Genetic Structure of Humpback Chub *Gila cypha* and Roundtail Chub *G. robusta* in the Colorado River Ecosystem

Final Report – May 2007

Principal Investigators:

Dr. Marlis R. Douglas
PH: 970-491-1819; FX: 970-491-5091
Email: Marlis.Douglas@colostate.edu

Dr. Michael E. Douglas
PH: 970-491-7265; FX: 970-491-5091
Email: Michael.Douglas@colostate.edu

Department of Fish, Wildlife and Conservation Biology
Colorado State University
Fort Collins, CO 80523-1474

GCMRC Contract 051884

CSU Project 5-31614

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Humpback Chub, *Gila cypha,* and Roundtail Chub, *G. robusta,* are found in the Upper Colorado River drainage (Green and Colorado rivers and their tributaries, in the states of CO, NM, UT, WY). The former species is also found in Marble and Grand canyons (AZ) of the lower basin Colorado River. Natural populations of a third species, Bonytail, *G. elegans* have been mostly extirpated from the wild. All three species have been affected by anthropogenic modifications of the Colorado River ecosystem, with two (*G. cypha* and *G. elegans*) listed as federally endangered, and the third considered for listing, as well as state-designated 'of concern' throughout its range. We evaluated molecular genetic diversity across four mitochondrial (mt) DNA regions (total 1,869 base pairs) for 336 specimens representing four upper basin populations of *G. cypha,* seven upper basin populations of *G. robusta*, and nine lower basin Marble/ Grand canyon "aggregates" of *G. cypha*. In addition, using larger sample sizes, we also surveyed these same populations and species at 16 microsatellite (msat) DNA loci for a total of 643 specimens.

Neither *G. cypha* nor *G. robusta* could be discriminated using mtDNA, although this marker was successful in separating both from *G. elegans* at 4.7—4.8% sequence divergence (± 0.0—0.1%). The recent coalescence of lineages in *G. cypha / G. robusta* is unusual, especially given (a) fossil history, (b) the broad geographic sampling conducted in this study, and (c) the number (and evolutionary rate) of the mtDNA regions examined. The most parsimonious explanation for these data is that both species were reduced to very small populations by an end-of-Pleistocene warming event and were subsequently forced together into refugial (and shrinking) riverine habitat, thus becoming syntopic with one another. They then hybridized, possibly backcrossing (progeny to parental forms) over an extended temporal span. Eventually, pluvial conditions returned, the aquatic environment subsequently expanded, and both species returned to familiar and exclusive niches within the river where they reproduced with like kind. Nevertheless, the large-scale population reduction would have re-set their mtDNA evolutionary clocks, and hybridization event(s) would have provided both species with the same mtDNA haplotype(s).

These data are also congruent with shallow mtDNA diversity uncovered in a basin-wide study of Flannelmouth Sucker (*Catostomus latipinnis*), giving credence to the fact that the aforementioned perturbation at end-of-Pleistocene was ecosystem-wide and thus affected more than a single species of fish. Two interesting points emerged from the *G. cypha / G. robusta* mtDNA data analysis. One is a lack of haplotype differentiation in Desolation Canyon, a location previously identified in a morphological analysis as exhibiting a 'locality effect' for *G. cypha* and *G. robusta,* where individuals of both species resembled one another more closely than either resembled conspecific populations elsewhere in the basin. Desolation Canyon specimens also reflected the lowest mtDNA k-value (i.e., an estimate of how much on average two random haplotypes differ), thus indicating that *G. cypha* and *G. robusta* are more similar to one another on the mtDNA level at this site than they are at other sites. These mtDNA results mirror the morphological results at Desolation Canyon.

Haplotype trees generated for all specimens graphically display the admixture of individuals with regard to mtDNA haplotype. This confusion extends not only across populations in each basin, but across upper and lower basins, and most importantly, across species as well. Although mtDNA is not well suited for hybrid analysis, our data nevertheless uncovered seven *G. cypha* with a *G. elegans* haplotype (with six of these in Marble and Grand canyons, the other in Desolation Canyon), whereas but a single *G. robusta* was found to possess a *G. elegans* haplotype (also from Desolation Canyon).

Analysis of nuclear DNA offered additional perspectives. Microsatellite (msat) DNA provided sufficient resolution to discriminate among populations and basins when *G. cypha* and *G. robusta* were evaluated. However, one difficulty with our msat analysis is that upper basin *G. cypha* and *G. robusta* are too similar to one another to adequately differentiate them. This problem is particularly manifest in upper basin *G. cypha*. Different software algorithms provide separate perspectives on the number of discrete populations within the upper basin and lower basin, and across both species.

We elect to take a conservative approach to delineating subgroups based on our msat data and select the analyses that recognize six groups. These six-clusters are: (1) *G. elegans*; (2) all Grand Canyon *G. cypha* 'aggregates'; (3) Desolation Canyon *G. cypha and G. robusta*; (4) upper basin G*. cypha* (excluding *G. cypha* from Desolation Canyon); (5) upper basin *G. robusta* (excluding *G. robusta* from Desolation Canyon and Yampa River) *;* and (6) upper basin *G. robusta* from the Yampa River.

Within Marble and Grand canyons, the northern and southern extremities of the distribution of *G. cypha* represent populations with greatest admixture overall. All other Grand Canyon aggregates (from HLCR through HHAV) are linked via gene flow, and this has kept them relatively homogeneous amongst themselves. It has also contributed to the high scores they reflect in our admixture evaluations. Marble and Grand canyon aggregates of *G. cypha* are relatively homogeneous amongst themselves (save for the populations at the upper and lower end of the canyon), yet as a group they are differentiated from upper basin *G. cypha*. However, the two extrema populations (30-mile and Western Grand Canyon) reflect very small sample sizes (eleven and ten, respectively). Given this, there is little statistical basis for making any claims with regard to supposed uniqueness.

Conclusions 1: Our data support the recognition of six 'Management Units' (MUs). The Marble/ Grand canyon aggregates of *G. cypha* would represent a distinct MU, as would Desolation Canyon *G. cypha* and *G. robusta* (as a group). Another MU would consist of *G. robusta* from the Yampa River. In addition, *G. robusta* from upper basin locations (save Desolation Canyon and Yampa River) could be recognized as an MU, as could *G. cypha* from upper basin locations (save Desolation Canyon). In addition, Bonytail would represent the sixth MU.

Conclusions 2: Marble and Grand canyon 'aggregates' of *G. cypha* were not overly distinct at the msat level. Aggregates appeared to be connected by geneflow, suggesting downstream drift of larvae and juveniles as a likely scenario. The Little

Colorado River population would be the primary source, but contribution from occasional local reproduction by mainstem aggregates cannot be excluded.

Conclusions 3: The *G. cypha* population at 30-mile in Marble Canyon was recorded as having two individuals with *G. elegans* haplotypes, and the msat profile for this population is intermediate between genotypes found in Desolation Canyon (a hypothesized hybrid population) and Grand Canyon. Although reproduction has been documented for the 30-mile population, it suffers from chronic low numbers (at least chronic low numbers of catchable fish). However, this is the only population in Grand Canyon that is upstream from the Little Colorado River and is least likely to receive migrants from downstream locations. Due to its potential distinctiveness it should be further studied.

Conclusions 4: Upper basin *G. robusta* are seemingly more distinct than are upper basin *G. cypha*, and *G. robusta* from the Yampa River fall out as distinct in the five- and six-cluster models. Only the four-cluster model groups upper basin *G. robusta* with upper basin *G. cypha.* While *G. robusta* in the Yampa River certainly appears distinct and potentially unique, our data are inconclusive about status of Yampa River *G. cypha*. Additional studies must be conducted on *G. cypha* from the Yampa River, in that only seven individuals were available to us. Such chronic low numbers preclude any statistically valid conclusions. Management decision should not be based on the assumption of Yampa River *G. cypha* being distinct, simply because this is neither supported nor rejected by our data.

Conclusions 5: Our last conclusion centers on *G. elegans*. The individuals we evaluated in this study were hatchery-derived, but in some analyses, they were grouped with the unique *G. cypha/ G. robusta* populations of Desolation Canyon. We identified a single individual from this population that retained both a mtDNA haplotpye and a msat DNA genotype of *G. elegans*. Given the propensity for Colorado River native fishes to hybridize throughout their long history, it is not surprising that alleles characteristic for

G. elegans apparently persist in wild populations of *G. cypha* and *G. robusta*. Similarly, alleles considered characteristic for the latter two species were occasionally detected in *G. elegans* we examined. Albeit historic, admixture among the Colorado River *Gila* species should be taken into account when *G. elegans* are reintroduced into the wild, particularly in locations where the other two species still persist.

INTRODUCTION

Threatened and endangered (T&E) species most often consist of disjunct, fragmented, and demographically constrained populations with low numbers of individuals and without gene flow that normally promotes homogenization (i.e., 'small population paradigm;' Caughley 1994). Such demographics often propel T&E species into an inbreeding spiral that, in turn, reduces individual survival and fecundity, and ultimately, population sizes. This process (i.e., the 'extinction vortex;' Gilpin & Soulé 1986) is self-perpetuating and often leads to extirpation. To stem this process, conservation biologists work not only to group constituent T&E populations into cohesive units for adaptive management (Sites & Crandall 1997), but also to clarify their historic and contemporary connectedness in a fragmented habitat (Waser & Strobeck 1998).

Direct measures of population connectedness are difficult to accrue. Mark/recapture studies are often time-consuming, expensive, and prolonged (Douglas & Marsh 1996, 1998). Radio-tracking, another direct measure (Reed & Douglas 2002), is surgically invasive, constrained by small sample sizes, and yields statistically dependent data (Reinert 1992). Furthermore, direct measures do not reflect gene flow *per se* because migrants must not only locate and reproduce in new demes but their progeny must survive to adulthood. This often fails, even under 'best-case' scenarios (Douglas & Brunner 2002).

Many of these difficulties, however, have been resolved through application of rapidly evolving, co-dominant microsatellite (msat) DNA loci (Schlötterer & Pemberton 1998). These are common in the genome, often highly polymorphic, and are now a focus of considerable research pertaining to population demography and divergence (Balloux & Lugon-Moulin 2002, Zhang & Hewitt 2003). Gene flow and dispersal are quantified by studying distributions of alleles within/among populations (Berry *et al.* 2004), while population abundance estimates are determined by mark-recapture methods using multi-locus genotypes as 'genetic tags' (Palsbøll 1999). Furthermore,

these techniques have been enhanced by Bayesian or coalescent-based approaches that not only assign individuals to demes (Pritchard *et al.* 2000), but also quantify first- or second-generation inter-demic movements (Wilson & Rannala 2003). One shortcoming of msat data is the ambiguous ancestral information they contain. Here, mitochondrial (mt)DNA data can contribute by not only clarifying evolutionary relationships but also by facilitating inferences regarding past demographic, ecological and climatological history (Douglas *et al.* 2006).

In this study, we applied coalescent and maximum likelihood analyses of mt- and msat DNA to infer contemporary vs historic gene flow and demographic parameters in populations of two Colorado River fishes, one federally listed as “endangered” and the second recognized as a “species of concern” (status reviewed in Bezzerides & Bestgen 2002). We particularly wished to evaluate if populations of these two species have diversified in light of two alternative hypotheses. One invokes late Pleistocene/early Holocene desertification and concomitant isolation of riverine populations (per Douglas *et al.* 2003), while the second reflects more recent and human-mediated impacts that have rendered remaining riverine habitat into a mosaic. We wished to determine which of these scenarios has impacted our study species most severely (if at all). Clearly, such data are necessary prerequisites for the management of the Colorado River ecosystem and its resident ichthyofauna. These hypotheses, in particular, impact one study species that resides within the Colorado River of Grand Canyon. The operation of Glen Canyon Dam is removing silt from the water, damping peak flows and water temperature variation. These anthropogenic effects have transformed the aquatic ecosystem within Grand Canyon (Blinn & Cole 1991), and have had direct impacts upon fish populations (Minckley 1991). How the operation of Glen Canyon Dam affects this endangered species is at the core of the project, particularly with regard to fragmentation of the river and consequences for interrelatedness of populations and their propensity for sharing genes.

BACKGROUND

Gila cypha in Grand Canyon

The mainstem Colorado River native fishes are renowned for their specialized morphologies (Minckley 1991), and this aspect reaches its culmination in the phenotype of *Gila cypha* (Miller 1964, Douglas 1993). Although it is endemic to the Colorado River and its major, swift-flowing tributaries (Holden & Minckley 1980), *Gila cypha* occurs but sporadically and in relatively small numbers. For example, it has been recorded from the gorge sections of the Green (Desolation Canyon) and Yampa rivers (Douglas *et al.* 1989), the Colorado River above Lake Powell (Cataract Canyon), and above its junction with the Green River (Black Rocks and Westwater canyons; McElroy & Douglas 1995) (see Fig. 1A). Historically, *G. cypha* also inhabited other canyon-bound reaches of the Colorado River, as documented from archaeological remains (Miller 1955, Miller & Smith 1984, Sigler & Miller 1963). Although *G. cypha* is quite distinct morphologically (McElroy *et al.* 1997), it was often confused with congenitors when management was attempted (per Douglas *et al.* 1999a). It was the last fish to be described from the mainstem Colorado River (Miller 1946), with the type specimen being caught in 1932 at Bright Angel Creek, Grand Canyon National Park (Carothers & Brown 1991:95).

The largest population of *G. cypha* in the Basin, estimated at approximately 6,000 individuals, inhabits the Little Colorado River in Grand Canyon (Douglas & Marsh 1996). While *G. cypha* is distributed patchily within Grand Canyon, its presence is consistent within certain specific mainstem areas, and suggests these populations may exist with some permanence within these Grand Canyon reaches. For example, specimens are often caught within reaches associated with tributaries or specific geomorphic features. This has led to recognition of nine ill-defined mainstem “aggregations” (H. Maddux unpubl., Valdez & Ryel 1995). Four of these represent individuals pooled by river reach, whereas data for three others are so sparse that population estimates cannot be calculated. Two more are located immediately above and below the confluence of the Little Colorado River, and recapture data (Douglas & Marsh 1996) indicate free exchange of individuals among the three locations. Hence, the areas above and below the Little Colorado River

should be included with the latter “aggregation.” Given this, only five (of nine) “aggregates” may have actual validity as independent assemblages of *G. cypha*. These are: 30-mile Springs (river mile 30), Shinumo Creek (river mile 108), Middle Granite Gorge (river mile 126), Havasu Creek (river mile 156), and Pumpkin Springs (river mile 213) (Figure 1B).

Uncertainty exists with regard to the origin and life history of these mainstem aggregates. Are they merely pre-dam relicts of a once more broadly distributed metapopulation? Are they maintained by sporadic local reproduction or are they instead rejuvenated through occasional dispersal of adults and/or drift of larvae spawned in the Little Colorado River? The only mainstem Colorado River population for which *in situ* reproduction has been documented is the complex at 30-Mile (Valdez & Masslich 1999). This aggregate is also the only one upstream of the Little Colorado River confluence, and is of interest because it cannot be influenced by the potential downstream drift of larvae or juveniles from Little Colorado River. As such, it may possibly represent the aggregate most similar to upper basin *G. cypha* stock.

Life history of *G. cypha* in Grand Canyon remains mostly enigmatic. Douglas & Marsh (1996) demonstrated long-term residency by this species within the Little Colorado River, particularly summer-through-winter. Several hypotheses were presented to accommodate these data. One suggests residency in the Little Colorado River is a pre-dam component of *G*. *cypha*'s life history. Another proposes that it is a post-dam alteration. A third advocates a combination of the two. With regard to the first, long-term residency by adults within the Little Colorado River may have always been an aspect of *G*. *cypha*'s life history. We know, for example, that *G. cypha* spawned within the pre-dam Little Colorado River during spring (Kolb & Kolb 1914:127, Carothers & Brown 1991:93). However, its duration of stay was unknown.

Alternative hypotheses assume that temperature and flow conditions in the Little Colorado River might be more similar to those of the pre-dam mainstem Colorado River, and are amenable to habitat requirements of indigenous fishes shaped over evolutionary time. The first suggests that the altered thermal regime of mainstem Colorado River has

forced *G*. *cypha* to adjust its life history such that it now accommodates lower mainstream temperatures primarily through avoidance. This hypothesis is anecdotally supported by three facts: First, active movement into/ from the Little Colorado River are primarily accomplished by larger (and presumably older) *G*. *cypha* (Douglas & Marsh 1996). The species attains great age (20+ years; Minckley 1991:150) and larger adults may thus represent mainstem-adapted individuals from pre-dam cohorts. Secondly, larvae and juvenile *G*. *cypha* are often transported via flood into the mainstem, but, until recently, small adults are seldom taken there (Douglas & Marsh 1996). Kaeding & Zimmerman (1983:585) similarly noted an absence of intermediate-sized individuals in the mainstem Colorado River above the Little Colorado River confluence, even though mature fish were present in this area. If *G*. *cypha* has altered its life history to accommodate dam-induced changes, then its long-term persistence within Grand Canyon is tied more intimately to the Little Colorado River than previously believed.

A second hypothesis assumes that *G. cypha* is actually composed of two co-existing forms within Grand Canyon, a main stem form and a tributary form, each with different life-history strategies. An alteration of its habitat (similar in tenor to the first hypothesis) could have forced the mainstem form to reproduce in the Little Colorado River. Closed population estimates calculated by month (Douglas & Marsh 1996, unpubl.) reveal two abundance peaks each spring plus an additional one in autumn that could be inferred as temporal variation in spawning between subpopulations or stocks. If indeed two life-history forms exist, the Little Colorado River population might represent an admixture, and as such, would mandate management that takes different life history forms into consideration.

Gila in the Colorado River Ecosystem

How anthropogenic alterations of the aquatic ecosystem affect Colorado River Basin fishes, particularly with regard to fragmentation of the river and consequences for geneflow among populations, must be evaluated because conflicting human demands and expectations with regard to uses of the Colorado River have provoked a classic confrontation within the Basin (Wydowski & Hamill 1991). Scientists and resource

agencies work to conserve the unique biological diversity within the Colorado River ecosystem (Fradkin 1984), while demands for consumptive use of its water are steadily increasing, concomitantly with population growth in the region.

At risk is a unique and endemic ichthyofauna of ancient origin, extending as far back as the Miocene (Miller 1959, 1961, Minckley *et al.* 1986, Douglas *et al.* 1999a). This fauna possesses remarkable morphological and anatomical modifications that are often explained as adaptations for survival in a historically turbulent and sediment-rich aquatic environment, one that has long experienced massive seasonal pulses in flow (Minckley 1991). However, the majority of these fishes have now become endangered (or candidates for such status), based upon their declining numbers (Bezzerides & Bestgen 2002). The interconnectedness of their populations is of major concern, in that the mainstem and its tributaries are now subdivided by dams and diversions, and the waters allocated to a myriad of agricultural and urban sinks (Miller 1982, Minckley & Douglas 1991).

Relationships among Gila and outgroups

When evaluating upper and lower basin (referred to as Grand Canyon) populations of *G. cypha*, and their relationships with upper basin *G. robusta* and hatchery-reared *G. elegans*, an understanding of relationships within the Cyprinidae (Minnows) is imperative so as to polarize mtDNA character states and provide a direction of evolution. In this regard, the cyprinids are divided into two subfamilies, the Cyprininae and the Leucisinae, with the former considered more ancestral and consisting of three clades (cyprinins, barbins and labeonins). *Cyprinus carpio* (Common Carp) is a member of the cyprinin clade (Cavender & Coburn 1992), and thus forms a suitable root for our analyses.

North American cyprinid genera fall within the more derived subfamily Leucisinae, which is divided into two broad lineages, the leucisins (with a single North American representative), and the phoxinins (which make up the remainder of the North

American cyprinid diversity) (Cavender & Coburn 1992). The western clade of phoxinins is the sister group to the remaining North American forms (Simons & Mayden 1998), and the genus *Gila* is the most derived clade of this western group. However, Simons *et al.* (2003) found that *Gila* does not comprise a monophyletic group, for *Ptychocheilus lucius* (the Colorado Pikeminnow) consistently falls between *G. coerulea* and the remainder of the genus. Similarly, *P. oregonensis* did not cluster with *P. lucius*, as expected, but instead falls outside of the *P. lucius* / *Gila* clade. Simons *et al.* (2003) suggested that historic introgression between *P. lucius* and *Gila sp*. may be a potential factor that has allowed *P. lucius* to retain *Gila*-like haplotypes. However, the Simons *et al.* analyses included but a single specimen of *P. lucius*. Our analyses included five *P. lucius*, and five *G. elegans* (Bonytail) as outgroups. Thus, our outgroup analyses should be able to assist in defining these outgroup relationships.

