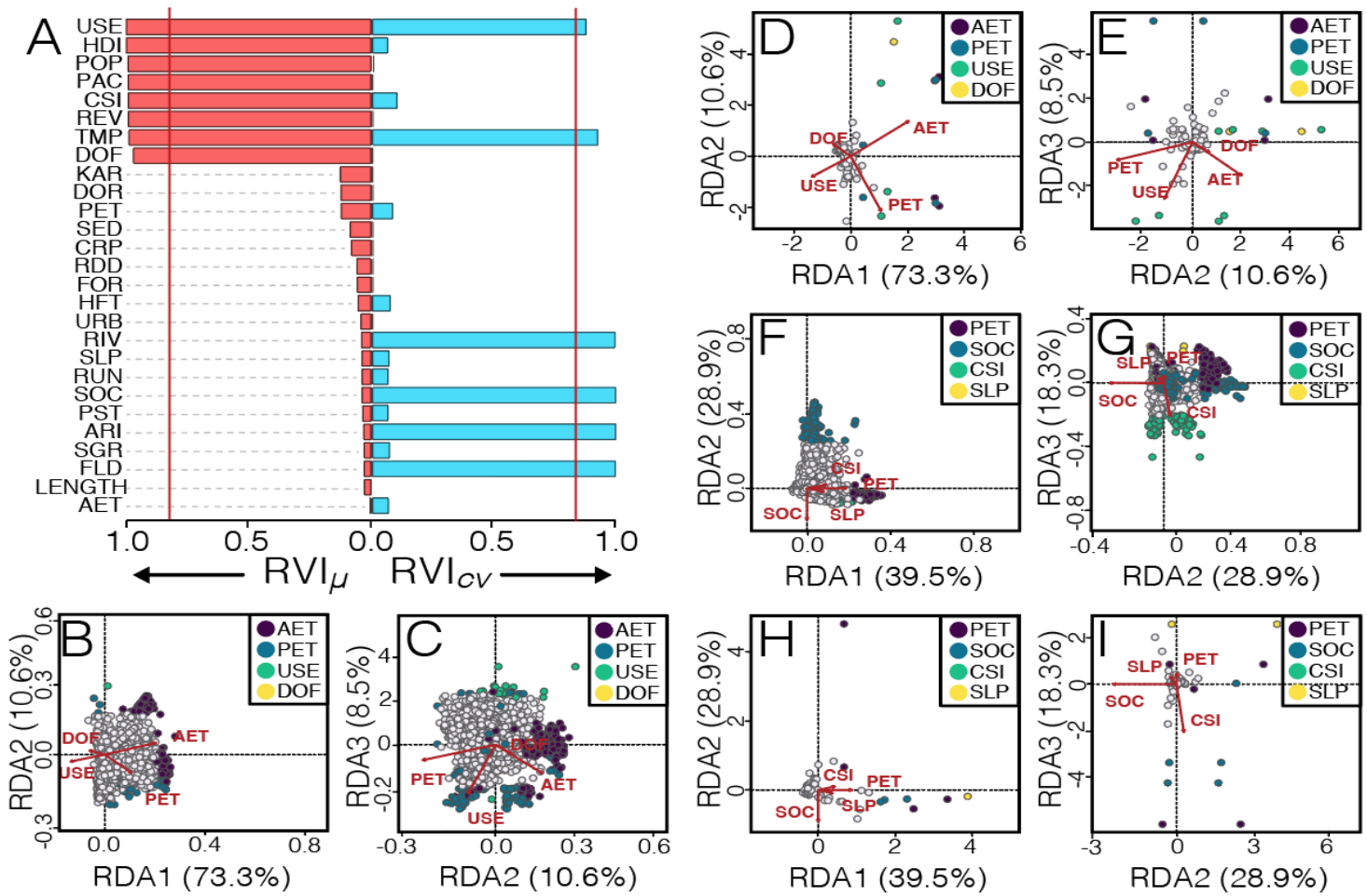
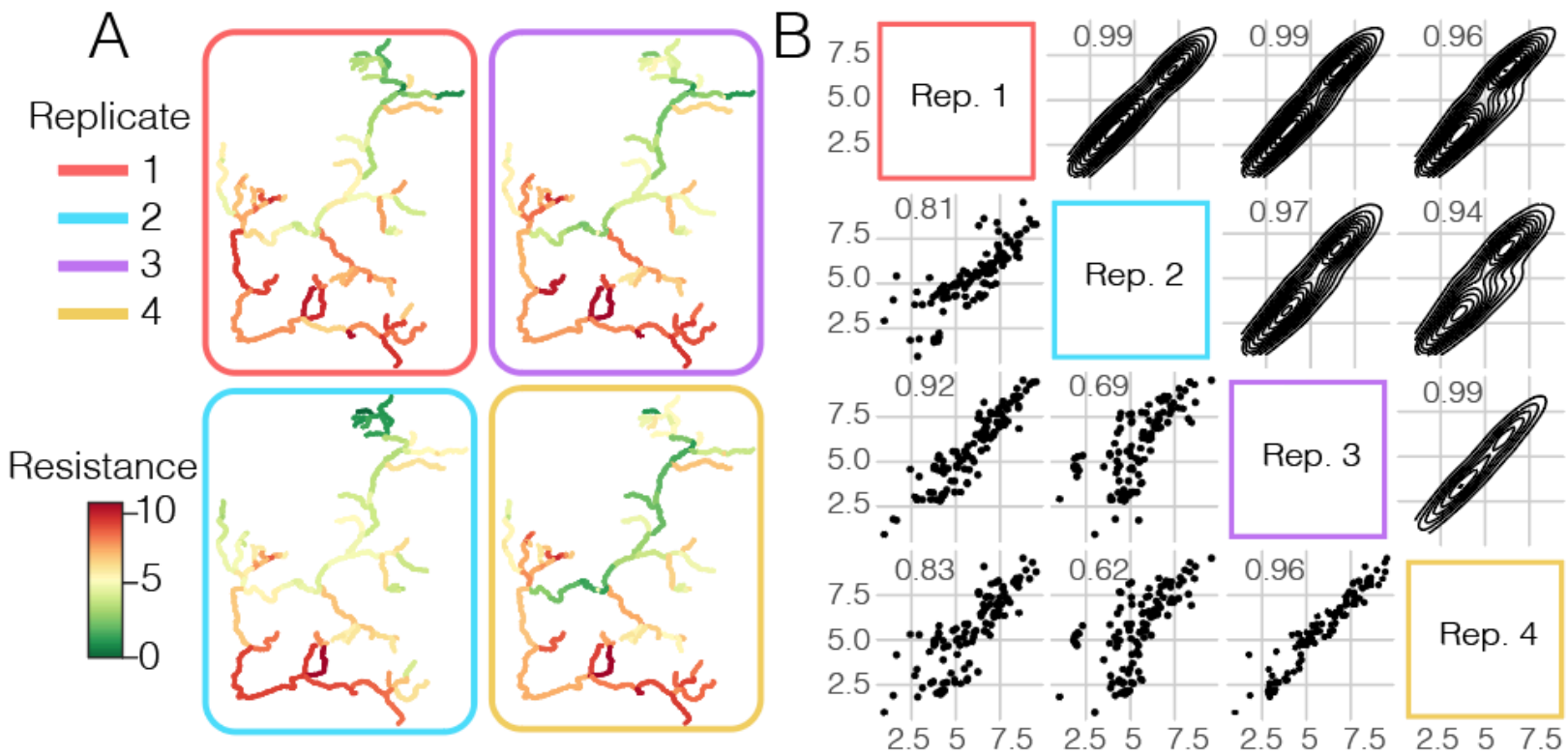