METHODS AND MATERIALS

MtDNA: Sampling, extraction and amplification

Samples of *G. cypha* and *G. robusta* were collected by individuals from state and federal agencies in Arizona, Colorado, Utah and Wyoming (see Acknowledgments). Additional samples had been previously collected by the Principal Investigators under federal and state permits, and under IACUC protocols issued by Arizona State University and Colorado State University. Total genomic DNA was isolated from these samples using the DNeasy® Tissue Kit (Qiagen Inc., Valencia CA) and stored in DNA hydrating solution (same kit). Three mitochondrial genes, ATPase 8 (ATP8), ATPase 6 (ATP6), NADH Dehydrogenase subunit 2 (ND2), as well as the non-coding displacement loop (D-Loop) were amplified using primers and conditions specified in Douglas *et al.* (2002).

Mitochondrial (mt) DNA accumulates mutations five to ten times faster than nuclear genes (Harrison 1989, Brown *et al.* 1979), and is thus favored for studies of intraspecific diversity and population subdivision because it can often resolve relationships among recently diverged lineages (Avise 2000). It is maternally inherited with no recombination, which necessitates fewer samples to encapsulate population variance. Methodologies to assess mtDNA variation are also well established, and universal primers allow a variety of mtDNA regions to be amplified, each with a varying evolutionary rate. However, parameters related to heterozygosity, inbreeding, and expression of deleterious recessive alleles cannot be addressed with this marker, but are often of interest for conservation of small populations (as herein).

Once the mtDNA had been amplified, we conducted double-stranded sequencing reactions on the product, using fluorescently-labeled dideoxy terminators according to manufacturer recommendations [Applied Biosystems Inc. (ABI), Foster City, CA]. The labeled extension products were electrophoretically analyzed with an automated DNA sequencer (ABI Prism 3100), and the resulting sequences aligned manually using

SEQUENCHER (Gene Codes, Ann Arbor, MI). We tested the effectiveness of combining sequences for analyses by using the ILD (incongruent length difference) test (Farris *et al. 1*994) (implemented as the partition homogeneity test in PAUP*: Swofford 2001).

MtDNA sequence evolution, neutrality, and regularity of substitutions

We determined the model of sequence evolution that best fit each marker (above) by using MODELTEST (Posada & Crandall 1998). To ensure selective neutrality, we applied the HKA test (Hudson *et al.* 1987) based on the fact that all mtDNA regions are found on the same chromosome. The HKA test (implemented in DNASP; Rozas *et al.* 2003) constructs a goodness-of-fit (Chi-square) statistic using observed and expected differences within two sister species, and the differences between the two taxa for the DNA regions being compared. We evaluated ATP8 vs. ATP6 in *Gila* vs. the same regions in *P. lucius*.

We also tested the neutrality of each mtDNA gene using the McDonald-Kreitman (MK) test (McDonald & Kreitman 1991). This test compares levels of polymorphism and divergence using two sets of data interspersed within each sequence. Unlike other tests of neutrality, the MK test does not assume panmixia, thus making it appropriate when population subdivisions may exist. Under neutrality, the ratio of fixed synonymous-to-nonsynonymous substitutions between sister taxa will be equal to the ratio of polymorphic synonymous-to-nonsynonymous substitutions. Selection is inferred when the variance of this ratio among loci exceeds expectations.

To ascertain clock-like behavior in our sequences, we applied Tajima's (1993) test to compare representative sequences from study clades vs. outgroup. Tajima's test is based on the expectation that under a uniform (i.e., clock-like) rate of substitution, numbers of sites shared by outgroup and one of two ingroups should be the same for both ingroups. We performed five random evaluations per clade, each involving three comparisons.

MtDNA diversity and demographic history at the drainage level

To examine broad-scale patterns of regional biodiversity, we first evaluated populations by species within sub-drainages (upper vs lower basin), after first excluding from analysis those individuals with *G. elegans* haplotypes. We then pooled these populations by sub-drainage (per Table 1). For each population and sub-drainage, we calculated numbers of haplotypes, total number of polymorphic sites in our sequence data, and how much (on average) two haplotypes differed from one another within a drainage, using DNASP. We also derived haplotype (*h*) and nucleotide diversity (π), where *h* is a measure of the frequencies and numbers of haplotypes across a particular drainage, and π is the average weighted sequence divergence among haplotypes within a population or sub-drainage, varying between 0% for no divergence to over 10% for deep divergence. Values provide an estimate of the probability that two randomly chosen homologous nucleotides are identical.

We also used our mtDNA data to generate sequence divergence (*p*) values among haplotypes, sub-drainages, and species, based on 1,000 bootstrapped sequences (using MEGA3; Kumar *et al. 2*004). We then calculated Tajima's D-statistic, and (if neutrality was sustained), applied the latter to infer demographic history (Tajima 1989). For a stable population, D = zero, whereas it is positive for an excess of high-frequency mutations, as after a population contraction or under balancing selection. Tajima's D is negative when there is an excess of low-frequency mutations, as after a population expansion, a recent selective sweep, weak negative selection, or when a sample comes from an admixed population. Given the possibilities, causation is difficult to ascertain when Tajima's D deviates significantly from zero. To clarify, we computed Fu's Fs (Fu 1997) which is particularly good at detecting population expansion. We derived *h,* π, Tajima's D and Fu's Fs in DNASP using the coalescent with 1000 replications.

MtDNA: Phylogenetic analyses

Using a total evidence approach, we combined all four mtDNA markers into a single sequence for subsequent analysis. We used maximum parsimony (MP) analysis to derive minimum length trees for these data, using TNT (Tree analysis using New Technology, ver.1; P.A. Goloboff, J.S. Farris, K. C. Nixon 2003; www.zmuc.dk/public/phylogeny), with the following parameters selected: Random Sectorial Searches (RSS)=43/45/45; Consensus-based Sectorial Searches (CSS)=3/10; Tree-Fusing (TF)=3 (see Goloboff 1999). We then used PAUP* (Swofford 2001) to produce a majority rule consensus of the 12 fused trees.

MP was selected as a method of analysis in that it is hypothesis-free. In this sense, it does not necessitate the incorporation of evolutionary rates for the different compiled genes or regions sequenced (even though we calculated these) (see Steel & Penny 2000). This stands in contrast with other analytical methods, such as Bayesian analysis (BA; Huelsenbeck *et al.* 2001) or maximum likelihood (ML) (Sober 2004), which require additional data.

Under a parsimony algorithm, the most widely used strategy for finding optimal trees is to employ random addition sequences in conjunction with tree bisection-reconnection branch swapping (Nixon 1999). This strategy usually works well when the numbers of taxa are reduced, but larger data sets (i.e., those > 50 taxa) have proven problematic (Goloboff 1999). This is because the latter contain numerous composite optima (or tree islands). These in turn make it difficult to identify a globally optimum tree. Instead, a rather large number of suboptimal trees are produced. Although these reflect identical tree lengths, they differ amongst themselves with regard to minor rearrangements. Their accumulation either fills system memory to capacity or overly taxes the patience of researchers. Larger data sets thus often require search strategies that specifically deal with the problem of composite optima. One (the parsimony ratchet) was demonstrated in Nixon (1999) while a second (employing tree-fusing, tree-drifting, and sectorial searches) was described in Goloboff (1999). We employed the latter to derive minimum length, fused trees for our data set.

Microsatellite (Msat) DNA: Background

Msat loci offer several advantages over other types of molecular markers, in that the former are abundant and highly variable (Ashley & Dow 1994, Burke 1994). Loci are characterized by tandem repeats of short, specific motifs of one to six nucleotides, and are among the fastest evolving DNA sequences (Weber & Wong 1993). They have been successfully used to investigate microgeographic divergence of fishes (Douglas *et al.* 1999b), and were able to detect high levels of genetic differentiation when mtDNA and allozymes failed to do so (Brunner *et al.* 1998). In addition, msat loci have become a useful tool for management of fish stocks (Bentzen *et al.* 1996, Ruzzante *et al.* 1997, Douglas *et al.* 1999b). Msats have an advantage over RAPDs (Random Amplified Polymorphic DNA), in that Mendelian data about specific loci are provided. Further, since msat repeats are randomly distributed, the entire genome rather than a localized area (such as MHC complex) can be assessed.

A disadvantage to msats (Burke 1994) is that loci must to be developed for a particular target species. This is often a time-consuming and expensive proposition. A second disadvantage is the limited and controversial application they offer for phylogenetic interpretations (Feldman *et al.* 1997, Goodman 1997). This limitation was addressed in our study by concomitant analyses of mtDNA sequence variation. Also problematic is the potential presence of null alleles (as determined from homozygote excess at a particular locus). Again, these were tested for in our study prior to analysis (see below).

Because msat loci have rapid mutation rates and often yield a large number of alleles, they are well suited for projects that examine genetic or demographic population structure. However, a large number of alleles also necessitates large sample sizes, so as to determine an accurate reflection of genotypic frequencies. Sample sizes between 30 and 50 individuals are considered necessary for precise estimation of genetic distances and this value also depends on numbers of loci, numbers of alleles, and the range in allele size (Ruzzante 1998, Estoup *et al.* 2000). While resampling procedures often allow larger

sample sizes to be simulated from the data, they do not overcome those potential biases induced by small sample sizes.

Msat DNA: Sampling, extraction and amplification

DNA was extracted as per mtDNA (above), but additional samples of both species were employed in the derivation of msat data (as listed in Table 5). Various sources were used to select msat loci that would amplify in our three study species and which would yield moderate levels of polymorphism. A total of 51 msat loci were tested, from which a subset of 20 loci (Table 6) was selected for genotyping across all samples. All loci were first tested via PCR/ agarose gel electrophoresis for successful amplification across 8-16 individuals representing populations from throughout the Colorado River basin. Some loci did not produce any identifiable PCR products and were subsequently excluded from further tests. Those loci that yielded multiple fragments were evaluated with a gradient PCR, where a range of annealing temperatures (~48-62°C) could be examined. To further investigate the extent of polymorphism in successfully amplified loci, PCR products were separated on Spreadex 500 gels, which yield much higher resolution than do conventional agarose gels. Loci that appeared monomorphic (i.e., with a single allele) were discarded. Only those loci that showed polymorphism were selected for multiplex tests.

Table 6 indicates our multiplex protocol. One primer from each pair listed in multiplexes 1-4 was modified by the addition of the M13 sequence (TGT AAA ACG ACG GCC AGT) to the 5' end, so as to allow for direct incorporation of a fluorescently-labeled M13 primer. PCR conditions were then re-optimized, because the addition of this tail modified the melting point (Tm) of the primer. This process proved to be time-consuming, and forward primers for multiplex set 5 (Table 6) were directly dye-labeled. Loci with the modified primers were amplified using a three-step annealing procedure, whereas loci with dye-labeled primers (set 5) required a less-involved two-step annealing procedure.

PCR reactions were in 10-µl volumes, each containing 20 ng of total DNA, 10 pmol of each primer, 0.3-mM of each dNTP, 2.0 mM $MgCl_2$, 1X *Taq* buffer, 0.4 ul BSA and 0.2 units of Go-*Taq* polymerase (Promega Corporation). Amplifications were carried out on an Applied Biosystems 9700 thermal cycler. PCR products were separated on an ABI Prism 3100 Genetic Analyzer, using POP4 and standard electrophoresis parameters. An internal size standard (Liz 500) was run with each sample.

Alleles were evaluated using GENESCAN 3.7, scored in GENOTYPER 3.7, and with allele sizes confirmed by visual inspection. Allele scores for each population across the 20 msat loci were compiled into spreadsheet databases. We then employed MICRO-CHECKER (Van Oosterhout et al. 2004) to identify scoring problems and to evaluate potential genotyping errors due to stutter peaks, short-allele dominance (large allele dropout), and non-amplified (null) alleles.

Msat DNA: Preliminary analyses

Populations were tested for departure from Hardy-Weinberg (HW) equilibrium at each locus, and over all loci. The HW test utilized a Markov chain randomization test (Guo & Thompson 1992) to derive exact 2-tailed probability values for each locus using GENEPOP 3.4 (originally described in Raymond & Rousset 1995a). An option in the same package allowed tests for either heterozygote deficiency or excess (as per Rousset & Raymond 1995). Linkage disequilibrium at each locus was also evaluated for each population using GENEPOP 3.4. Statistical significance was determined using a Bonferroni adjustment (Rice 1989).

Unequal sample sizes can bias estimates of allelic diversity and we corrected for this by employing MSA (MSAT ANALYSIS: Dieringer & Schloetterer 2003) to derive minimum, maximum and variance in allele size (corrected for sample size). Genetic polymorphism was calculated in GENALEX 6 (Peakall & Smouse 2006) for each population, and the following parameters were derived: Numbers of alleles per locus (N_A), mean numbers of alleles (A), and effective numbers of alleles (N_e). The same package also compiled frequencies of less common alleles.

Msat DNA: Statistical estimates population differentiation

Population differentiation can be estimated using either statistical or clustering methods. With regard to the former, allelic goodness-of-fit tests are most appropriate, particularly when sample sizes are unequal. Thus, to test our populations (Table 5) for differences in allele frequencies, we used a randomization test in GENEPOP to evaluate the significance (if any) of genetic differences among species and populations, with results combined over loci using Fisher's exact test. The latter has proven robust and with a high resolving power (Ryman *et al.* 2006).

Contingency tests of allele frequency heterogeneity followed the methods of Raymond and Rousset (1995), which uses Markov chain Monte Carlo (MCMC) methods to provide an unbiased estimate of the exact probability for each single locus comparison. Calculations were performed using program TFPGA, which computes a multilocus *P*-value for each comparison using Fisher's exact method for combining probabilities across loci. Approximately 30 independent batches were run, each with 2,000 dememorization steps and 10,000 permutations. The Bonferroni adjustment for each population-comparison was $P = 0.0025$. Populations identified homogeneous at this probability level were considered to comprise but a single gene pool (i.e., panmictic population).

For populations showing overall heterogeneity, the number of different populations represented by the samples was derived in the following manner. TFPGA was used to determine which of the pair-wise comparisons differed at the Bonferroni-adjusted probability value, and these were then linked. A group of samples were considered as the same population if every pair in the group could be connected through a chain of non-significant results (per Waples and Gaggiotti, 2006).

Msat DNA: Clustering estimates of population differentiation

In contrast to the statistical procedures (above) that estimate differentiation among geographically defined and locale-specific populations (per Table 5), clustering methods instead create groups of individuals simply based upon multi-locus genotypes, and without considering the precise geographic locale of the individual. We utilized this approach by employing two different algorithms. The first was program BAPS (Bayesian Analysis of Population Structure, ver. 4.13: Corander & Marttinen 2006). This program analyzes both the allele frequencies of the msats and the number of genetically diverged groups as random variables. The estimation of membership coefficients (i.e., membership in a cluster) proceeds simultaneously along with the estimation of allele frequencies. BAPS does not employ a Markov Chain Monte Carlo (MCMC) approach to determining the posterior distribution. Instead, it uses a 'greedy algorithm' (a stochastic optimization approach) to search for the most likely k-value. One benefit is that the BAPS analysis is performed quite quickly, which is an added bonus when the size and complexity of the project are considerable (as herein). BAPS also allows individuals to be of mixed ancestry (i.e., admixed), and proportionally assigns individual genomes to clusters.

For our analyses, we initially specified seven different distributions of potential populations in our data ranging from 1—30, in increments of 5 (i.e., k =1, 5, 10, etc., to 30). Our thinking was that the number of populations in our study could potentially run from k=1 (a single panmictic population) through k=21 [where every designated population (i.e., n=20 + *G. elegans*) was a distinct entity]. We then extended the ceiling from 21 to 30, so as to allow as much room for expansion as reasonably possible. If the results of our analyses fixed on 30 as the optimum number of populations in our data, our strategy would then be to extend the ceiling even further in subsequent runs.

Once BAPS had determined the number of clusters (k), the results serve as input to an admixture analysis that will diagrammatically display the proportion of mixing that has occurred among the identified clusters. Here, we identified minimum population size = 5, with 1000 iterations to identify admixture coefficients for individuals, and 200

individuals/ per population as reference (per software instructions), with 100 iterations done to evaluate admixture for the references (deemed of less importance for the analysis). Output is a horizontal bar-graph with colored partitions representing clusters, and populations (per input data) identified below the bar and separated within the bar by vertical partitions. Each individual is represented by a vertical bar, colored according to group membership, but the vertical bar is split into several colors when there is evidence of admixture. Each color of the bar represents the source of the admixture (cluster), and the proportion of each color represents the proportion of the genome estimated to be represented.

The clustering procedure to determine the maximum number of clusters within our msat data was replicated by applying a second (but computationally slightly different) Bayesian-based program (STRUCTURE, ver. 2.1: Pritchard *et al.* 2000). As with BAPS, this program not only identifies the number of genetically distinct clusters that maximizes the probability of the data, but it also assigns individuals to clusters and identifies migrants and admixed individuals. STRUCTURE also plots the estimated membership coefficients for each individual in each cluster. When running this program, we utilized the Admixture Model with a uniform prior on degree of admixture (prior mean = 0.01, prior SD = 0.05; lambda = 1.0), and with allele frequencies assumed to be correlated among subpopulations. These configurations are considered optimal for detecting subtle population structure (per Falush *et al*. 2003). For each of our k-values, we ran the MCMC algorithm for 1,500,000 iterations, preceded by an initial (discarded) burn-in of 50,000 iterations. The MCMC generates an estimate of the posterior distribution of the sample partition. Assignment of individuals to clusters was then accomplished using the greatest likelihood-of-assignment (q) value (Pritchard *et al.* 2000), while optimal number of clusters was determined using Δk (per Evanno *et al.* 2005). Admixture in program STRUCTURE was evaluated graphically in the same manner as presented for program BAPS.

We also separately evaluated as a single group the nine populations from Marble and Grand canyons, so as to determine the extent of gene flow among these

aggregates (if any). Here, we hypothesized that larval drift and potential migration of juveniles/ adults may have played a role in uniting locations one to another in a downstream manner (but excluding the 30-mile population which is upstream of the Little Colorado River). We surmised that population linkages (if present) would be most apparent within Marble and Grand canyons, and more diffuse within the upper basin, where sample locations are separated by much greater distances.

FSTAT (Goudet 2002) was used to calculate estimates of the inbreeding coefficient [F_{IS}: ranging from -1.0 (all individuals are homozygous) to +1.0 (all individuals heterozygous)] and the fixation index [F_{ST}: ranging from 0.0 (no differentiation) to 1.0 (complete differentiation)]. These were computed among species (N=2) and geographically defined populations (N=20), with significance estimated via permutation tests. GENEPOP was used to derive R_{ST} values (where mutational differences account for allelic diversity) across the same subsets, which were then contrasted against F_{ST} values (where migration and drift produce allelic diversity) so as to determine which process was potentially more influential in contributing to the observed msat diversity. .

Msat DNA: Statistical differentiation among clusters

Based on the STRUCTURE analysis (above), we grouped populations into six clusters, so as to statistically evaluate genetic diversity within and among these clusters. Analyses were performed in ARLEQUIN 3.1(Excoffier *et al.* 2005).

Distribution of microsatellite DNA diversity was quantified using the analysis of molecular variance model (AMOVA). Genetic correlation measures are estimated by taking the variance in allele size into account (Michalakis & Excoffier 1996). Components of genetic variance were computed at three hierarchical levels: Within populations (F_{ST}), among populations within clusters (F_{SC}), and among clusters (F_{CT}). The resulting values were tested for significant departure from zero using permutations of alleles within- and between-clusters, as well as among clusters.

Genetic divergence among clusters was also evaluated by calculating F_{ST} values and average number of differences between pairs of clusters. Computations were performed in ARLEQUIN 3.1 using permutations to test for significance.

RESULTS

MtDNA: Sampling/ amplification, sequence evolution/ neutrality and clock-like behavior

Samples are listed by population and drainage in Table 1, and are defined cartographically in Figure 1. Populations (n=21) of two species were sampled from two different sub-basins across four states (Arizona, Colorado, Utah, and Wyoming; Appendix 1). Sample sizes per population ranged from 3 to 46 (mean = 15.8, stdv. = 11.9). Total numbers of *G. cypha* were 214 (62 from upper basin, 152 from Grand Canyon) while total numbers of *G. robusta* were 117 (all upper basin). When both species were combined, we totaled 331 individuals. Three out-groups were also used [*Gila elegans* (n=5); *Ptychocheilus lucius* (n=5); *Cyprinus carpio* (n=1)]. Thus, total numbers of OTUs (operational taxonomic units) in this aspect of our study = 5, and total number of individuals = 342.

The primers amplifying the ATPase 8 and 6 mtDNA genes produced a single fragment, because the two genes overlap one another at 10 bp. The complete length of the ATP-8 gene (168 bp), and a portion from the 5' end of the ATP-6 gene (464 bp) were sequenced, yielding a total of 642 bp of sequence data. Some 17 haplotypes were produced using the ATPase genes, with an average of 19.5 individuals/ haplotype. The distribution of ATP haplotypes is shown in Appendix 1 by population, species, and sub-basin. Two haplotypes (A1 and A2) representing 230 individuals (or 69.5% of the total) spanned both sub-basins and species, while two other haplotypes (A4 and A6; total 18 individuals) spanned upper basin *G. robusta* and *G. cypha*. Interestingly, one haplotype (A9) spanned *G. robusta* (one individual) and Grand Canyon *G. cypha* (two individuals); it was likely not detected in upper basin *G. cypha* due to its low frequency (just below 1% of all individuals in the study). Four haplotypes (23.5% of total) were found to represent *G. elegans* (detected across eight individuals, or 2.4% of all individuals in the study), and one of these haplotypes (A5) occurred in five individuals and was shared between upper basin and Grand Canyon *G. cypha*. In all, six (35.3%) of

the haplotypes spanned sub-basins and/or species [represented across 256 (77.3%) of all individuals].

For the ND2 gene, partial sequence of 589 bp from the 5' were generated. These produced 23 haplotypes (with an average of 14.4 individuals/ haplotype). One haplotype (N1) spanned both sub-basins and species, and was represented in all populations accounting for 250 individuals (75.5% of the total) (Appendix 2).Two haplotypes (N4 and N6) were shared between upper basin *G. cypha* and *G. robusta* (two individuals each), and another haplotype (N12) was shared between upper basin *G. robusta* and Grand Canyon *G. cypha* (three individuals total). As per ATP, four haplotypes (17.4% of total) represented *G. elegans* (detected across the same eight individuals). In all, only five of 23 haplotypes (21.7%) were shared either among species or among basins, but they represented 258 individuals (77.9% of the total).

Approximately 638 bp were sequenced across the D-loop, and these data partitioned into 23 haplotypes (again, 14.4 individuals/ haplotype). Two haplotypes (D2 and D3), representing 248 individuals (74.9% of total) were spread across both species and sub-basins (Appendix 3). Another three haplotypes (D4, D5, and D6), found in12 individuals (3.6% of all individuals) represented both upper basin *G. cypha* and *G. robusta*, whereas a single haplotype (D15), found in two individuals, spanned upper basin *G. robusta* and Grand Canyon *G. cypha*. Six total haplotypes (26%) thus spanned sub-basins and/ or species, accounting for 262 individuals (79.2% of total). Five haplotypes (8 individuals) represented *G. elegans*, but none of these spanned species or sub-basins.

Combination of sequence data obtained for ATP 8/6, ND2 and D-Loop fragments was supported by a non-significant partition homogeneity test (PAUP*: $P > 030$). All sequences were found to be evolving neutrally [HKA: ($0.28 < P < 0.75$), MK: ($0.31 < P < 0.72$)] and in a rate-uniform manner (non-significant Tajima's test). MODELTEST indicated that our markers were evolving under the General Time Reversible (GTR) model of sequence evolution (Tavaré, 1986). Composite haplotypes were 1,869 bp in

length, and yielded 54 haplotypes (Appendix 4) or one haplotype every 16.3 individuals. Three of these (5.5%) spanned both sub-basins and species, and accounted for 141 individuals (42.6% of total individuals). Another six haplotypes (11.1%) spanned upper basin *G. cypha* and *G. robusta*, and involved 18 individuals (5.4%). Five composite haplotypes (8 individuals) represented *G. elegans*. Parameters below were all calculated based on these composite haplotypes.

Comparisons made across the combined 1,869 bp for the 323 *G. cypha* and *G. robusta* individuals (less eight *G. elegans* haplotypes) revealed 1,396 monomorphic sites (74.7%) and 473 polymorphic sites (25.3%). Of the latter, 300 were parsimony-uninformative (16%), whereas 173 were parsimony-informative (9.3%). Overall, haplotype (gene) diversity (*h*) = 0.860 (S.D. = 0.012), while nucleotide diversity (π) = 0.00128 (S.D. = 0.00004). Individual haplotypes differed from one another on average at 2.36 bp (Table 2). Base frequencies were: A = 0.28; G = 0.16; C = 0.26; and T = 0.30.

MtDNA: Molecular divergence at the population and sub-drainage level

Haplotype and sequence divergences at the population, locality and sub-drainage level are provided in Table 2. While comparisons among populations are often difficult due to variance in sample sizes, the lack of haplotype differentiation in Desolation Canyon is conspicuous for both *G. cypha* and *G. robusta*. For each, this site also reflects the lowest k-value (i.e., how much on average two random haplotypes differ). These deficits are again reflected when an overview of Desolation Canyon is performed by combining individuals of both species (less those individuals reflecting *G. elegans* haplotypes) (Table 2).

Populations of *G. cypha* in Grand Canyon seemingly reflect a similarity in haplotype numbers, polymorphic sites, and h/π values (Table 2), particularly given the disparity found in sample sizes. For example, Stevens Aisle Area (SAA) had a k-value of 0.867 whereas Lava-to-Hance reach (LAH) reflected k= 3.18, but both had very low

sample sizes (n=6 and 8, respectively). Of interest is the fact that six (of eight) individuals with *G. elegans* haplotypes were found within Grand Canyon, whereas the other two were from Desolation Canyon (one each in *G. cypha* and *G. robusta*).

Percent sequence divergences were calculated among all populations (Table 3). Individuals with *G. elegans* haplotypes were excluded from these calculations and *G. cypha* and *G. robusta* from Upper Basin localities were considered separately. Simply put, populations differ minimally from one another, regardless of whether they are found in different sub-basins or represent different species. In contrast, sequence divergence between these populations and other species were clearly larger. *Gila elegans*, for example, differs from *G. cypha* or *G. robusta* populations on average at 4.7—4.8% sequence divergence (± 0.0—0.1). Similarly, *P. lucius* differs from *G. cypha* or *G. robusta* populations on average by 5.3—5.4% sequence divergence (± 0.5). In comparison, *G. elegans* and *P. lucius* differ from one another at 6.1% sequence divergence (± 0.5).

The demographic histories of our study populations were explored using Tajima's D-statistic and Fu's Fs (Table 4). These parameters can indicate when a particular area has undergone population expansion. All Grand Canyon populations exhibited no such signal, whereas two (of four) upper basin *G. cypha* populations, and five (of seven) upper basin *G. robusta* populations showed a significant value for Fu's Fs, indicating a recent (i.e., post-Pleistocene) population expansion.

MtDNA: Phylogenetic analyses and haplotype networks

Figure 2 depicts a majority-rule MP consensus tree of all 49 *G. cypha* /*G. robusta* haplotypes (below referred to as in-group), five *G. elegans* and three *P. lucius* haplotypes (referred to as out-groups), and is rooted at *Cyprinus carpio* (Carp). Following *C. carpio*, the next clade in the tree (1) is composed of three *P. lucius* (CPM) haplotypes, which is sister to the remainder of the tree. Clade 2 is composed of five *G. elegans* (BTC) haplotypes, and is sister to the remainder of the tree. Within the in-

group, haplotypes are grouped into two large clades (i.e., 3 and 4) that are relatively ill-defined, particularly clade 3, which is supported by a 58% bootstrap value. Yet, both clades are large, with clade 3 encompassing 24 haplotypes across 182 individuals (56.3%), while clade 4 tallies 19 haplotypes across 123 individuals (38.1%). An additional six haplotypes representing 18 individuals (5.6%) are located at the base of the in-group in a undifferentiated comb.

The 49 *G. cypha* /*G. robusta* haplotypes are arranged in an undirected network (Figure 3) that links them to one another and establishes relational links. The number of each haplotype in Figure 3 corresponds to those found in Appendix 4. Five haplotypes representing *G. elegans* (haplotype #11, 19, 20, 30, and 38) are excluded. The representation of *G. robusta,* upper basin *G. cypha* and Grand Canyon *G. cypha* in the various haplotypes can be determined in Appendix 4, and is also presented as color-coded (grey-scale) haplotype network in Figure 4. The color-coding of each haplotype in the figure is determined by the genealogy of the individuals that comprise it (i.e., upper basin *G. cypha*, upper basin *G. robusta*, and Grand Cayon *G. cypha*). The haplotypes in Figures 3 and 4 can be configured into three groups. The top group radiates from haplotype 2, while the middle group is centered at haplotype 10. The bottom group emanates from haplotype 4. These three groups (top, bottom, middle) correspond to clades 4, 3, and the six unclassified haplotypes at the base of the in-group in Figure 2.

Msat DNA: amplification and scoring

Of the 20 microsatellite loci screened across all populations included in the msat analysis (Table 5), two were eventually excluded from analyses. One (C02) did not consistently yield fragments, while a second (G211) displayed an unscorable allele pattern (Table 6). Null alleles appeared to strongly influence allelic patterns across most populations in two other loci (G03 and C13), and these were excluded as well. This left 16 total loci for analyses.

Total number of alleles at each locus was generally high (Table 6), ranging from 3 (Gel14) to 51 (G99), and averaging 29.3. Mean allelic diversity ranged from 4.9 ($\pm$ 0.4)

in HKAN to 20.1 (± 2.1) in HWWC. The lowest value for populations with more than 10 individuals was 7.8 (± 0.6) in HHAV (Table 7). Populations from Grand Canyon generally showed lower allelic diversity than did populations of similar size from the Upper Basin. The latter likely reflects higher rates of introgression in Upper Basin populations. However, allele frequencies in samples with less than 20 individuals may not accurately reflect the alleles present in these populations.

Most loci revealed multi-modal allele frequency distributions, a pattern partially due to allele size differences characteristic for particular populations or taxa (Appendix 5). For example, locus G99 revealed alleles ranging from 337 to 561 bp, with *G. elegans* exhibiting mostly alleles larger than 500 bp. The appearance of large alleles at this locus for *G. cypha* or *G. robusta* populations is likely due to introgressive hybridization. If the introgression event occurred some time ago, intermediate alleles could have evolved, which appears to be the case.

Msat DNA: preliminary analyses

Tests for linkage disequilibrium between all pairs of loci across all populations were significant for 15 of 120 comparisons (6 expected by chance alone). Some loci were more affected than others, with five significant tests involving locus C01, four each for loci G38, G34 and G02, and two each for loci G39, C03, G87, Gel16, while approximately 0.8 significant tests per locus would be expected by chance alone. However, significant linkage disequilibrium tests are not surprising, since these were performed across various putative taxa. Loci with taxa-specific allele distributions should exhibit distinct patterns, making such loci appear to be linked in pair-wise comparisons.

When each locus pair was tested for each population separately, 19 out of 2,520 tests were significant (126 expected by chance alone). Significant tests were found in populations HHAV and RWWC (1 each), HLCR (2), BTC (6) and HRAN (9), whereas six per population would be expected by chance alone. Given these results, the 16 loci appear not to be linked. The nine significant tests in HRAN were thus surprising, but

may be due to biased sampling (i.e., samples consisting of closely related individuals) or inbreeding (which would cause individuals to be closely related). Indeed, tests for deviation from HW equilibrium run under the alternative null hypothesis of heterozygote deficiency revealed significant results for four out of 16 loci in the HRAN population (whereas 0.8 would be expected by chance alone). Null alleles might also cause hetereozygote deficiency in populations, but significant tests would be concentrated within a particular locus, rather than a population. Three significant tests did involve locus G34, making the presence of a null-allele likely for this locus.

Msat DNA and heterozygosity

Levels of expected heterozygosity for the 16 microsatellite loci were generally high in the 21 populations, ranging on average from 0.222 (Gel14) to 0.873 (C03), with only six populations fixed for a particular allele at one locus (Appendix 6).

After Bonferroni correction (Rice 1989), six out of 336 locus-population comparisons showed significant departure from Hardy-Weinberg equilibrium, while 16.8 would be expected by chance alone (Appendix 6). Significant deviations were found for one locus each in populations HLCR (C03), RC15 (C01), and BBTC (G13), with 0.8 expected by chance alone. Population RYAM showed two significant deviations (G34 and G02), whereas HRAN showed three (G34, C03, and G87).

Testing with the alternative hypothesis of heterozygosity deficiency (a more powerful test than the probability test above) revealed 10 significant observations, whereas 16.5 would be expected by chance alone. Four populations were affected at one locus each: HWWC (G34), RWWC (Gel06), RYAM (G34), and BBTC (G99). In addition HLCR was affected at two loci (C03 and G87), whereas HRAN was affected at four (G34, C03, G87, and G99). This confirmed that that the departure from Hardy-Weinberg equilibrium in HRAN was due to an excess in homozygotes, and further supports the hypothesis of potential inbreeding (or biased sampling) in this population, particularly given its large sample size (N=80; Table 7). No population showed

significant deviations from Hardy-Weinberg equilibrium with the alternative hypothesis of heterozygote excess.

Msat DNA and private alleles

Seventy private alleles (*i.e*., alleles unique to a particular population and not found in any other population examined) were identified across the 21 study populations (Table 8). Grand Canyon populations revealed few private alleles (seven total), suggesting high levels of gene flow and overall genetic similarity among populations. One exception was population HRAN with five private alleles (out of seven total for Grand Canyon), again suggesting a certain level of uniqueness for this population, with little incoming gene flow. Among upper basin *G. cypha* populations, HWWC stood out with 15 private alleles (although HBKR with eight and HDES with seven are impressive as well). Similarly, RYAM stood out among upper basin *G. robusta* populations with 12 private alleles. Interestingly, RWWC had only three private alleles, unlike the situation in HWWC where 15 private alleles were detected.

Msat DNA: Statistical estimation of population differentiation

The pair-wise RxC test (Table 9) detected 8 total populations (i.e., 8 separate gene pools). These were: Gene pool 1 = HBKR and HWWC; Gene pool 2 = HDES and RDES; Gene pool 3 = HYAM; Gene pool 4= H30M; Gene pool 5 = HLCR, HLAH, HSHN, HSAA, HRAN, HKAN, HHAV and HWGC; Gene pool 6 = RBKR, RC15 and RWWC; Gene pool 7 = RYAM, RYDT and RLYC; Gene pool 8 = BBTC. These make intuitive sense, in that *G. cypha* from Black Rocks and Westwater canyons would form a single population, as would *G. cypha* and *G. robusta* from Desolation Canyon. Likewise, all *G. cypha* populations within Grand Canyon grouped together save for the population at 30-mile (upriver from the Little Colorado River). *Gila robusta* from Black Rocks Canyon, Colorado River 15-mile, and Westwater Canyon also formed a single population, as did *G. robusta* from three sites within the Yampa River.

Msat DNA: Clustering estimation of population differentiation

Our multi-locus msat analyses using program BAPS [an analysis based upon individuals rather than geographic locations (= samples)] consistently yielded four clusters regardless of the number of populations specified *a priori*. The composition of these clusters makes intuitive sense: One cluster was represented by *G. elegans*; a second was represented by all individuals from Grand Canyon; a third was represented by individuals from upper basin Desolation Canyon (both species)*;* the fourth contained all remaining upper basin G. *cypha* and *G. robusta*. Figure 5 represents the 4-cluster configuration, with cluster names atop the figure and populations (= locations) below the figure. Upper basin *G. cypha* populations (dark blue) are relatively admixed, particularly when compared with Grand Canyon. Upper basin *G. robusta* exhibit less admixture. Of interest is the fact that the greatest admixture within Grand Canyon populations are those at the upstream and downstream ends of the canyon, i.e., 30-mile (H30M) in Marble Canyon and Western Grand Canyon (HWGC). Also of interest is some individuals within the Little Colorado River (HLCR) population show reasonable admixture with Desolation Canyon (DES; species unknown), whereas the same situation occurs at Randy's Rock (RAN), but with greatest admixture stemming from the more generic 'upper basin' [either *G. cypha* (UB-H), *G. robusta* (UB-R), or both].

In contrast, program STRUCTURE had difficulty in fixing upon a given k-value, alternately selecting either K=5 or K=6. This dilemma is underscored diagrammatically in Figure 6, which depicts the optimal number of clusters as derived using *ΔK* (Evanno *et al.* 2005). Clearly the *ΔK* is rather uniform for either configuration (K=5 or K=6). Table 10 depicts F_{ST} values calculated for K=4 through K=6 solutions. Of interest is the fact that *G. cypha* in the upper basin shows a consistently low F_{ST} (i.e., < 0.018), demonstrating one reason why the software cannot fix upon a precise solution. This F_{ST} level (discussed below) is simply far too weak of a signal to adequately differentiate this cluster.

The five-cluster analysis is depicted in Figure 7, and yielded the following groups: (1) *G. elegans*; (2) Grand Canyon; (3) Desolation Canyon (both species)*;* (4) upper

basin *G. cypha;* and (5) *G. robusta*. This figure again depicts the relative unanimity of the Grand Canyon aggregates. However, as in the K=4 solution, there is considerable admixture between upper basin *G. cypha* and *G. robusta*, with a tendency for the latter species to admix within the former. *Gila robusta* in the Yampa River is grouped as a distinct cluster, but shows considerable admixture with other upper basin *G. robusta* populations. Although *G. cypha* and *G. robusta* in Desolation Canyon reveal levels of admixture, they are still grouped with the very little admixed *G. elegans.*

The K=6 solution depicted by program STRUCTURE (Figure 8) differs from the 5-group solution in that *G. elegans* is now separated from Desolation Canyon. Again, the Grand Canyon populations reflect the least admixture of all, and this is demonstrated in the proportion admixed scores presented for the K=5 and K=6 solutions in Table 11. Nevertheless, we select the K=6 solution to perform our AMOVA analyses (below).

Msat DNA and statistical differentiation among clusters

Small sample sizes (particularly for Grand Canyon populations other than HLCR and HRAN) rendered inappropriate the calculation of divergence between all pairs of populations. Instead, populations were grouped into clusters, as suggested from our STRUCTURE analysis (Table 12). Further, allele frequencies showed multi-modal distributions, suggesting potential non-stepwise mutations. Consequently, non-SMM-based estimators were applied, which merely underestimate population divergence in those cases where SMM processes were in fact responsible in shaping variation (Slatkin 1995). They represent a more conservative approach when SMM-based models may be inappropriate.

AMOVA (Table 13) revealed that most genetic variation was found within populations (94.4%), rather than among populations within clusters (0.93%), or among clusters (4.16%). These values are significant (per Table 13).

Similarly, between-cluster F_{ST} values were relatively shallow, particularly for microsatellite data, although all differed significantly from zero (Table 14). The Grand

Canyon cluster (GC-H) was more differentiated from Upper-Basin *G. cypha* (UB-H), than the latter was from Upper-Basin *G. robusta* (UB-R and YR-R). Based on F_{ST} analysis, *G. elegans* is most divergent from all other clusters, including *G. cypha* and *G. robusta* from Desolation Canyon (cluster DC). *Gila elegans* groups with Desolation Canyon fishes when K < 6 clusters are considered (i.e., K=4 or K=5). These conflicting results suggest a high level of introgression by *G. elegans* alleles into Desolation Canyon individuals, and is recorded differently by clustering *vs* population-differentiation analyses, such as the ones performed herein. *Gila robusta* from the Upper Basin (including Yampa River) are more differentiated from Desolation Canyon individuals than they are from upper basin *G. cypha*, suggesting a much lower influence of *G. elegans* alleles in *G. robusta*. It is likely that pure *G. robusta* are phenotypically much easier to distinguish from *G. elegans* than are *G. cypha* from *G. elegans*. Expressing divergence among populations as average pair-wise differences again revealed overall genetic distinctiveness, albeit this degree of separation is shallow, particularly when corrected for within-cluster differences (Table 15).