Figure 6: Composite plot of RESISTNET results across four independent replicates modelling riverscape resistance using 27 environmental variables in a river network of *N*=78 Speckled Dace (*Rhinichthys osculus*) sampling sites in the Colorado River Basin. Each replicate explored model space using a genetic algorithm with a model population size 512, with the final *N* top models selected as those with a cumulative Akaike weight of 0.95. Resistance values for each stream segment (A) represent model-averaged estimates (using Akaike weights). Replicates compared using Pearson's correlation coefficient (r) of edge-wise resistances (B; below diagonal) and Mantel test of pairwise resistances among sites (B; above diagonal).



1122 **Supplementary Material**

1123 **Table S1**: Locality information (=Locality) and sample sizes (=N) for *N*=78 sampling localities
1124 of Speckled Dace (*Rhinichthys osculus*) in sub-basins (=Sub-Basin) of the Colorado River Basin.
1125

Site	Locality	Sub-basin	*N*
182	Grand Canyon RM182	Grand Canyon	10
ANR	Animas River	Upper CRB	6
ARU	Aravaipa Creek; San Pedro tributary	Lower CRB	12
BAC	Grand Canyon RM87-117	Grand Canyon	10
BBC	Big Beaver Creek – White River	Upper CRB	9
BBL	Meadow Valley Wash	Virgin	9
BDM	Beaver Dam Wash	Virgin	6
BDS	Beaver Dam Wash	Virgin	10
BDW	Mainstem Virgin River	Virgin	10
BFT	Blacks Fork	Upper CRB	10
BIG	Big Sandy River	Upper CRB	10
BOU	Grand Canyon RM87-117	Grand Canyon	10
CAM	Camp Spring – White River	Virgin	18
CCN	Beaver Dam Wash	Virgin	4
CHE	Chevelon Creek; Little Colorado trib.	Little Colorado	12
CHU	Grand Canyon RM225-249	Grand Canyon	9
CLC	Meadow Valley Wash	Virgin	10
CYC	Chinle Wash; Colorado R. tributary	Upper CRB	10
DES	Desolation Canyon; Green R. mainstem	Upper CRB	7
DIA	Grand Canyon RM225-249	Grand Canyon	10
DOL	Dolores River	Upper CRB	12
EAG	Eagle Creek	Lower CRB	12
ECC	East Clear Creek; Little Colorado trib	Little Colorado	12
EFL	East Fork, Little Colorado River	Little Colorado	10
ELK	Elk River (Yampa River Tributary)	Upper CRB	10
ELV	Grand Canyon RM87-117	Grand Canyon	10
EVR	Meadow Valley Wash	Virgin	7
FER	Ferron Creek – San Rafael River	Upper CRB	8
FOS	Fossil Creek; Verde River tributary	Lower CRB	10
FRA	Francis Creek; Bill Williams R. tributary	Lower CRB	12
FRE	Fremont River; Green River tributary	Upper CRB	10
GLN	San Francisco River at Glenwood	Lower CRB	9
HAM	Ham's Fork; Green River	Upper CRB	10
HAV	Havasu Creek, Grand Canyon	Grand Canyon	10
HFK	Ham's Fork; Green River	Upper CRB	10
ISP	Indian Springs; White River; Virgin-Muddy tributary	Virgin	8
KAN	Kanab Creek; Grand Canyon	Grand Canyon	10
KWA	Grand Canyon RM225-249	Grand Canyon	10
LAV	Mainstem Virgin River	Virgin	5

LCR	Little Colorado River confluence	Grand Canyon	10
LIT	Mainstem Virgin River	Virgin	7
LYC	Little Yampa Canyon; Yampa River	Upper CRB	10
MAT	Grand Canyon RM143-189	Grand Canyon	10
MAY	Maynard Spring – Pahranagat Valley	Virgin	7
MBD	Mainstem Virgin River	Virgin	8
MCE	McElmo Creek – San Juan River	Upper CRB	10
MOA	Moapa River; Virgin-Muddy tributary	Virgin	22
MOD	Mainstem Virgin River	Virgin	8
MVW	Meadow Valley Wash	Virgin	10
MUC	Muddy Creek – Blacks Fork River	Upper CRB	10
NCC	San Simon	Lower CRB	10
NFV	Upper Virgin River	Virgin	10
NTH	Upper Virgin River	Virgin	10
NVC	Vermillion Creek	Upper CRB	10
PAH	Pahranagat River; Virgin-Muddy basin	Virgin	10
PAR	Paria River	Upper CRB	12
PIP	Grand Canyon RM87-117	Grand Canyon	10
RYC	Ryckman Creek upstream of Hill Creek Confluence	Upper CRB	10
SCG	Mainstem Virgin River	Virgin	9
SIC	Silver Creek; Little Colorado tributary	Little Colorado	10
SHN	Shinumo Creek; Colorado R. tributary	Grand Canyon	10
SMF	Smith's Fork; Green River	Upper CRB	7
SPG	Spring Creek; Verde River tributary	Lower CRB	10
SUN	Sunnyside Creek; White River; Virgin-Muddy	Virgin	10
SUR	Grand Canyon RM225-249	Grand Canyon	10
SYC	Sycamore Creek; Agua Fria tributary	Lower CRB	10
TSA	Chinle Wash; Colorado R. tributary	Upper CRB	10
TUL	Tularosa Creek – San Francisco River	Lower CRB	10
VEL	Burn's Ranch – Pahranagat Valley	Virgin	7
VSC	Mainstem Virgin River	Virgin	10
WEN	East Fork, Little Colorado R.	Little Colorado	10
WFA	Mainstem Virgin River	Virgin	8
WFG	West Fork of the Gila River at cliff dwellings	Lower CRB	10
WIL	East Clear Creek; Little Colorado trib	Little Colorado	10
WIT	Grand Canyon RM143-189	Grand Canyon	10
WPC	Island Park below Whirlpool Canyon; Yampa River	Upper CRB	10
WHY	Chinle Wash; Colorado R. tributary	Upper CRB	10
YAM	Yampa River (Green R. tributary)	Upper CRB	10

1127 **Table S2**: Outlier SNPs/ edges in pRDA analysis of means (μ) and variation coefficients (c_v).
1128
1129

Variable	*N* SNPs (%)	*N* Edges (%)
c_v – CSI	105 (30.1)	-
c_v – PET	112 (32.1)	3 (42.9)
c_v – SLP	2 (0.6)	1 (14.3)
c_v – SOC	130 (37.2)	3 (42.9)
μ – AET	158 (71.5)	2 (22.2)
μ – DOF	-	1 (11.1)
μ – PET	11 (5.0)	4 (44.4)
μ – USE	52 (23.5)	2 (22.2)

1130

Table S3: Fitness metrics for the 'best' model selected in each of four independent RESISTNET replicate runs. Values represent: Log-likelihood (=*loglik*), *AIC* values, marginal R^2 of MLPE mixed model fit (=R^2_m), and change in likelihood as compared to the null model (=ΔAIC_{null}).

Replicate	*loglik*	*AIC*	R^2_m	ΔAIC_{null}
1	-3087.17	6182.35	0.337	1151.52
2	-2991.06	5990.13	0.343	1201.49
3	-3086.74	6181.49	0.296	1006.52
4	-3053.56	6115.12	0.332	1141.73

Figure S1: Shapes of transformations used in RESISTNET at several shape parameter values (1,3 ,5, 10, 20, 50, 100). Transformed values maintain the scale of the input data (with a maximum of 10.0 and minimum of 0.0), and approach linearity as the shape parameter increases.