DISCUSSION

Gila and the problem of interspecific hybridization

Gila in the Colorado River ecosystem (and particularly forms in the Upper Basin), have always been thought to comprise numerous hybrids. Valdez & Clemmer (1982), for example, suggested that apparent hybridization between *G. cypha* and *G. robusta* was due to recent human-constructed impoundments that have diverted and altered the flow of the river. In this sense, historic flows were perceived as an important reproductive isolating mechanism for native fishes (also discussed by Douglas & Douglas 2000) now seriously altered by water development projects. Conventional wisdom was that diminution of peak flows and the disturbance of the natural hydrograph, concomitant with temperature and sediment alterations (particularly manifested in the Colorado River of Grand Canyon) have conspired to alter the niches of native fishes such that they now overlap, especially during critical reproductive periods (see Karp & Tyus 1990, Kaeding *et al*. 1990). This was particularly evident with the species that comprised the *G. robusta* complex (i.e., *G. robusta, G. cypha,* and *G. elegans*).

Traditionally, species diagnosis was based on morphology and it is here where suspected interbreeding amongst *Gila* species caused much confusion. Particularly hybridization between *G. elegans* and *G. cypha* was deemed problematic. When these two species were captured together, fishery managers often considered them phenotypically indistinguishable. However, when two species were quantitatively evaluated, a high degree of within-group homogeneity was demonstrated (Holden & Stalnaker 1970; Douglas *et al*. 1998). The inability of researchers to adequately diagnose *G. elegans*, especially when taken with *G. cypha,* led early fishery biologists to simply lump the two under the generic label “Bonytail” (reviewed by Holden 1991; Douglas *et al*. 1998). This inability to discriminate at the time of capture led to considerable confusion decades later, when distributions and numbers of these fishes (as recorded in the early pre-impoundment collections) were evaluated and contrasted.

Given the concerns with regard to hybridization in *Gila* (as above), considerable emphasis was placed early on to determine interspecific relationships among the species using morphology (i.e., were the species actually distinct, and could they be sufficiently discriminated so as to be classified?). Although hump and snout characteristics are instrumental in discriminating *G. cypha* and *G. elegans* (Douglas *et al*. 1998), these traits are often less diagnostic in younger and smaller fishes. For example, Douglas *et al*. (1998) could only delineate 60% and 66% of juvenile *G. cypha* and *G. elegans,* respectively. They deemed their analyses as inconclusive and could not reject the hypothesis that adult characteristics diagnose juveniles.

While numerous morphological analyses have been performed over the years, the most innovative techniques were implemented by McElroy & Douglas (1995) and Douglas *et al*. (2001). McElroy & Douglas (1995) examined 363 specimens (215 *G. robusta* and 148 *G. cypha*) from seven upper basin locations and the Little Colorado River in Grand Canyon in a geometric morphometric analysis. At the population level, all samples differed significantly from one another, and there was no relationship between morphological similarity and geographic proximity. This suggests the populations are distinct, and each is on its own evolutionary trajectory. At the species level, both *G. cypha* and *G. robusta* were clearly separated from one another, whether they were in sympatry or allopatry (a conclusion underscored by other researchers as well).

McElroy & Douglas (1995) and Douglas *et al*. (2001) found that although *G. robusta* and *G. cypha* are morphologically distinct, each species phenotypically collapsed towards the other at two locations, Desolation Canyon and Cataract Canyon (Utah: Figure 1). At these sites, a ‘locality effect’ contributed significantly to the inter-relationships among populations, such that *G. robusta* and *G. cypha* resembled one another more closely than either did with conspecifics elsewhere in the basin (where a ‘species effect’ dominated). However, it should be noted that *G. robusta* and *G. cypha* could, in fact, be discriminated from one another at Desolation and Cataract canyons, even though they did not cluster with conspecifics in a global analysis. McElroy &

Douglas (1995) offered three hypotheses to explain this phenomenon: (1) extensive introgressive hybridization; (2) similar selective pressures on distinct, but sympatric populations; and (3) retention of a high proportion of ancestral characteristics at both locations, such that their perceived morphological similarities may instead reflect shared, ancestral traits. McElroy & Douglas (1995) went on to point out that complementary genetic studies might be useful in evaluating these alternatives. Our analyses consistently grouped *G. cypha* and *G. robusta* from Desolation Canyon into a distinct cluster (Figures 5 and 6), confirming findings of morphological analyses.

Hybridization or introgression clearly poses problems for management and conservation of populations, specifically if supplementation must be implemented to mitigate declines or extirpation of natural populations (Ryman 1991, Leary *et al.* 1995). Usually, the conservation goal is to sort out this genetic mixing so as to maintain distinctness while assuring viable populations for the future. However, hybridization might also be considered as an important element of the current and historical evolution of a group. Cases are known where the potential to hybridize might result in increased population viability (Vrijenhoek *et al.* 1996). Introgressive hybridization was previously found among *G. cypha*, *G. robusta,* and *G. elegans* (Dowling & DeMarais 1993), and this further suggests a long history of potential hybridization within the genus.

In our mtDNA analyses, only eight *G. elegans* haplotypes (i.e., 2.4% of total) were detected in individuals phenotypically identified as *G. cypha* or *G. robusta*. In contrast, in our genetic assessment of 312 Flannelmouth Sucker (*Catostomus latipinnis*) from Grand Canyon, 5.6% of the individuals showed introgressed haplotypes with Razorback Sucker (*Xyrauchen texanus*) (Douglas *et al.* 2000, 2001). The geographic distribution of introgressed *G. cypha* and *G. robusta* (i.e., those showing *G. elegans* haplotypes) is also of interest here. Only two (of eight) were from the Upper Basin, and both were found in Desolation Canyon (one in *G. cypha* and the second in *G. robusta*). The other six were from Grand Canyon (two from 30-mile, two from the Little Colorado River, and two from Randy's Rock area). However, when these eight individuals were evaluated using the 16 msat loci, only a single individual demonstrated a characteristic *G. elegans* nuclear

genotype (a specimen phenotypically identified as *G. robusta* from Desolation Canyon). This in turn suggests the hybridization event(s) that produced the eight *G. elegans* haplotypes in *G. robusta* and *G. cypha* occurred in deep history, and subsequent backcrossing occurred with conspecifics such that introgressed nuclear alleles are maintained in the population, albeit at reduced to low numbers, and thus were not detected among the 16 microsatellite loci assessed in this study. Gerber et al. (2001) came to a similar conclusion based on their basin-wide mtDNA survey of *Gila* species. These authors concluded that levels of gene exchange were too extensive to be attributed to recent, human-induced perturbations, a notion our results support.

Population differentiation using mtDNA

Assessment of intraspecific relationships using mtDNA can often be hampered by lack of variation (Goldstein *et al.* 1995). This is because mtDNA lacks the resolution to diagnose recent (i.e., post-Pleistocene) events (Avise *et al.* 1984, Brunner *et al.* 2001). Unfortunately, much of the population subdivision of interest to managers has occurred during the intervening temporal span. We find this to be true in this study as well, in spite of the fact that we have employed three of the more rapidly evolving mtDNA genes and regions (Kumar 1996). This lack of differentiation, manifested at single or combined mtDNA regions is revealed repeatedly in haplotype and maximum parsimony trees (Figs. 2, 3 and 4) and in percent sequence divergence tables (Table 3). Populations of *G. cypha* and *G. robusta* appear to be linked by 'ancestral polymorphism', which represents variation that was originally present in the ancestral population (presumably the one represented by the collapsed and hybridized population at end-of-Pleistocene). This variation now remains and is retained within the current populations of both species. Not enough evolutionary time has elapsed for these populations to sort out, even though they may now be geographically isolated. Thus, ancestral polymorphism often links populations together even though they are geographically isolated. 'Lineage sorting' then changes the frequency of ancestral alleles within each population over time. Unique mutations occur within isolated populations and these move to fixation. Ultimately, over time, ancestral polymorphism diminishes and the populations become more and more distinct in their molecular characterization and their unique mutations

sort. In this study, shared or similar haplotypes link populations within and across basins. In our dataset, the Upper Basin has more unique haplotypes, but also a higher frequency of haplotypes shared among populations. Grand Canyon populations reflect fewer unique haplotypes, but they also display a higher frequency of basin-specific haplotypes.

The failure of mtDNA to discriminate populations in this study is even more intriguing when one considers that the level of differentiation among populations of *G. cypha* is identical to the level of differentiation they show when contrasted against populations of *G. robusta*. The two are recognized as separate biological species, yet neither can be separated from one another at the mtDNA level. Of additional interest is the fact that this extensive admixture includes *G. cypha* and *G. robusta*, but not *G. elegans* (albeit low levels of introgression by *G. elegans* into *G. cypha* as well as *G. robusta* have been detected). Based upon phenotypic similarities, the *G. elegans* has been historically interpreted as a close sister-species to *G. cypha*, yet it differs from *G. cypha* / *G. robusta* at 4.7—4.8% sequence divergence (± 0.0—0.1). To place this value into context, *G. elegans* is separated from *P. lucius* at 6.1% (± 0.5).

One hypothesis that explains this complex evolutionary situation is that *G. cypha* and *G. robusta* populations were reduced to very low abundances and subsequently forced together into refugial habitats by an end-of-Pleistocene warming event (Betancourt 2004). While so reduced and syntopic, they broadly hybridized, possibly repeatedly over a temporal span. Eventually, pluvial conditions returned and the aquatic environment subsequently expanded again. Both species then returned to familiar and exclusive niches, and to reproduce again with like kind. Nevertheless, the large-scale population reduction both incurred would have re-set their mtDNA evolutionary clocks, and the resulting hybridization event(s), combined with founder-events would have provided the same mtDNA haplotype(s) for both species, a scenario also supported by Gerber et al. (2001). Two (of four) *G. cypha* populations in the upper basin and all seven *G. robusta* populations revealed a signature of post-Pleistocene population expansion, whereas none of the nine Grand Canyon aggregates reflected such a signal.

Douglas & Douglas (2003) explained results from basin-wide analysis of Flannelmouth Sucker (*Catostomus latipinnis*) in much the same context as above. These researchers were surprised to uncover patterns of low genetic diversity within a reasonably abundant and widespread native species. Their analyses also revealed a significant and positive geographic cline ($P < 0.001$) in nucleotide diversity as one moved upstream in the Colorado River from the southwest (circa Lake Powell) to the northeast. This cline divided the Colorado River basin into three reaches: lower (i.e., Virgin River/ lower Grand Canyon); upper (i.e., Yampa/ upper Green rivers), and middle (i.e., mid--Grand Canyon through lower Green and upper Colorado rivers). Populations from the three reaches were estimated to have diverged from one another 3,400--11,000 y.b.p. and haplotype distribution suggests that populations are expanding in the upper Colorado River.

Douglas & Douglas (2003) suggested that the lack of genetic variation and the recent coalescence of lineages in Colorado River basin *C. latipinnis* were unusual, especially given (a) its fossil history, (b) the broad geographic sampling conducted in the study, and (c) the number (and evolutionary rate) of the mtDNA genes examined. These researchers argued that the most parsimonious explanation for these data was a rapid expansion in *C. latipinnis* following a period of low effective population size at end-of-Pleistocene. In other words, the species experienced a severe population bottleneck (seemingly at the Pleistocene-Holocene interface), then rebounded and expanded northward into the basin following this event.

Our evaluations of mtDNA variation in *G. cypha* from populations throughout its range revealed relatively high haplotype diversity and low levels of nucleotide diversity (Table 2). These data are shallow in their depth (as above), and essentially reveal a plethora of haplotypes separated by but a few base pair differences (see Figs. 3, 4). Grant and Bowen (1998) explored the problem of shallow genetic architecture in marine fishes by grouping the latter into four classifications based upon different combinations of haplotype (h) and nucleotide diversity (π). These categories are defined by demographic events that alter the likelihood of mtDNA lineage survival and the time to

ancestral coalescence of lineages. The first category (and the most important for this discussion) included species with small values for both parameters (i.e., $h < 0.5$; $\pi = 0.5\%$). Similar low values are represented in Table 3 (below) and may result either from recent founder effects, region-wide bottlenecks, or metapopulation structure.

Also of interest is the lack of haplotype differentiation in Desolation Canyon. This site reflects the lowest k-value (i.e., how much on average two random haplotypes differ) for both *G. cypha* and *G. robusta*. These deficits are again demonstrated when individuals of both species are combined (but not including the two individuals reflecting *G. elegans* haplotypes).

Population differentiation using msat DNA

Our mtDNA data were unable to resolve populations, subdrainages (upper basin vs Grand Canyon), or even species, and we hypothesized that this may be due to the short temporal duration that has elapsed since the proposed "collapse/ admixture" event at end-of-Pleistocene. This length of time was simply of too short for ancestral polymorphism to be superseded, and for lineages to adequately sort out. However, msat DNA did provide sufficient resolution to discriminate among populations and basins (see below), and this capability is congruent with similar results obtained in other studies using these markers (Brunner *et al.* 1998, Douglas *et al.* 1999b).

However, one difficulty with our msat DNA analysis is that upper basin *G. cypha* and *G. robusta* are too similar to one another to adequately differentiate them. This situation is depicted in Figure 5, which shows admixture in upper basin *Gila*, particularly *G. cypha*. Evidence is provided in Table 10, which depicts the F_{ST} values depicted for each cluster. Note there are several that fall between 0.01 and 0.02. Latch *et al*. (2006) evaluated both STRUCTURE and BAPS with regard to the diagnosis of clusters within simulated msat data (which were much less admixed than ours), and concluded that both programs were skilled in discerning the number of clusters, as long as certain conditions were met. For example, STRUCTURE could not detect populations when $F_{ST} = 0.01$, and could only designate four (of five) at $F_{ST} = 0.02$. However, it did well at F_{ST}

levels = 0.03 or above. Similarly, Latch *et al*. (2006) found that *Δk* (Evanno et al. 2006) worked well at F_{ST} levels = 0.03 or above, but when F_{ST} levels fell below this, the statistic did not exhibit a clear trend, suggesting that a clear pattern of genetic structure was not evident in the data. This is what emerges from our analyses as well.

With regard to program BAPS, Latch *et al*. (2006) found that, again, F_{ST} levels played a large part in delineating clusters. When F_{ST} = 0.02, most individuals were grouped into two to four main clusters, indicating an overall inability to correctly delineate subpopulations. When F_{ST} levels reached 0.03 (or greater), BAPS correctly inferred the true value of K. Table 10 demonstrates this lack of resolution within our data, in that *G. cypha* from the upper basin consistently reflects an F_{ST} value < 0.018. As above, this is simply below the threshold for adequate detection, and adds to the confusion with regard to designating the number of groups present in the data.

Waples and Gaggioti (2006) found STRUCTURE to be reliable in estimating the true number of populations only when gene flow was relatively low (number of migrants Nm = 5), and with large sample sizes and highly polymorphic loci (i.e., ideal conditions). Performance was much worse when sample size or number of loci were reduced, and when gene flow was elevated (Nm >= 25). These researchers also noted that the alternative method of estimating number of populations argued for by Evanno et al. (2005) was not an improvement over the usual approach found in STRUCTURE. Both performed well when genetic divergence was strong (Nm = 1), and the standard method was better than the alternative when gene flow was moderate (Nm = 5).

The Bayesian approach to determining population numbers (i.e., clustering groups of individuals) proved to be very conservative (Waples and Gaggioti, 2006), in that different gene pools could only be detected under very restrictive migration (Nm = 1). However, multilocus contingency tests (such as that in TFPGA) proved quite powerful in estimating numbers of populations, even with reduced samples of individuals and loci (Table 9). In our sample, TFPGA detected eight gene pools, but two consisted of small populations (HYAM and H30M), both revealing high levels of

admixture. The latter is reflected by unclear assignment of either population in Bayesian clustering estimates (Figures 7 and 8).

Given these remarks, we elect to take a conservative approach to delineating subgroups based on our msat data. We thus fix upon the partitions in Figure 8 (the six-group configuration from software STRUCTURE). We select K=6 because the F_{ST} value is largest there for upper basin *G. cypha* (although still < 0.02). This solution also delineates *G. elegans* as a distinct entity, a consideration that is supported by its high F_{ST} value (in fact, the highest of all three K-value analyses at 0.1579; Table 10). The six-cluster solution is: (1) *G. elegans*; (2) all Grand Canyon *G. cypha* 'aggregates'; (3) Desolation Canyon *G. cypha and G. robusta*; (4) upper basin G*. cypha* (excluding *G. cypha* from Desolation Canyon); (5) upper basin *G. robusta* (excluding *G. robusta* from Desolation Canyon and Yampa River) *;* and (6) upper basin *G. robusta* from the Yampa River.

Msat DNA and the estimation of population differentiation within Marble/ Grand canyons

Again, at the northern and southern extremities of the distribution of *G. cypha* in Grand Canyon are the populations with greatest admixture overall. Clearly, the Grand Canyon aggregates are linked via gene flow, and this had kept them relatively homogeneous amongst themselves and has contributed to the high scores they reveal in the admixture evaluation. It would also make biological sense that the extremities of the distribution show the greatest flux. The 30-mile population of Marble Canyon (H30M) is relatively isolated from the remainder of Grand Canyon, and consistently shows an upper basin influence that borders on 45—50%. Similarly, a gradual diminution of gene flow from the Little Colorado River would be manifested within Western Grand Canyon (HWGC), and the relatively small populations upstream of this reach would not be able to muster enough reproduction to allow for downstream drift of larvae and/ or juveniles. Thus, individuals within the Western Grand Canyons reach (HWGC) probably do not represent a random sample of genetic diversity within Grand Canyon, and, given the supposed lack of gene flow from upriver, would be expected to reflect a genetic composition that is more similar to the ancestral Grand Canyon

populations. Interestingly, our sample representing HWGC contained several juveniles sampled from a backwater at rivermile 193, potentially suggesting a rare event of local reproduction.

A second point to consider in this regard is that both H30M and HWGC reflect very small sample sizes (eleven and ten, respectively). Given this, there is little statistical basis for making any claims with regard to supposed uniqueness. There are indeed trends, particularly with H30M, but little else can be said.

Historic versus contemporary geneflow

Extent of admixture between *G. cypha* and *G. robusta*, as well as occasional involvement of *G. elegans*, revealed by our mtDNA and msat analyses suggests an historic event, likely occurring during late-Pleistocene population bottlenecks and subsequent recolonization events. However, these signals would mask more recent events of introgressive hybridization and prevent their detection. Thus, it cannot be concluded whether human-induced habitat perturbations affected reproductive behavior of these species. In light of strong evidence that these events occurred naturally in these populations, as well as similar observations in other fishes of the arid southwest and potential significance of such introgressive hybridization for evolution and diversification in these groups (as discussed in Gerber *et al.* 2001), admixture of big river *Gila* species can be considered natural and not detrimental to the long-term survival of the species. Other factors (e.g., habitat alterations, predation and competition by introduced non-natives) should be of greater concern.

Overview with regard to management units

We recognize difficulty with defining conservation units among native fishes of the Colorado River ecosystem. These questions have consequence beyond mere taxonomy and evolutionary biology, long dubbed ‘academic exercises,’ and instead impinge upon the manner in which we will adaptively manage the Colorado River ecosystem. Here, sound knowledge about species identity and distinctiveness is needed. In spite of the

complicated evolutionary history exhibited by these species, they are on their own evolutionary trajectories.

Gila cypha, G. robusta, and *G. elegans* have long been recognized as distinct from one another. Vanicek & Kramer (1969) noted length/ weight differences and growth characteristics that defined *G. robusta* as separate from *G. elegans*, and argued these factors clearly delimited the two as separate species. Holden & Stalnaker (1970) likewise felt that *G. robusta* and *G. elegans* were distinct morphologically and represented separate evolutionary lineages. Holden & Stalnaker (1970) were more ambivalent with regard to *G. cypha*. Yet, Smith *et al*. (1979) were able to statistically differentiate all three species, as did Douglas *et al*. (1998). Other researchers found clear morphological distinctions between *G. cypha* and *G. robusta*, when this species-pair was contrasted (*e.g*., Douglas *et al.* 1989, Kaeding *et al*. 1990, McElroy & Douglas 1995, Douglas *et al*. 2001).