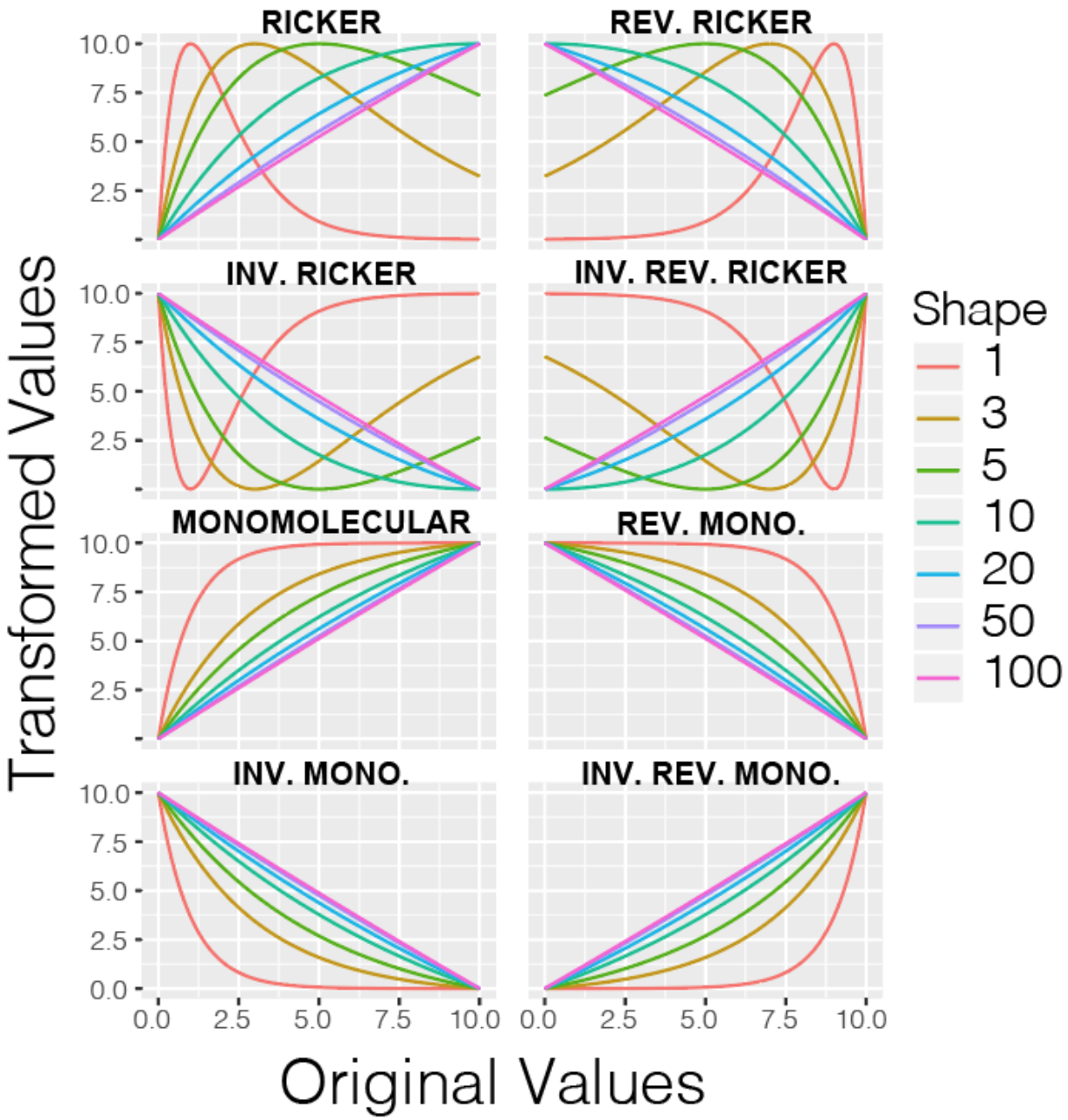


Figure S2: Graph representation of the minimized stream network for Speckled Dace (*Rhinichthys osculus*) sampling sites (N=78) in the Colorado River Basin, with sample site (blue) and junction nodes (black) represented as circles, and stream segments (edges) as connecting lines.