A classic model for the current situation resides within *G. seminuda* (DeMarais *et al*. 1992), a hybrid species that has arisen by introgressive hybridization between *G. elegans* and *G. robusta*. This example serves as a reflection of the current situation, and in fact, we hypothesize that other fish species in the Colorado River ecosystem may have been experiencing levels of introgressive hybridization that could lead to speciation.

How much divergence is required to deduce if populations are distinct enough for protection? Conservation genetics (sensu Vrijenhoek *et al.* 1986) has a role in identifying cryptic variation towards which management action can be directed (as per Avise 1989, Quattro *et al.* 1996). Many have argued that genetically divergent populations are appropriate for conservation regardless of their taxonomic status (Moritz 1994a). Because existing taxonomies may not adequately reflect the extent of genetic diversity within lineages, the concept of “evolutionarily significant unit” (ESU) was developed (Moritz 1994a). However, ESUs are concerned with historical population structure as reflected in a molecular phylogeny, and serve long-term conservation needs. By definition, ESUs should be reciprocally monophyletic for mtDNA alleles, and show significant divergence in allele frequencies at nuclear loci (Moritz 1994b). Our situation does not reflect the reciprocal

monophyly of mtDNA sequences, but populations deviate from one another at the msat level.

Conservation of genetic variation at the local level was not initially addressed by the ESU concept. To correct this, Moritz (1994a,b) stressed the importance of distinguishing a second type of conservation unit, also using genetic criteria. The “management unit” (MU) is concerned with current (rather than historical) population structure, is based on allele frequencies (rather than phylogenies), and is applied to short- (rather than long-term) management issues. They are important, for example, in recognizing situations were there has been rapid speciation, and where mtDNA lineages may not have sorted between otherwise discrete taxa (as per Douglas *et al.* 1999b). Thus, MUs recognize the dynamic aspects of intraspecific genetic variation, and provide a category by which these entities can be defined, conserved and managed. In the present situation, Grand Canyon ‘aggregates’ of *G. cypha* represent an MU, as does the Desolation Canyon mix of *G. cypha* and *G. robusta*. Similarly, upper basin *G. cypha* (less Desolation Canyon) would also be considered an MU, as would upper basin *G. robusta* from Yampa River. In addition, the remaining populations of *G. robusta* from the upper basin (less Desolation Canyon and Yampa River) would also be distinct.

CONCLUSIONS

Conclusions 1: Our data support the recognition of six 'Management Units' (MUs). The Marble/ Grand canyon aggregates of *G. cypha* would represent a distinct MU, as would Desolation Canyon *G. cypha* and *G. robusta* (as a group). Another MU would consist of *G. robusta* from the Yampa River. In addition, *G. robusta* from upper basin locations (save Desolation Canyon and Yampa River) could be recognized as an MU, as could *G. cypha* from upper basin locations (save Desolation Canyon). In addition, Bonytail would represent the sixth MU.

Conclusions 2: Marble and Grand canyon 'aggregates' of *G. cypha* were not overly distinct at the msat level. Aggregates appeared to be connected by geneflow, suggesting downstream drift of larvae and juveniles as a likely scenario. The Little Colorado River population would be the primary source, but contribution from occasional local reproduction by mainstem aggregates cannot be excluded.

Conclusions 3: The *G. cypha* population at 30-mile in Marble Canyon was recorded as having two individuals with *G. elegans* haplotypes, and the msat profile for this population is intermediate between genotypes found in Desolation Canyon (a hypothesized hybrid population) and Grand Canyon. Although reproduction has been documented for the 30-mile population, it suffers from chronic low numbers (at least chronic low numbers of catchable fish). However, this is the only population in Grand Canyon that is upstream from the Little Colorado River and is least likely to receive migrants from downstream locations. Due to its potential distinctiveness it should be further studied.

Conclusions 4: Upper basin *G. robusta* are seemingly more distinct than are upper basin *G. cypha*, and *G. robusta* from the Yampa River fall out as distinct in the five- and six-cluster models. Only the four-cluster model groups upper basin *G. robusta*

with upper basin *G. cypha.* While *G. robusta* in the Yampa River certainly appears distinct and potentially unique, our data are inconclusive about status of Yampa River *G. cypha*. Additional studies must be conducted on *G. cypha* from the Yampa River, in that only seven individuals were available to us. Such chronic low numbers preclude any statistically valid conclusions. Management decision should not be based on the assumption of Yampa River *G. cypha* being distinct, simply because this is neither supported nor rejected by our data.

Conclusions 5: Our last conclusion centers on *G. elegans*. The individuals we evaluated in this study were hatchery-derived, but in some analyses, they were grouped with the unique *G. cypha/ G. robusta* populations of Desolation Canyon. We identified a single individual from this population that retained both a mtDNA haplotpye and a msat DNA genotype of *G. elegans*. Given the propensity for Colorado River native fishes to hybridize throughout their long history, it is not surprising that alleles characteristic for *G. elegans* apparently persist in wild populations of *G. cypha* and *G. robusta*. Similarly, alleles characteristic for the latter two species were occasionally detected in the *G. elegans* we examined. Albeit historic, this naturally occurring admixture among the Colorado River *Gila* species should be taken into account when *G. elegans* are reintroduced into the wild, particularly in locations where the other two species still persist.

LITERATURE CITED

Angers B, Estoup A, Jarne P (2000) Msat size homoplasy, SSCP, and population structure: A case study in the freshwater snail *Bulinus truncates*. *Molecular Biology and Evolution*, **17**, 1926—1932.

Ardren WR, Borer S, Thrower F, Joyce JE, Kapuscinski AR (1999) Inheritance of 12 msat loci in *Oncorhynchus mykiss*. *Journal of Heredity*, **90**, 529—536.

Ashley MV, Dow BD (1994) The use of msat analysis in population biology: Background, methods and potential applications, p. 185-201. In: *Molecular Ecology and Evolution: Approaches and Applications* (Schierwater B, Streit B, Wagner GP, DeSalle R, eds.). Birkhäuser Verlag, Basel, Switzerland.

Avise JC (1989) A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution*, **4**, 279—281.

Avise JC (1994) *Molecular Markers, Natural History, and Evolution*. Chapman and Hall publishers, New York, NY.

Avise JC (2000) *Phylogeography, the History and Formation of Species*. Harvard University Press, Cambridge, MA.

Avise JC, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *Journal of Molecular Evolution*, **20**, 99—105.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with msat markers. *Molecular Ecology*, **11**, 155—165.

Bentzen P, Taggart CT, Ruzzante DE, Cook D (1996) Msat polymorphism and the population structure of Atlantic cod (*Gadus morhua*) in the northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 2706—2721.

Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology*, **13**, 551—561.

Betancourt JL (2004) Arid Lands Paleobiogeography: The Rodent Midden Record in the Americas, p. 27–65. In: *Frontiers in Biogeography* (Lomolino MV, Heaney LR, eds.), Sinauer Associates, Sunderland, MA.

Bezzerides N, Bestgen KR (2002) Status of Roundtail Chub *Gila robusta*, Flannelmouth Sucker *Catostomus latipinnis*, and Bluehead Sucker *Catostomus discobolus* in the Colorado River basin. *Report to the U. S. Bureau of Reclamation, Salt Lake City, Utah*. 139 p. *Larval Laboratory Contribution* 118.

Blinn DW, Cole GA (1991) Algal and invertebrate biota in the Colorado River: Comparison of pre- and post-dam conditions, p. 102—123. In: *Colorado River Ecology and Dam Management* (Committee, eds.) National Academy Press, Washington DC.

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Science USA*, **76**, 1967—1971.

Brunner PC, Douglas MR, Bernatchez L (1998) Msat and mitochondrial DNA assessment of population structure and stocking effects in Arctic charr *Salvelinus alpinus* (Teleostei: Salmonidae) from central Alpine lakes. *Molecular Ecology*, **7**, 209—223.

Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L (2001) Holarctic phylogeography of Arctic Charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, **55**, 573—586.

Burke T (1994) Spots before the eyes: molecular ecology. *Trends in Ecology and Evolution*, **9**, 355—357.

Carothers SW, Brown BT (1991) *The Colorado River through Grand Canyon: Natural history and human change*. University of Arizona Press, Tucson, AZ.

Caughley G (1994) Directions in conservation biology. *Journal of Animal Ecology,* **63**, 215—244.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **32**, 550—570.

Cavender TM, Coburn MM (1992) Phylogenetic relationships of North American cyprinids. p. 293—327. In: *Systematics, Historical Ecology, and North American Freshwater Fishes* (Mayden RL, ed.). Stanford University Press, Stanford, CA.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657—1659.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology,* **15**, Online publication date: 29-Jun-2006.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, **15**, 290—295.

DeMarais BD (1992) Genetic relationships among fishes allied to the genus *Gila* (Teleostei: Cyprinidae) from the American Southwest. *Unpublished Ph.D. dissertation*, Arizona State University, Tempe AZ.

DeMarais BD, Dowling TE, Douglas ME, Minckley WL, Marsh PC (1992) Origin of *Gila seminuda* (Teleostei: Cyprinidae) through introgressive hybridization: Implications for evolution and conservation. *Proceedings of the National Academy of Science USA*, **89**, 2747—2751.

Dieringer D, Schlötterer C (2003) MSAT ANALYZER (MSA): A platform independent analysis tool for large msat data sets. *Molecular Ecology Notes,* **3**, 165—167.

Dimsoski P, Toth GP, Bagley MJ (2000) Microsatellite characterization in Central Stoneroller *Campostoma anomalum* (Pisces: Cyprinidae). *Molecular Ecology*, **9**, 2155—2234.

Douglas ME (1993) An analysis of sexual dimorphism in an endangered cyprinid fish (*Gila cypha* Miller) using video image technology. *Copeia*, **1993**, 334—343.

Douglas ME, Douglas MR, Lynch JM, McElroy DM (2001) Use of geometric morphometrics to differentiate *Gila* (Cyprinidae) within the Upper Colorado River basin. *Copeia*, **2001**, 389—400.

Douglas ME, Douglas MR, Schuett GW, Porras LW, Holycross AT (2002) Phylogeography of the Western Rattlesnake (*Crotalus viridis*) complex, with emphasis on the Colorado Plateau, p. 11-50. In: *Biology of the Vipers* (Schuett GW, Hōggren M, Douglas ME, Greene HW, eds.). Eagle Mountain Publishing, LC, Eagle Mountain, UT.

Douglas ME, Marsh PC (1996) Population estimates/ population movements of *Gila cypha*, an endangered cyprinid fish in the Grand Canyon region of Arizona. *Copeia,* **1996**, 15—28.

Douglas ME, Marsh PC (1998) Population and survival estimates for *Catostomus latipinnis* in northern Grand Canyon, with distribution and abundance of hybrids with *Xyrauchen texanus*. *Copeia*, **1998**, 915—925.

Douglas ME, Miller RR, Minckley WL (1998) Multivariate discrimination of Colorado Plateau *Gila spp*.: The "art of seeing well revisited." T*ransactions of the American Fisheries Society,* **127**, 163—173.

Douglas ME, Minckley WL, Tyus HM (1989) Qualitative characters, identification of Colorado River chubs (Cyprinidae: genus *Gila*) and the "art of seeing well." *Copeia*, **1989**, 653—662.

Douglas ME, Minckley WL, DeMarais BD (1999a) Did vicariance mold phenotypes of western North American fishes? Evidence from Gila River cyprinids. *Evolution*, **53**, 238—246.

Douglas ME, Douglas MR, Schuett GW, Porras LW (2006) Evolution of Rattlesnakes (Viperidae: *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. *Molecular Ecology*, **15**, 3353—3374.

Douglas MR, Brunner PC (2002) Biodiversity of Central Alpine *Coregonus* (Salmoniformes): Impact of one-hundred years of management. *Ecological Applications,* **12**, 154—172.

Douglas MR, Brunner PC, Bernatchez L (1999b). Do assemblages of *Coregonus* (Teleostei: Salmoniformes) in the Central Alpine region of Europe represent species flocks? *Molecular Ecology*, **8**, 589—604.

Douglas MR, Douglas ME (2000) Late season reproduction by big river Catostomidae in Grand Canyon. *Copeia*, **2000**, 238—244.

Douglas MR, Douglas ME, Brunner PC (2003) Drought in an evolutionary context: Molecular variability in Flannelmouth Sucker (*Catostomus latipinnis*) from the Colorado River Basin of Western North America. *Freshwater Biology*, **48**, 1254—1273.

Dowling TE, DeMarais BD (1993) Evolutionary significance of introgressive hybridization in cyprinid fishes. *Nature*, **362**, 444—446.

Estoup A, Largiader CR, Cornuet JM, Gharbi K, Presa P, Guyomard R (2000) Juxtaposed msat systems as diagnostic markers for admixture: An empirical evaluation with brown trout (*Salmo trutta*) as model organism. *Molecular Ecology*, **9**, 1873—1886.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, **14**, 2611—2620.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN VER. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47—50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479—491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics,* **164**, 1567—1587.

Farris JS, Källersjö M, Kluge AG, Bult C (1994) Testing significance of incongruence. *Cladistics,* **10**, 315—319.

Feldman MW, Bergman A. Pollock DD, Goldstein DB (1997) Msat genetic distances with range constraints: Analytic description and problems of estimation. *Genetics,* **145**, 207—216.

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package), version 3.65. Department of Genetics, University of Washington, Seattle, Washington.

Fradkin PL (1984) A River No More – the Colorado River and the West. University of Arizona Press, Tucson, AZ.

Fu YX (1997) Statistical test of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915—925.

Gerber AS, Tibbets Am Dowling TE (2001) The role of introgressive hybridization in the evolution of the *Gila robusta* complex (Teleostei: Cyprinida). *Evolution*, 55, 2028—2039.

Gilpin ME, Soulé ME (1986) Minimum viable populations: processes of species extinction, p. 19—34. In: *Conservation Biology: the Science of Scarcity and Diversity* (Soulé ME ed.). Sinauer Press, Sunderland, MA.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on msats and the origin of modern humans. *Proceedings of the National Academy of Sciences USA*, **92**, 6723—6727.

Goloboff PA (1999) Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics*, **15**, 415—428.

Goodman, S. J. 1997. RST CALC: a collection of computer programs for calculating estimates of genetic differentiation from msat data and determining their significance. *Molecular Ecology*, **6**, 881—885.

Goudet J (2002) FSTAT. A program to estimate and test gene diversisties and fixation indices (version 2.9.3). <http://www.unnil.ch/izea/softwares/fstat.html>.

Grant WS, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *Journal of Heredity*, **89**, 415—426.

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

Haberl M, Tautz D (1999) Comparative allele sizing can produce inaccurate allele size differences for msats. *Molecular Ecology*, **8**, 1347—1350.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618—620.

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution*, **4**, 6—11.

Holden PB (1991) Ghosts of the Green River: Impacts of Green River poisoning on management of native fishes, p. 434—54. In: *Battle Against Extinction: Native Fish Management in the American West* (Minckley WL, Deacon JE, eds). University of Arizona Press, Tucson, AZ.

Holden PB, Minckley WL (1980) *Gila cypha* Miller, Humpback Chub, p. 165. In*: Atlas of North American freshwater fishes*, Lee DS, Gilbert CR, Hocutt CH, Jenkins RE, McAllister DE, Stauffer JR, Jr (eds.). North Carolina State Museum of Natural History, Raleigh NC.

Holden PB, Stalnaker CB (1970) Systematic studies of the cyprinid genus *Gila* in the upper Colorado River basin. *Copeia*, **1970**, 409—420.

Huelsenbeck JP, Ronquist F, Nielsen R, Bolback J (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310—2314.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153—159.

Kaeding LR, Burdick RD, Schrader PA, McAda CW (1990) Temporal and spatial relations between the spawning of humpback chub and roundtail chub in the upper Colorado River. *Transactions of the American Fisheries Society*, **119**, 135—144.

Kaeding LR, Zimmerman MA (1983) Life history and ecology of the humpback chub in the Little Colorado and Colorado Rivers of the Grand Canyon. *Transactions of the American Fisheries Society*, **112**, 577—594.

Karp CA, Tyus HM (1990) Humpback Chub (*Gila cypha*) in the Yampa and Green rivers, Dinosaur National Monument, with observations on Roundtail Chub *(Gila* robusta) and other sympatric fishes. *Great Basin Naturalist,* **50**, 257—264.

Keeler-Foster CL, Spies IB, Bondu-Hawkins V, Bentzen P (2004) Development of microsatellite markers in Bonytail (*Gila elegans*) with cross-species amplification in Humpback Chub (*Gila cypha*). *Molecular Ecology Notes*, **4**, 23—25.

Kolb E, Kolb E (1914) Experiences in the Grand Canyon. *National Geographic*, **26**, 99—184.

Kumar S (1996) Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. *Genetics*, **S**, 537—548.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Bioinformatics*, **5**,150–163.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295—302.

Leary RF, Allendorf FW, Sage GK (1995) Hybridization and introgression between introduced and native fish, p. 91-101. In: *Uses and Effects of Cultured Fishes in Aquatic Ecosystems* (Schramm HL, Piper RG, eds.). American Fisheries Society Symposium **15**, Bethesda, Maryland, USA.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *adh* locus in Drosophila. *Nature*, **352**, 652—654.

McElroy DM, Douglas ME (1995) Patterns of morphological variation among endangered populations of *Gila robusta* and *G. cypha* (Teleostei: Cyprinidae) in the upper Colorado River basin. *Copeia*, **1995**, 636—649.

McElroy DM, Shoemaker JA, Douglas ME (1997) Discriminating *Gila robusta* from *Gila cypha*: Risk assessment and the Endangered Species Act. *Ecological Applications*, **7**, 958—967.

Meredith EP, May B (2002) Microsatellite loci in the Lahontan Tui Chub, *Gila bicolor obesa*, and their utilization in other chub species. *Molecular Ecology Notes*, **2**, 156—158.

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. *Genetics*, **142**, 1061—1064.

Miller MP (1997) TFPGA (Tools for Population Genetic Analysis, version 1.3). Northern Arizona University, Flagstaff AZ. <www.marksgeneticsoftware.net/tfpga.htm>

Miller RR (1946) *Gila cypha*, a remarkable new species of cyprinid fish from the Colorado River in Grand Canyon, Arizona. *Journal of the Washington Academy of Science*, **36**, 409—415.

Miller RR (1955) Fish remains from archaeological sites in the lower Colorado River Basin, Arizona. *Papers of the Michigan Academy of Sciences, Arts and Letters*, **40**, 125—136.

Miller RR (1959) Origin and affinities of the freshwater fish fauna of western North America, p. 187-222. In: *Zoogeography,* Hubbs CL (ed.). American Association for the Advancement of Science, Publication **51**, Washington, DC.

Miller RR (1961) Man and the changing fish fauna of the American southwest. *Papers of the Michigan Academy of Science, Arts & Letters*, **46**, 365—404.

Miller RR (1964) Fishes of Dinosaur. *Naturalist*, **15**, 24—29.

Miller RR, Smith GR (1984) Fish remains from Stanton's Cave, Grand Canyon of the Colorado, Arizona, with notes on the taxonomy of *Gila cypha*, p. 61-65. In: *The Archaeology, Geology, and Paleobiology of Stanton's Cave, Grand Canyon National Park, AZ*. (Euler RC, ed.). Grand Canyon Natural History Assoc., Monogr. 6, Flagstaff, AZ.

Miller WH (1982) Concluding remarks, p. 130-131. In: *Fishes of the Upper Colorado River: Present and Future*. (Miller WH, Tyus HM, Carlson CA, eds.). American Fisheries Society, Western Division, Bethesda, MD..

Minckley WL (1991) Native fishes of the Grand Canyon region: An obituary? p. 124-177. In: *Colorado River Ecology and Dam Management* (Committee, eds.) National Academy Press, Washington DC.

Minckley WL, Douglas ME (1991) Discovery and extinction of Western fishes: A blink of the eye in geologic time, p. 7-17. In: *Battle Against Extinction: Native Fish Management in the American West* (Minckley WL, Deacon JE, eds). University of Arizona Press, Tucson, AZ.

Minckley WL, Hendrickson DA, Bond CE (1986) Geography of Western North American fishes: Description and relations to intracontinental tectonism, p. 519—613. In: *The Zoogeography of North American Freshwater Fishes* (Hocutt CH, Wiley EO, eds.). John Wiley and Sons, New York, NY.

Moritz C (1994a) Application of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, **3**, 401—411.

Moritz C (1994b) Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution*, **9**, 373—375.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583—590.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **105**, 207—217.

Nixon KC (1999) The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, **152**, 407—414.

Palsbøll PJ (1999) Genetic tagging: contemporary molecular ecology. *Biological Journal of the Linnean Society***, 68**, 3—22.

Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes,* **6**, 288—295.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945—959.

Quattro JM, Leberg PM, Douglas ME, Vrijenhoek RC (1996) Molecular evidence for unique evolutionary lineage of endangered Sonoran desert fish (Genus *Poeciliopsis*). *Conservation Biology*, **10**, 128—135.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetic software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248—249.

Reed RN, Douglas ME (2002) Ecology of the Grand Canyon Rattlesnake (*Crotalus viridis abyssus*) in the Little Colorado River Canyon, Arizona. *Southwestern Naturalist,* **47**, 30—39.

Reinert HK (1992) Radiotelemetric field studies of pitvipers: data acquisition and analysis, p. 185—197. In: *Biology of the Pitvipers*. (Campbell JA, Brodie ED Jr, eds.). Selva Publishing, Tyler, TX.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223—225.

Rozas J, Sánchez-De I, Barrio JC, Messenguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496—2497.

Ruzzante DE (1998) A comparison of several measure of genetic distance and population structure with msat data: Bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1—14.

Ruzzante DE, Taggart CT, Cook D, Goddard SV (1997) Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: A test and evidence of temporal stability. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 2700—2708.

Ryman N (1991) Conservation genetic considerations in fishery management. *Journal of Fish Biology* **39** (Suppl. A), 211—224.

Ryman N, Palm S, André C, Carvalho GR, Dahlgren TG, Jorde PE, Laiker L, Larsson LC, Palmé, A, and Ruzzante DE (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology*, **15**, 2031—2045.

Schlötterer C, Pemberton J (1998) The use of msats for genetic analysis of natural populations – a critical review, p. 72—86. In: *Molecular Approaches to Ecology and Evolution* (DeSalle R, Schierwater B, eds.). Birkhauser Verlag, Basel, Switzerland.

Sigler WF, Miller RR (1963) *Fishes of Utah*. Utah Division of Wildlife, Salt Lake City, UT.

Simons AM, Berendzen PB, Mayden RL (2003) Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. *Zoological Journal of the Linnean Society*, **139**, 68—80.

Simons AM, Mayden RL (1998) Phylogenetic relationships of the western North American phoxinins (Actinopterygii: Cyprinidae) as inferred from mitochondrial 12S and 16S ribosomal RNA sequences. *Molecular Phylogenetics and Evolution*, **9**, 308—329.

Sites JW Jr, Crandall KA (1997) Testing species boundaries in biodiversity studies. *Conservation Biology*, **11**, 1289—1297.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457—462.

Smith GR, Miller RR, Sable WD (1979) Species relationships among fishes of the genus *Gila* in the Upper Colorado River drainage, p. 613-623. In: *Proceedings of the Conference on Scientific Research in the National Parks*, U.S. National Park Service Transactions and Proceedings Series **5**.

Sober E (2004) The contest between parsimony and likelihood. *Systematic Biology,* **53**, 644—653.

Steel M, Penny D (2000) Parsimony, likelihood, and the role of models in molecular phylogenetics. *Molecular Biology and Evolution,* **17**, 839—850.

Swoffor DL (2001) PAUP*, *phylogenetic analysis using parsimony (and other methods)*, Version 4.04b. Sinauer Publishers, Sunderland, MA.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from msat DNA. *Genetics*, **144**, 389—399.

Tajima F (1993) Simple method for testing the molecular evolutionary clock hypothesis. *Genetics*, **135**, 599—607.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585—595.

United States Fish and Wildlife Service and United States National Marine Fisheries Service (1996) Policy regarding the recognition of Distinct Vertebrate Population Segments under the Endangered Species Act. *Federal Register*, **61**, 4722—4725.

Valdez RA, Clemmer GH (1982) Life history and prospects for recovery of the hympback chub and bonytail chub, p. 109—119. In: *Fishes of the Upper Colorado River system: Present and Future* (Miller WH, Tyus HM, Carlson CA, eds.). American Fisheries Society, Western Division, Bethesda, MD.

Valdez RA, Masslich WJ (1999) Evidence of reproduction by Humpback Chub in a warm spring of the Colorado River in Grand Canyon, Arizona. *Southwestern Naturalist*, **44**, 384—387.

Valdez RA, Ryel RJ (1995) Life history and ecology of the Humpback Chub (*Gila cypha*) in the Colorado River, Grand Canyon, Arizona. Final report to the Bureau of Reclamation (Contract 0-CS-40-09110), Salt Lake City, Utah. BIO/WEST Report TR-250-08. 268 pp.

Valdez, RA, Ryel RJ, Carothers SW (2000) Recovery goals of Humpback Chub (*Gila cypha*). *Draft final report to U.S. Fish and Wildlife Services*, Denver, CO. SWCA, Inc. Logan Utah.

Vanicek CD, Kramer RH (1969) Life history of the Colorado River Squawfish, Ptychocheilus lucius, and the Colorado Chub, Gila robusta, in the Green Reiver in Dinosaur National Monument, 1964-1966. *Transactions of the American Fisheries Society*, **98**, 193—208.

Van Oosterhout C, Hutchinson WF, Willis DPM, Shipley P (2004) MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535—538.

Vrijenhoek RC (1996) Conservation genetics of North American desert fishes, p. 367-397. In: *Conservation genetics: Case histories from nature* (Avise JC, Hamrick JL, eds.). Chapman and Hall Publishers, NYC, NY.

Vrijenhoek RC, Douglas ME, Meffe GK (1986) Conservation genetics of endangered fish populations in Arizona. *Science*, 229:400-402.

Waser PM, Strobeck C (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution,* **13**, 43—44.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**, 1123—1128.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177—1191.

Wilson IJ, Balding DJ (1998) Genealogical inference from msat data. *Genetics*, **150**, 499—510.

Wydoski RS, Hamill J (1991) Evolution of a cooperative recovery program for endangered fishes in the upper Colorado River basin, p. 123-140. In: *Battle Against Extinction: Native*

Fish Management in the American West (Minckley WL, Deacon JE, eds). University of Arizona Press, Tucson, AZ.

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany,* **129**, 157.

Zhang DX, Hewitt GM (2003) Nuclear DNA analysis in genetic studies of populations: practice, problems and prospects. *Molecular Ecology,* **12**, 563—58

FIGURE HEADINGS

Figure 1. Sampling locations for study populations of *Gila* within the Colorado River basin. (A) Box showing location of Grand Canyon. Panel (B) reveals location of *G. cypha* aggregates within Grand Canyon, whereas panel (C) reveals sampling locations in relation to the Upper Basin Colorado River (UB). Refer to Table 1 for population names and numbers.

Figure 2. Majority rule consensus tree based upon 12 equally parsimonious trees of length 608. Leaves of tree represent numbers of composite haplotypes (see Appendix 4) or species abbreviations (where BTC = *Gila elegans*; CPM = *Ptychocheilus lucius;* and CARP=*Cyprinus carpio*). Data based on 1,869 base pairs of sequence across three mtDNA regions (ATPase 8 & 6, ND2, and D-loop), screened across 331 individuals of *Gila* collected within the Colorado River basin. Refer to Table 1 for sampling size and collecting location of analyzed populations.

Figure 3. Haplotype network for *Gila cypha* and *G. robusta* from upper and lower Colorado River basin. Numbers correspond to composite haplotypes as listed in Appendix 4; *G. elegans* haplotypes not shown.

Figure 4. Haplotype network for *Gila cypha* and *G. robusta* from upper and lower Colorado River basin. Each haplotype is color-coded to represent the contribution(s) of individuals subsumed within each. HBC-GC= *G. cypha* from Grand Canyon; HBC-UB= *G. cypha* from upper basin; and RTC-UB= *G. robusta* from upper basin.

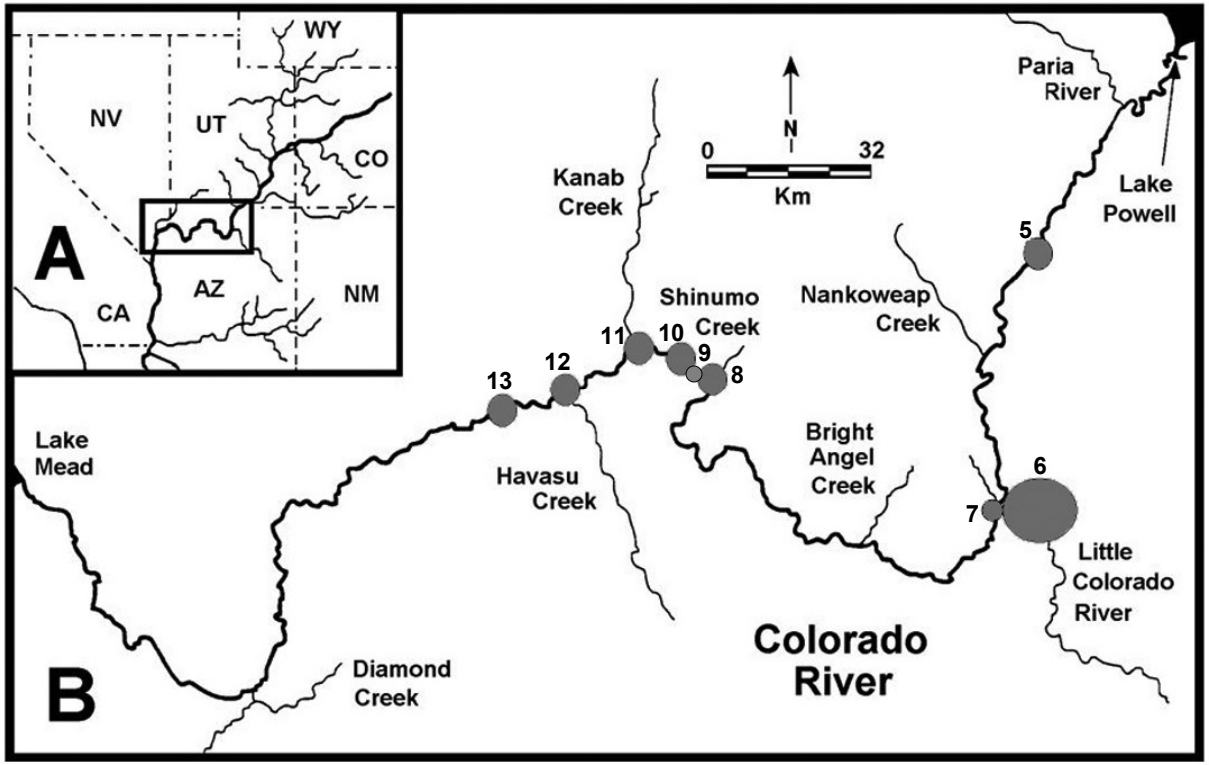
Figure 5. Admixture analysis from program BAPS that diagrammatically displays the proportion of mixing that has occurred among identified clusters of *Gila cypha, G. robusta,* and *G. elegans*. Population size = 4, with 1000 iterations to identify admixture coefficients for individuals, 200 individuals/ per population as reference, and 100 iterations to evaluate reference admixture. Clusters are identified at top-of-figure, whereas populations are listed at bottom of figure. UB = upper basin *G. cypha and G.*

robusta; DC = Desolation Canyon *G. cypha* and *G. robusta*; GC = Grand Canyon *G. cypha*; BT = *G. elegans.* Population names as per Table 1.

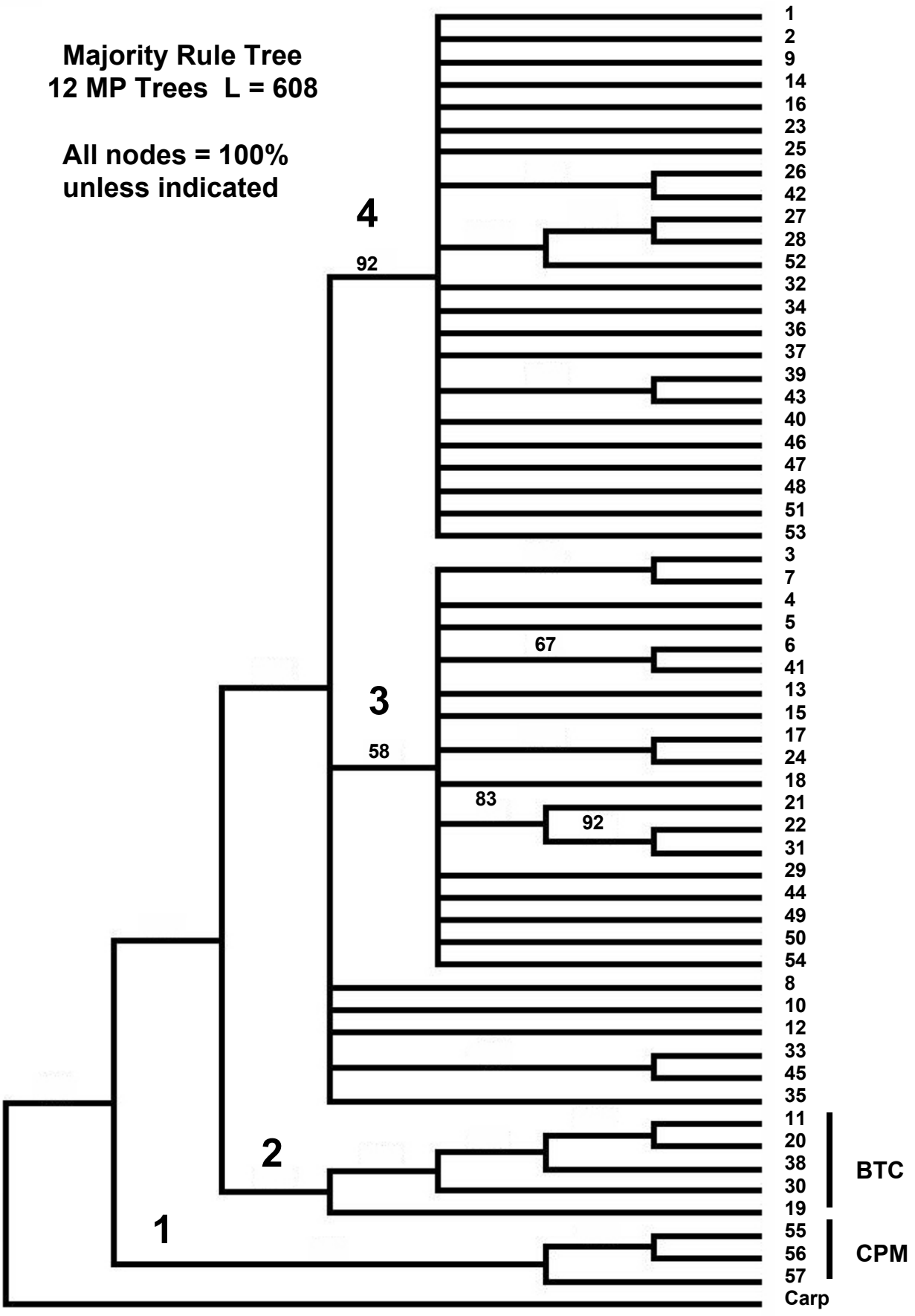
Figure 6. Plots to delimit ΔK and thus, to determine the number of clusters contained within the microsatellite (msat) DNA data amplified for 643 Colorado River *Gila cypha, G. robusta,* and *G. elegans* across 16 msat loci. For included samples refer to Table 5.

Figure 7. Admixture analysis from program STRUCTURE that diagrammatically displays the proportion of mixing that has occurred among identified clusters of *Gila cypha, G. robusta,* and *G. elegans*. Population size = 5, with the Markov Chain Monte Carlo algorithm performing 50,000 burn-in generations, followed by 1,500,000 iterations so as to develop the posterior probabilities. Clusters are identified at top-of-figure, whereas populations are listed on bottom (refer to Table 1 for population acronyms). UB-H = upper basin *G. cypha*; UB-R = upper basin *G. robusta*; DC/BT = Desolation Canyon *G. cypha and G. robusta*, as well as *G. elegans*; GC-H = Grand Canyon *G. cypha*; YR-R = *G. robusta* from Yampa River.

Figure 8. Admixture analysis from program STRUCTURE that diagrammatically displays the proportion of mixing that has occurred among identified clusters of *Gila cypha, G. robusta,* and *G. elegans*. Population size = 6, with the Markov Chain Monte Carlo algorithm performing 50,000 burn-in generations, followed by 1,500,000 iterations so as to develop the posterior probabilities. Clusters are identified at top-of-figure, whereas populations are listed on bottom (refer to Table 1 for population acronyms UB-H = upper basin *G. cypha*; UB-R = upper basin *G. robusta*; DC = Desolation Canyon *G. cypha and G. robusta*; GC-H = Grand Canyon *G. cypha*; YR-R = *G. robusta* from Yampa River; BT = *G. elegans*.