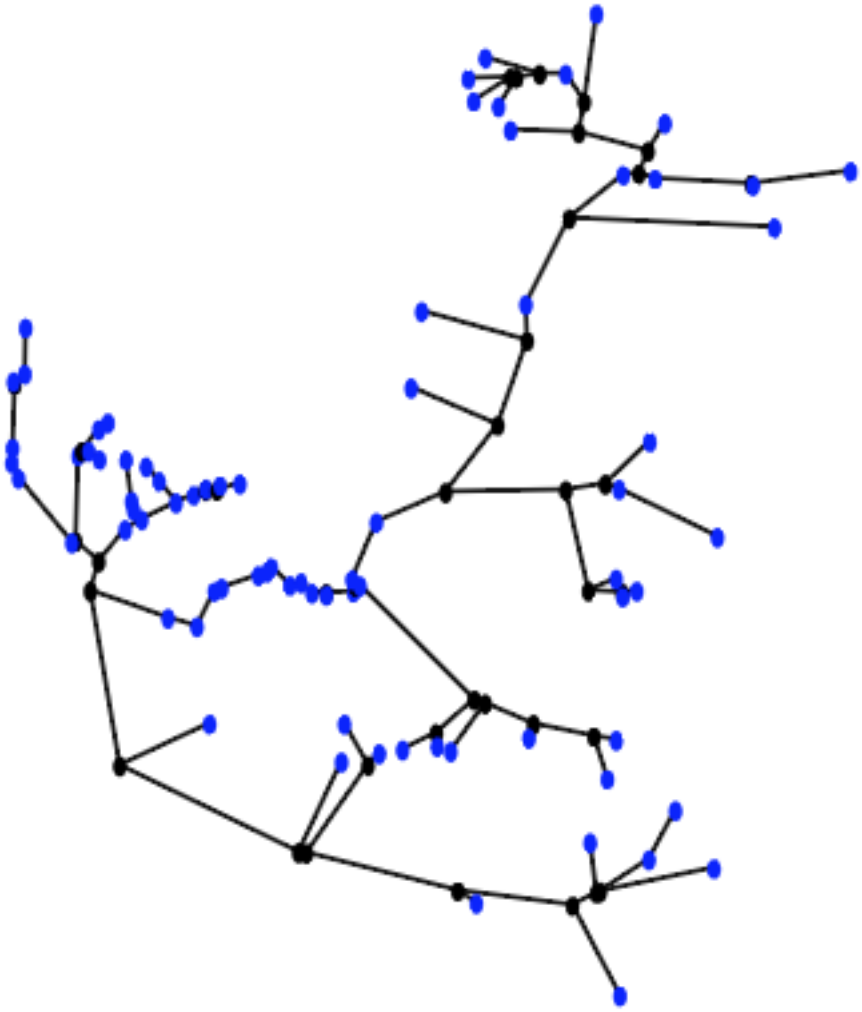


Figure S3: Benchmarking results in RESISTNET showing relationships of peak memory usage and total runtime with increasing numbers of processors. All tests used a dataset of 112 network edges (=2,248 individual contiguous stream segments) and N=78 populations. The model population was composed of 50 models examining 10 randomly selected environmental variables, with the genetic algorithm proceeding for 10 generations. All runs received the same random number seed to reduce stochasticity due to model choice. Tests were performed on a 16-core Linux computer with 256gb memory.

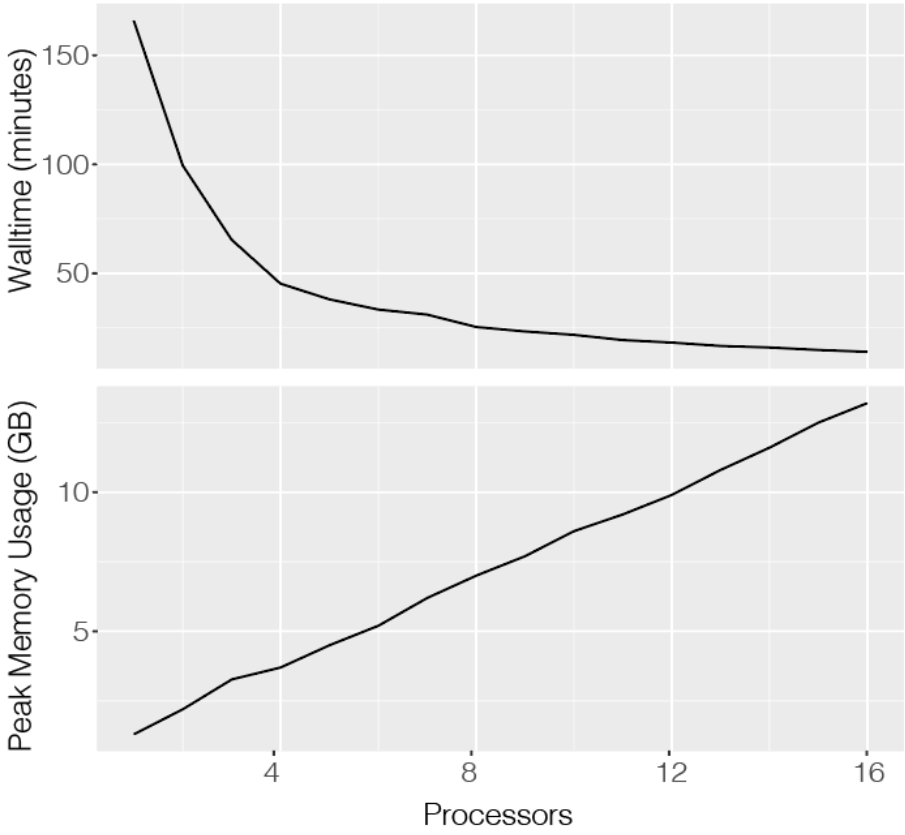
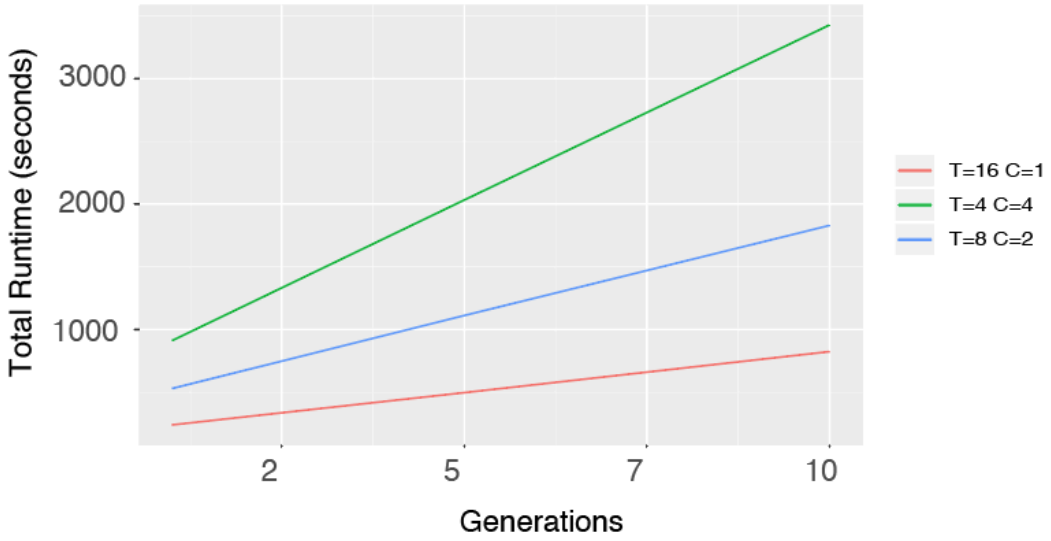
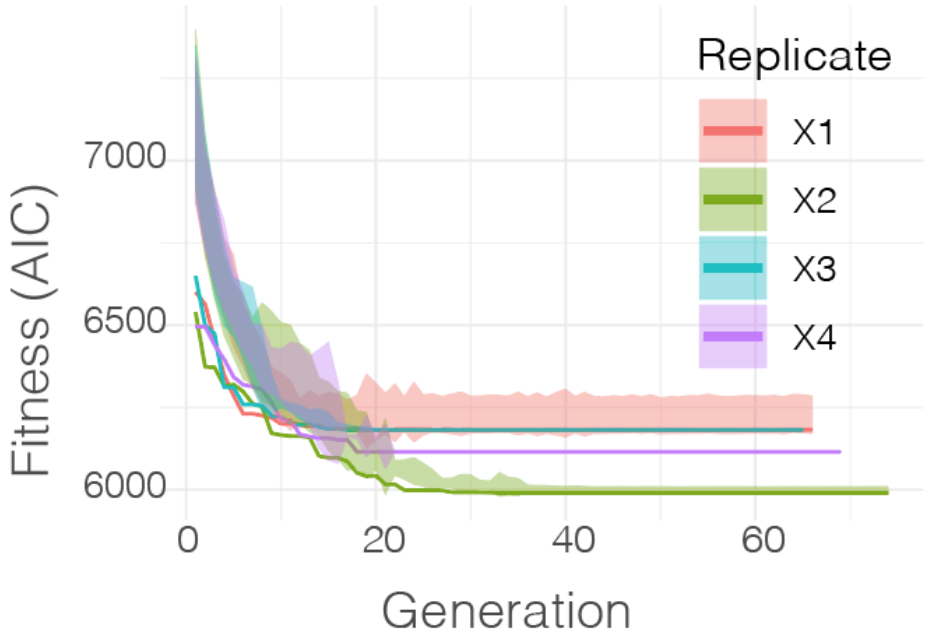