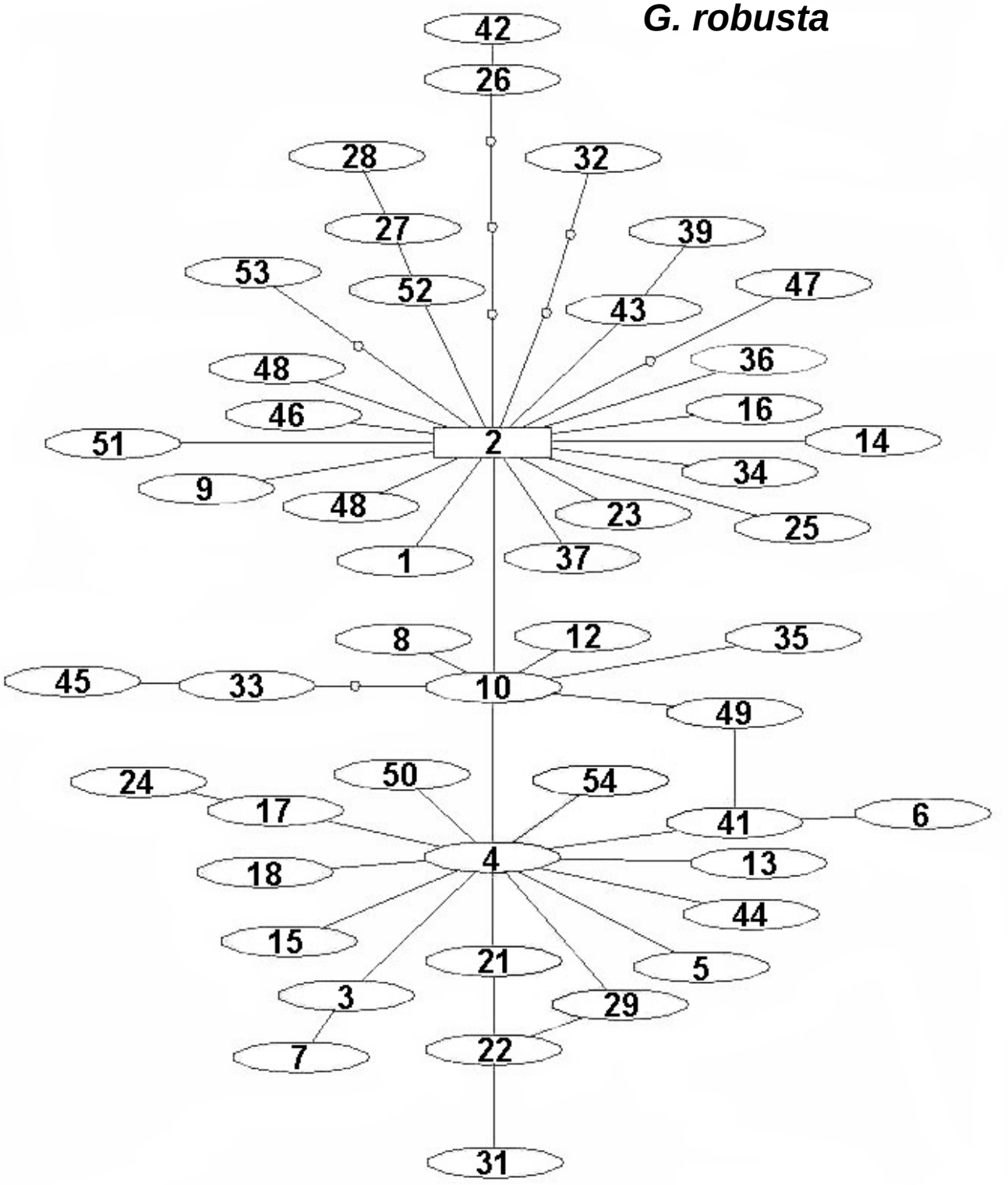

A
WY
NV
UT
CO
CA
AZ
NM
B
N
0
32
Km
Paria
River
Lake
Powell
Kanab
Creek
Shinumo
Creek
Nankoweap
Creek
Bright
Angel
Creek
Havasu
Creek
Lake
Mead
Diamond
Creek
Little
Colorado
River
Colorado
River
5
6
7
8
9
10
11
12
13


C
WYOMING
NEVADA
UTAH
FLAMING GORGE
RESERVOIR
YAMPA CANYON
YAMPA R.
DESOLATION CN
GREEN
RIVER
COLORADO
BLACK
ROCKS
CATARACT
CANYON
GUNNISON R.
VIRGIN R.
WESTWATER
CANYON
LAKE
POWELL
DOLORES R.
LAKE
MEAD
COLORADO
LITTLE
COLORADO
RIVER
SAN
JUAN RIVER
LAKE MOHAVE
GRAND
CANYON
LAKE HAVASU
NEW MEXICO
ARIZONA

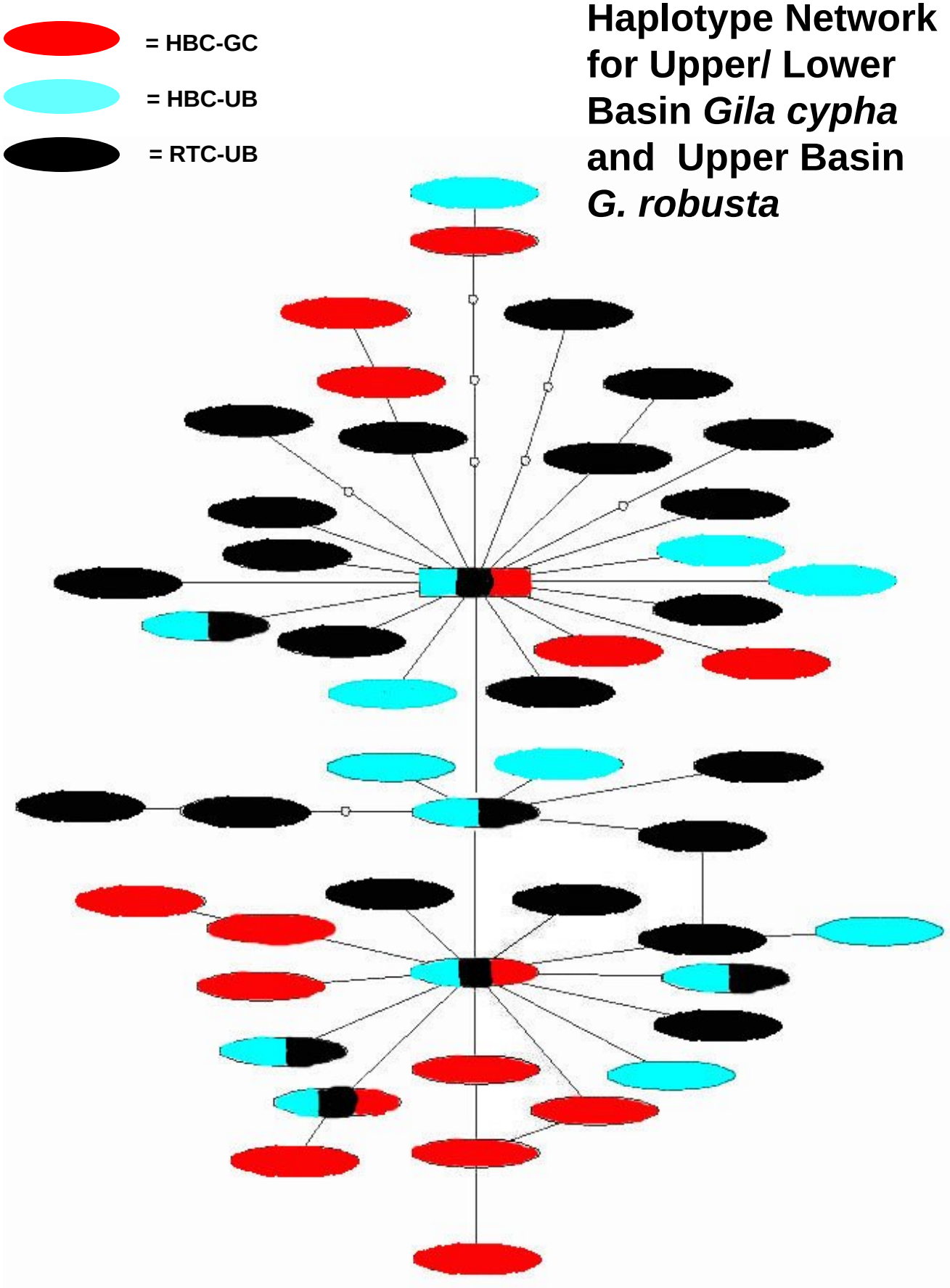

Majority Rule Tree
12 MP Trees L = 608
All nodes = 100%
unless indicated
4
92
3
58
67
83
92
2
1
1
2
9
14
16
23
25
26
42
27
28
52
32
34
36
37
39
43
40
46
47
48
51
53
3
7
4
5
6
41
13
15
17
24
18
21
22
31
29
44
49
50
54
8
10
12
33
45
35
11
20
38
30
19
55
56
57
Carp
BTC
CPM

F3

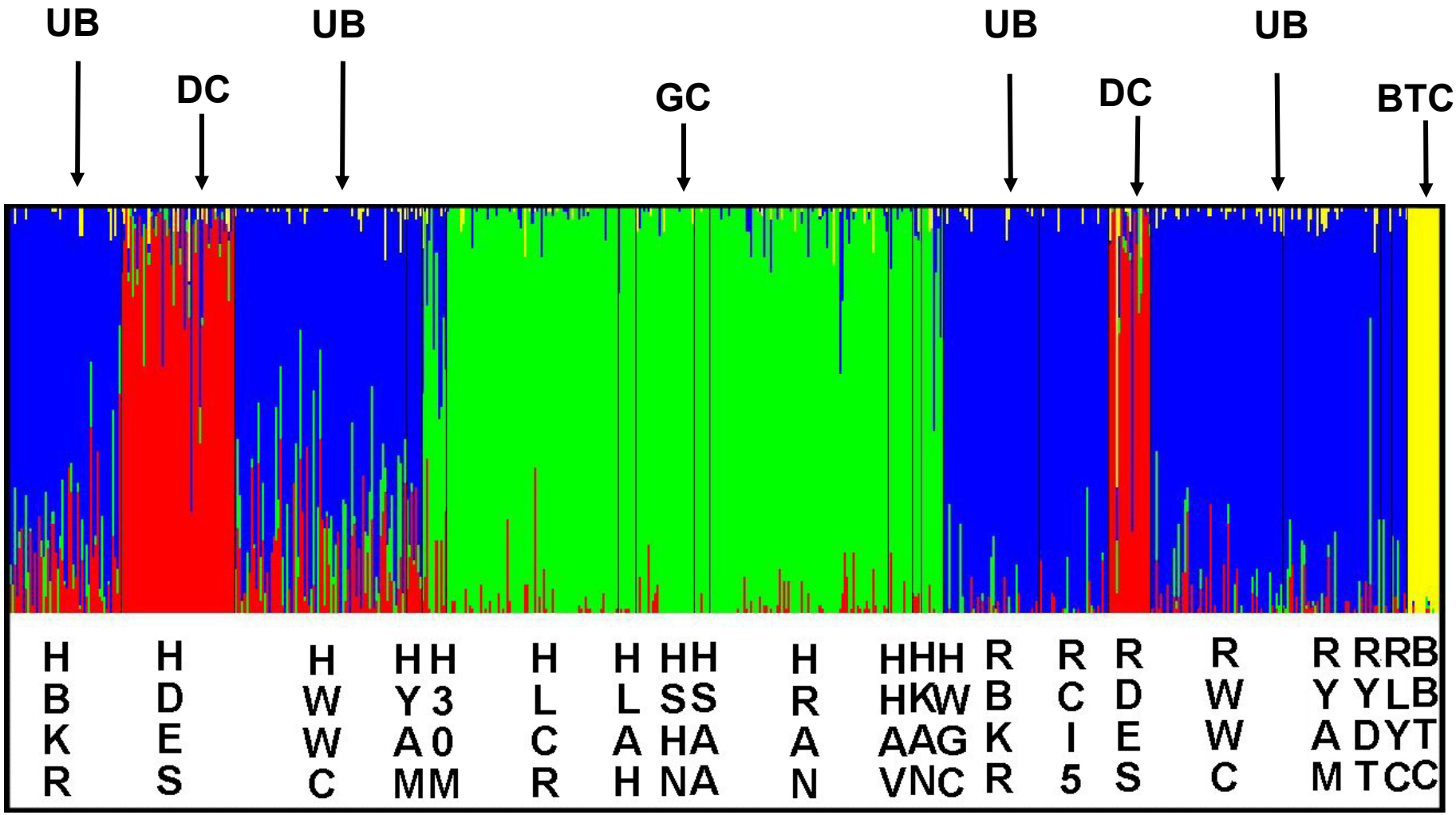
Haplotype Network for Upper/ Lower Basin *Gila cypha* and Upper Basin *G. robusta*

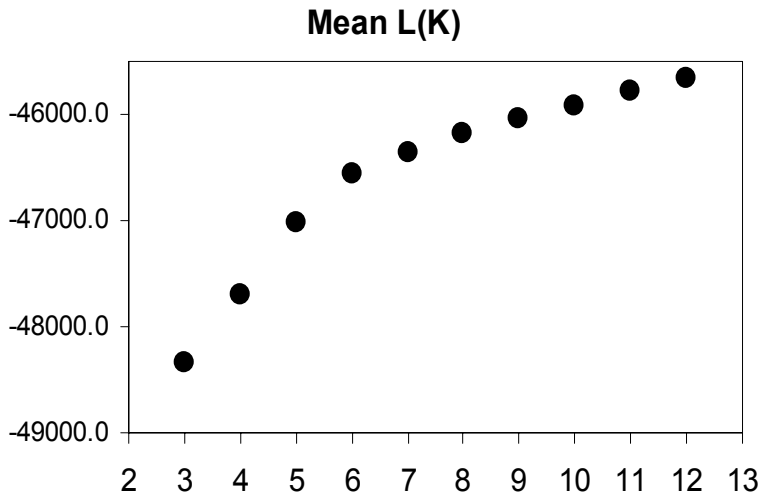


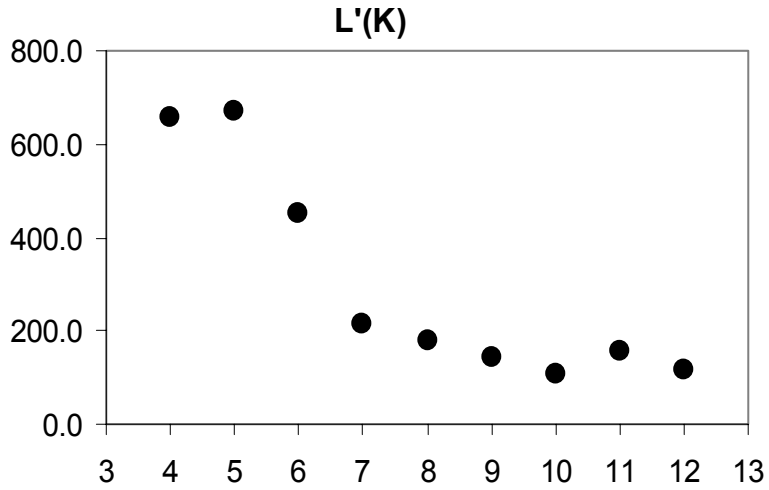
Haplotype Network for Upper/ Lower Basin *Gila cypha* and Upper Basin *G. robusta*

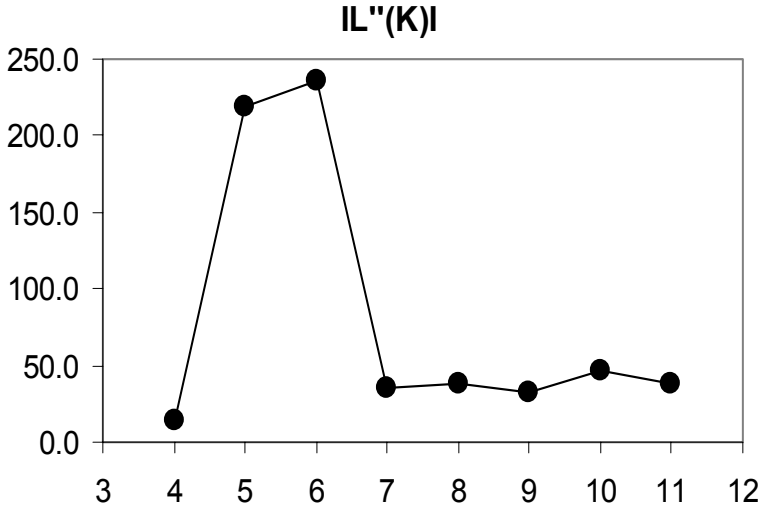


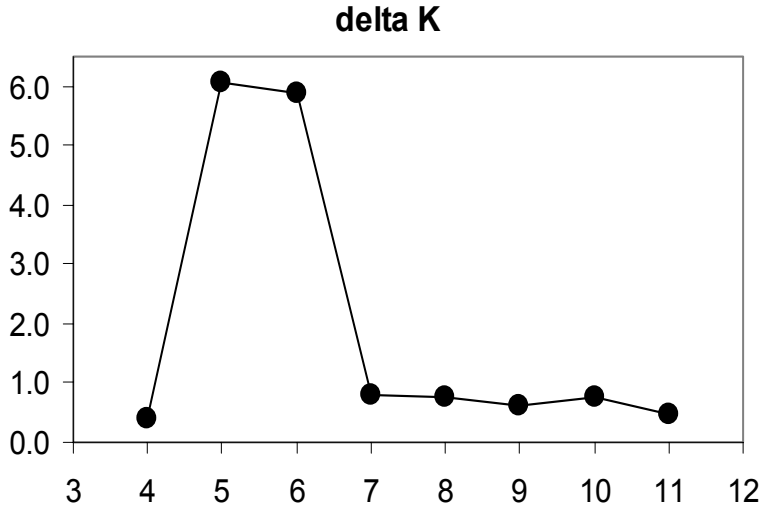
Four-Group Clustering (Program BAP)



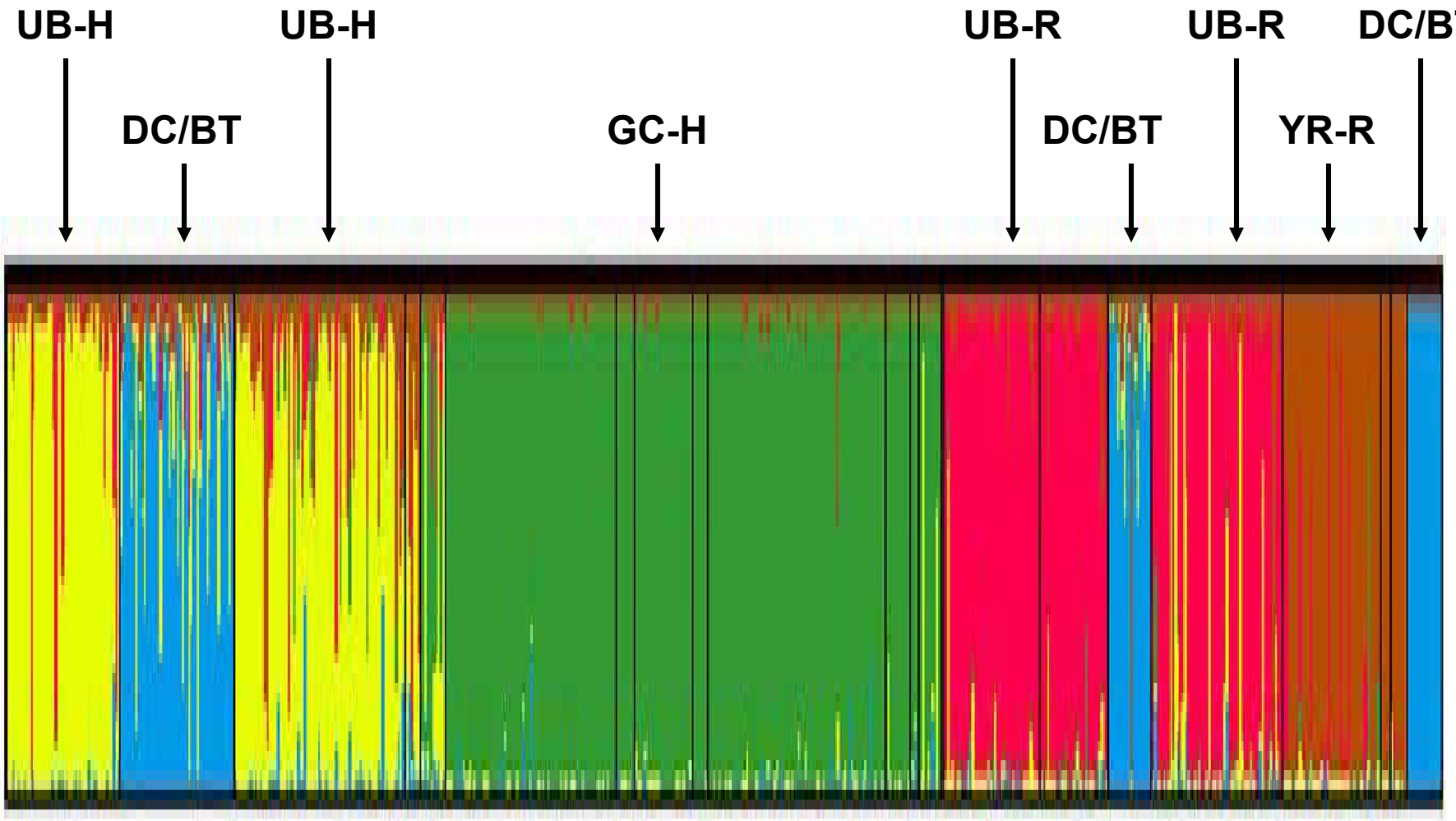

Mean L(K)
-46000.0
-47000.0
-48000.0
-49000.0
2 3 4 5 6 7 8 9 10 11 12 13


L'(K)
800.0
600.0
400.0
200.0
0.0
3 4 5 6 7 8 9 10 11 12 13


IL''(K)I
250.0
200.0
150.0
100.0
50.0
0.0
3 4 5 6 7 8 9 10 11 12


delta K
6.0
5.0
4.0
3.0
2.0
1.0
0.0
3 4 5 6 7 8 9 10 11 12

Five-Group Clustering (Program STRUCTURE)



Six-Group Clustering (Program STRUCTURE)