Figure S4: Benchmarking results showing increase in runtime as a function of number of generations when 16 total threads are spread as: 1) 16 parallel processes in RESISTNET, with each dedicating a single process to Circuitscape (T=16, C=1); 2) 8 RESISTNET processes, each dedicating 2 threads to Circuitscape (T=8, C=2); and 3) 4 RESISTNET processes, each dedicating 4 threads to Circuitscape (T=4, C=4). All runs evaluated an identical model population size of 50, with a fixed random number seed.



1165 **Figure S5**: Fitness through time for four RESISTNET replicate runs. Solid lines show the fitness
1166 (AIC value) of the 'best' model in each generation, and the shaded region shows the range for
1167 the population of models (mean +/- standard deviation).
1168



1169

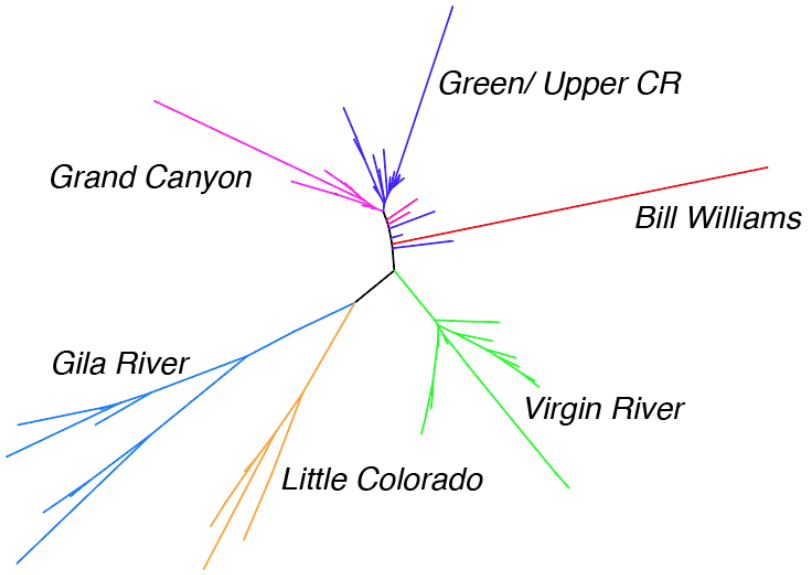



Figure S6: Unrooted species-tree generated using a polymorphism-aware model (PoMo) implemented in IQ-TREE using a GTR+G nucleotide substitution model with 4 rate categories, and a virtual population size of 19 implemented in the polymorphism model.