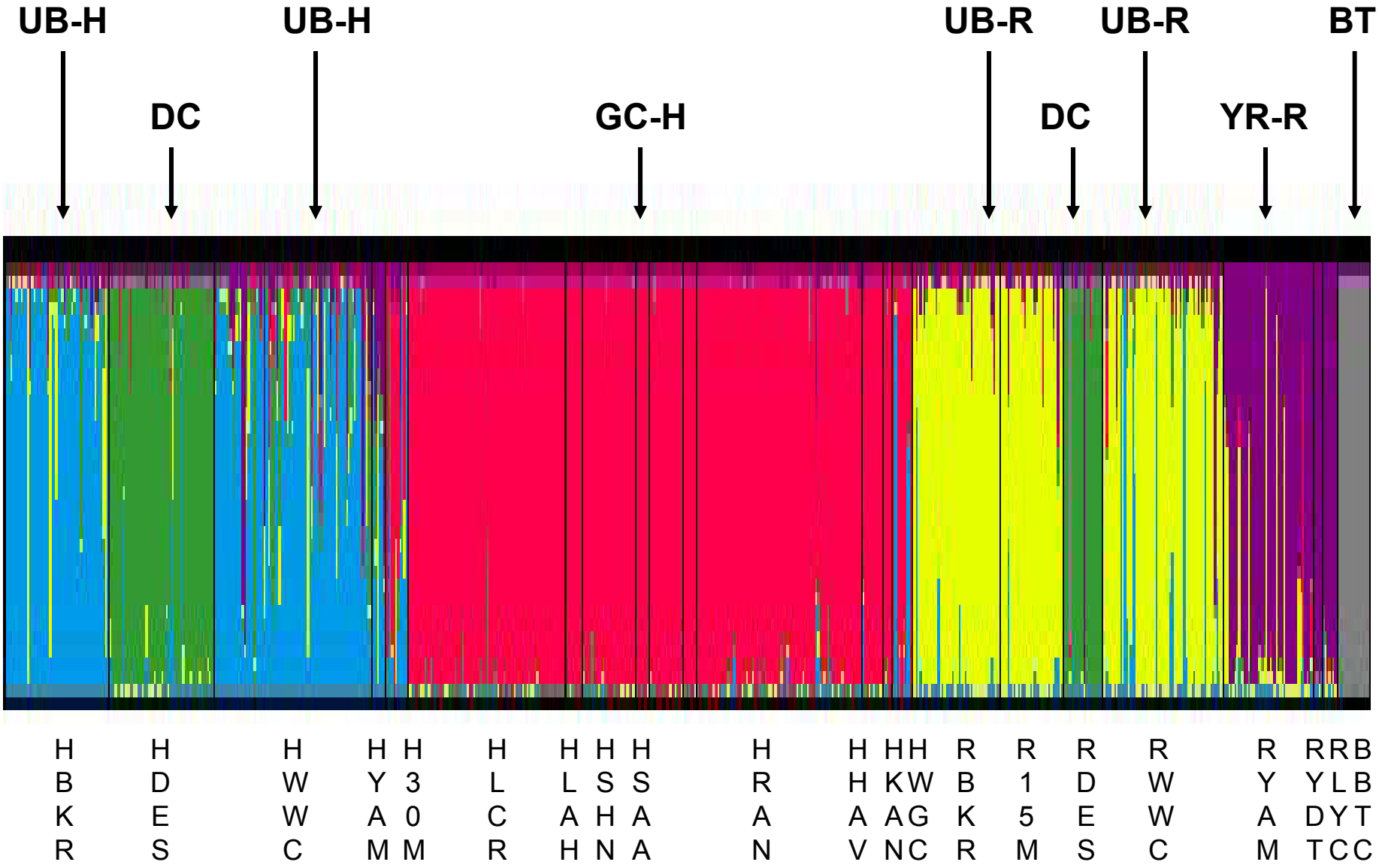


Table 1. Locations of the 331 samples used in mtDNA sequence analysis. Sampling localities are listed by phenotype/ basin. Provided for each population are name (=Pop), number (=Pop#), acronym (=Code) and sample size (=N). Abbreviations are: HBC = Humpback Chub, *Gila cypha*; RTC = Roundtail Chub, *Gila robusta*; UB = Upper Basin; GC = Grand Canyon. Total number of samples for each phenotype/basin are provided in parentheses.

Phenotype/Basin	Pop	Pop#	Code	N
HBC-UB	Black Rocks	1	**HBKR**	16
(62)	Desolation Cn	2	**HDES**	21
	Westwater Cn	3	**HWWC**	20
	Yampa R	4	**HYAM**	5
HBC-GC	30-Mile area	5	**H30M**	9
(152)	Little Colorado	6	**HLCR**	46
	Lava to Hance	7	**HLAH**	8
	Shinumo Ck	8	**HSHN**	24
	Stephen's Aisle Area	9	**HSAA**	6
	Randy's Rock	10	**HRAN**	44
	Kanab Ck	11	**HKAN**	3
	Havasu Ck	12	**HHAV**	9
	Western Grand Cn	13	**HWGC**	3
RTC-UB	Black Rocks	14	**RBKR**	19
(117)	15-Mile Reach	15	**RC15**	20
	Desolation Cn	16	**RDES**	22
	Westwater Cn	17	**RWWC**	21
	Yampa R	18	**RYAM**	20
	Yampa Duffy Tunnel	19	**RYDT**	3
	Little Yampa Cn	20	**RLYC**	7
	Wyoming	22	**RWYO**	5
			Total	331

Table 2: Molecular diversity across four mtDNA regions (ATPase 8 & 6, ND2, and D-Loop; total 1,869 base pairs) sequenced for *Gila cypha* and *G. robusta* populations sampled from upper and lower Colorado River basins. Population information provided in Table 1. Abbreviations are: Code=population acronym; N=sample size; BTC=number of *G. elegans* haplotypes removed from the population for calculation purposes; Nf=final number of individuals; H=number of haplotypes; Ps=Total number of polymorphic sites; h= haplotype diversity (standard deviation in parentheses); π= nucleotide diversity (standard deviation in parentheses); K=average within-group nucleotide difference (i.e., how much on average two haplotypes differ; sampling variance in parentheses, no recombination). Table lists molecular diversity 2a) at the population level, 2b) for upper basin localities (*G. cypha* and *G. robusta* pooled), and 2c) for species/subdrainages (populations for each species pooled by basin).

Table 2a

Code	N	BTC	Nf	H	Ps	*h*	π	K
HBKR	16		16	8	9	0.858 (0.063)	0.00116 (0.00016)	2.15 (0.194)
HDES	21	1	20	3	2	0.353 (0.123)	0.00020 (0.00007)	0.37 (0.013)
HWWC	20		20	8	8	0.742 (0.096)	0.00087 (0.00015)	1.61 (0.096)
HYAM	5		5	4	5	0.901 (0.161)	0.00110 (0.00037)	2.00 (0.650)
H30M	9	2	7	4	4	0.857 (0.102)	0.00093 (0.00018)	1.71 (0.340)
HLCR	46	2	44	9	10	0.783 (0.047)	0.00095 (0.00010)	1.76 (0.049)
HLAH	8		8	3	9	0.679 (0.122)	0.00172 (0.00059)	3.179 (0.881)
HSHN	24		24	7	9	0.775 (0.063)	0.00122 (0.00022)	2.25 (0.136)
HSAA	6		6	3	2	0.600 (0.215)	0.00047 (0.00018)	0.867 (0.144)
HRAN	44	2	42	6	5	0.769 (0.034)	0.00080 (0.00006)	1.487 (0.040
HKAN	3		3	3	3	1.000 (0.272)	0.00108 (0.00038)	2.000 (1.238)
HHAV	9		9	5	9	0.722 (0.159)	0.00135 (0.00055)	2.500 (0.471)
HWGC	3		3	3	3	1.000 (0.272)	0.00108 (0.00038)	2.000 (1.238)
RBKR	19		19	5	8	0.696 (0.095)	0.00108 (0.00022)	2.000 (0.144)
RC15	20		20	8	7	0.779 (0.083)	0.00083 (0.00012)	1.526 (0.089)
RDES	22	1	21	5	5	0.424 (0.131)	0.00030 (0.00011)	0.562 (0.021)
RWWC	21		21	9	11	0.795 (0.077)	0.00120 (0.00020)	2.210 (0.151)
RYAM	20		20	9	9	0.858 (0.054)	0.00109 (0.00013)	2.005 (0.137)
RYDT	3		3	3	3	1.000 (0.272)	0.00011 (0.00008)	2.000 (1.238)
RLYC	7		7	7	11	1.000 (0.076)	0.00201 (0.00243)	3.714 (1.245)
RWYO	5		5	4	4	0.900 (0.161)	0.00087 (0.00026)	1.600 (0.453)
Total	**331**	**8**	**323**	**49**	**49**	**0.860 (0.012)**	**0.00128 (0.00004)**	**2.361 (0.010)**

Table 2b

Code	N	BTC	Nf	H	Ps	*h*	π	K
ALL-BKR	35		35	11	15	0.780 (0.059)	0.00117 (0.00016)	2.168 (0.086)
ALL-DES	43	2	41	5	5	0.384 (0.092)	0.00025 (0.00007)	0.466 (0.008)
ALL-WWC	41		41	14	16	0.762 (0.063)	0.00103 (0.00014)	1.910 (0.060)
ALL-YAM	35		35	15	19	0.877 (0.038)	0.00125 (0.00016)	2.309 (0.095)

Table 2c

Code	N	BTC	Nf	H	Ps	*h*	π	K
HBC-UB	62	1	61	15	15	0.695 (0.00398)	0.00083 (0.00010)	1.530 (0.028)
HBC-GC	152	6	146	15	17	0.772 (0.025)	0.00102 (0.00008)	1.880 (0.016)
RTC-UB	117	1	116	30	35	0.761 (0.039)	0.00103 (0.00009)	1.899 (0.020)

Table 3. Pair-wise percent sequence divergence between populations of *Gila cypha* and *G. robusta* sampled from upper and lower basin locations as defined in Table 1 (populations RYAM, RYDT and RLYC pooled as RYAM). Values in lower triangle are mean percent sequence divergence (p-distances) corrected for within-group variance, while those in upper triangle are standard errors. BTC = *Gila elegans*; CPM = *Ptychocheilus lucius.*

	HBKR	HDES	HWWC	HYAM	H30M	HLCR	HLAH	HSHN	HSAA	HRAN	HKAN	HHAV	HWGC	RBKR	RC15	RDES	RYAM	RWWC	RWYO	BTC	CPM
HBKR	**X**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0.5
HDES	0	**X**	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0	0.4	0.5
HWWC	0	0	**X**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0.5
HYAM	0	0	0	**X**	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0.5	0.5
H30M	0.1	0.1	0.1	0.1	**X**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0.4
HLCR	0	0.1	0	0.1	0	**X**	0	0	0	0	0	0	0	0	0	0.1	0	0	0.1	0.5	0.5
HLAH	0	0.1	0	0.1	0.1	0	**X**	0	0	0	0	0	0	0	0	0.1	0	0	0.1	0.5	0.5
HSHN	0	0.1	0	0.1	0.1	0	0	**X**	0	0	0	0	0	0	0	0.1	0	0	0.1	0.5	0.5
HSAA	0	0.1	0.1	0.1	0.1	0	0	0	**X**	0	0	0	0	0.1	0	0.1	0	0	0.1	0.5	0.5
HRAN	0	0.1	0.1	0.1	0.1	0	0	0	0	**X**	0	0	0	0	0	0.1	0	0	0.1	0.5	0.5
HKAN	0.1	0.2	0.1	0.1	0.2	0	0	0	0	0	**X**	0	0	0.1	0.1	0.1	0	0	0.1	0.5	0.5
HHAV	0	0.1	0	0.1	0.1	0	0	0	0	0	0	**X**	0	0	0	0.1	0	0	0.1	0.5	0.5
HWGC	0	0	0	0	0.1	0	0	0	0	0	0	0	**X**	0	0	0	0	0	0	0.5	0.5
RBKR	0	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	**X**	0	0	0	0	0	0.5	0.5
RC15	0	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	**X**	0	0	0	0	0.5	0.5
RDES	0	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	**X**	0	0	0	0.5	0.5
RYAM	0	0	0	0	0.1	0	0	0	0	0	0.1	0	0	0	0	0	**X**	0	0	0.5	0.5
RWWC	0	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	**X**	0	0.5	0.5
RWYO	0	0	0	0	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0	0	0	0	0	0	**X**	0.5	0.5
BTC	4.7	4.3	4.7	4.7	2.8	4.4	4.8	4.8	4.8	4.4	4.8	4.8	4.7	4.7	4.7	4.3	4.7	4.7	4.8	**X**	0.5
CPM	5.3	5.2	5.4	5.4	4.6	5.1	5.3	5.3	5.3	5.1	5.3	5.3	5.3	5.4	5.4	5.2	5.3	5.4	5.4	6.1	**X**

Table 4. Molecular estimates of population expansion/ contraction as calculated across 1,869 base pairs of combined sequence from four mtDNA regions (ATPase 8 & 6, ND2, and D-Loop) calculated for *Gila cypha* and *G. robusta* populations sampled from upper and lower Colorado River basins locations. Population information provided in Table 1. Code=population acronym; N=sample size; BTC=number of *G. elegans* haplotypes removed from the population for calculation purposes; Nf=final number of individuals; T-D = Tajima`s D; Fu-F = Fu`s F.

Code	N	BTC	Nf	T-D	Fu-F
HBKR	16		16	-0.76140	-2.5959 (P<0.041)
HDES	21	1	20	-0.81235	-0.7746
HWWC	20		20	-0.97106	-3.0110 (P<0.017)
HYAM	5		5	-1.12400	-1.0116
H30M	9	2	7	0.23902	-0.4280
HLCR	46	2	44	-0.68630	-1.8860
HLAH	8		8	-0.41280	2.6909
HSHN	24		24	-0.21280	-0.4769
HSAA	6		6	-0.05002	-0.4268
HRAN	44	2	42	0.70313	-0.0449
HKAN	3		3	n/a	n/a
HHAV	9		9	-1.12780	-0.3540
HWGC	3		3	n/a	n/a
RBKR	19		19	-0.43173	0.6046
RC15	20		20	-0.74450	-3.2079 (P<0.019)
RDES	22	1	21	-1.79547	-2.593 (P<0.011)
RWWC	21		21	-0.97113	-2.7474 (P<0.036)
RYAM	20		20	-0.72060	-3.305 (P<0.013)
RYDT	3		3	n/a	n/a
RLYC	7		7	-0.93421	-3.7838 (P<0.004)
RWYO	5		5	-1.09380	-1.4048

Table 5. Overview of populations genotyped across 16 nuclear microsatellite loci. Population abbreviations are as in Table 1. Listed are populations of *Gila cypha* from upper basin (= HBC-UB), *G. cypha* from Grand Canyon (= HBC-GC), and *G. robusta* from upper basin (= RTC-UB). N = sample size for each population, Total = sample size for each region (by taxon), and the overall total.

Taxa	POP	Pop#	N	Total
HBC-UB	**HBKR**	1	51	
	HDES	2	51	**186**
	HWWC	3	77	
	HYAM	4	7	
HBC-LB	**H30M**	5	11	
	HLCR	6	77	
	HLAH	7	8	
	HSHN	8	26	
	HSAA	9	7	**234**
	HRAN	10	80	
	HHAV	11	11	
	HKAN	12	4	
	HWGC	13	10	
RTC-UB	**RBKR**	14	43	
	RC15	15	31	
	RDES	16	19	
	RWWC	17	59	**208**
	RYAM	18	44	
	RYDT	19	5	
	RLYC	20	7	
BTC	**BBTC**	21	15	**15**
		Overall Total		**643**

Table 6. Characteristics of 20 microsatellite loci screened across 21 populations. Loci were run in sets of four (i.e., Multiplex), with a specific fluorescent dye (= Dye) identifying each locus (i.e., 6FAM = blue, VIC = green, NED = yellow, PET = red). Differences between consecutive alleles (= Repeat) was either four base-pairs (= tetra) or two (= di). Listed are total number of alleles, minimum and maximum allele size (in base-pairs), and variance in repeat number across all genotyped individuals. Also provided are the name of each locus employed in this study, while their source (= reference) is listed in Literature Cited.

Multiplex	Locus	Dye	Repeat	Alleles	Allele Size		Var_{RP}	Source	
				total	min	max		name	reference
MP1	**G38**	6FAM	Tetra	42	295	479	54.0	Gbi-G38	Meredith & May 2002
	G13	VIC	Tetra	29	215	327	21.2	Gbi-G13	Meredith & May 2002
	G34	NED	Tetra	33	204	332	59.8	Gbi-G34	Meredith & May 2002
	G211*	PET	Tetra	n/a	n/a	n/a	n/a	Gbi-G211	Meredith & May 2002
MP2	**G39**	6FAM	Tetra	19	214	286	8.4	Gbi-G39	Meredith & May 2002
	C01	VIC	Di	42	118	212	88.5	Ca1	Dimsoski et al. 2000
	G02	NED	Tetra	19	213	293	13.6	Gbi-G2	Meredith & May 2002
	G294	PET	Tetra	10	213	257	3.5	Gbi-G294	Meredith & May 2002
MP3	**Gel14**	6FAM	Tetra	3	74	90	1.2	Gel_267	Keeler-Foster et al. 2004
	C13*	VIC	Di	n/a	n/a	n/a	n/a	Ca13	Dimsoski et al. 2000
	C02*	NED	Di	n/a	n/a	n/a	n/a	Ca2	Dimsoski et al. 2000
	C03	PET	Tetra	37	245	405	31.7	Ca3	Dimsoski et al. 2000
MP4	**G03***	6FAM	Tetra	n/a	n/a	n/a	n/a	Gbi-G3	Meredith & May 2002
	C12	VIC	Tetra	21	198	278	13.1	Ca12	Dimsoski et al. 2000
	G87	NED	Tetra	32	201	535	31.6	Gbi-G87	Meredith & May 2002
	G99	PET	Tetra	51	337	561	104.4	Gbi-G99	Meredith & May 2002
MP5	**Gel16**	6FAM	Tetra	35	160	320	41.4	Gel_300	Keeler-Foster et al. 2004
	Gel06	VIC	Tetra	45	122	302	58.4	Gel_225	Keeler-Foster et al. 2004
	Gel11	NED	Tetra	27	155	267	13.5	Gel_234	Keeler-Foster et al. 2004
	Gel09	PET	Tetra	24	200	292	16.3	Gel_228	Keeler-Foster et al. 2004

Table 7. Mean allelic patterns for 20 chub populations (= Pop) genotyped across 16 microsatellite loci (*Gila elegans* not shown). Population abbreviations are as in Table 1. Listed are (with variance in parentheses): N = sample size; Na = number of alleles; Ne = number of effective alleles; Ap = private alleles. Also listed are less common alleles with frequency <= 5% (<5%), 25% (<25%), and 50% (<50%).

Pop	N	Na	Ne	Ap	<5%	<25%	<50%
HBKR	51	18.1	10.5	0.5	6.6	4.6	9.6
		(1.8)	(1.3)	(0.2)	(0.5)	(0.8)	(1.4)
HDES	51	17.0	8.7	0.4	5.6	4.6	9.3
		(1.8)	(1.1)	(0.2)	(0.5)	(0.8)	(1.3)
HWWC	77	20.1	11.2	0.9	7.0	5.4	10.6
		(2.1)	(1.4)	(0.3)	(0.6)	(1.1)	(1.5)
HYAM	7	7.3	5.2	0.1	7.3	1.8	4.1
		(0.6)	(0.4)	(0.1)	(0.6)	(0.5)	(0.6)
H30M	11	8.7	5.7	0.0	4.4	1.5	4.6
		(0.8)	(0.6)	(0.0)	(0.3)	(0.3)	(0.8)
HLCR	77	14.3	6.5	0.1	6.2	3.4	7.9
		(1.5)	(0.7)	(0.1)	(0.5)	(0.6)	(1.1)
HLAH	8	6.7	5.1	0.0	6.7	0.9	3.1
		(0.7)	(0.5)	(0.0)	(0.7)	(0.2)	(0.5)
HSHN	26	10.6	5.9	0.1	6.1	2.1	5.8
		(1.1)	(0.6)	(0.1)	(0.4)	(0.5)	(0.9)
HSAA	7	6.4	4.8	0.0	6.4	1.1	3.3
		(0.6)	(0.5)	(0.0)	(0.6)	(0.3)	(0.7)
HRAN	80	14.4	6.5	0.3	6.0	3.5	8.1
		(1.5)	(0.7)	(0.2)	(0.5)	(0.8)	(1.2)
HHAV	11	7.8	5.3	0.0	4.9	1.2	3.9
		(0.6)	(0.5)	(0.0)	(0.3)	(0.3)	(0.6)
HKAN	4	4.9	4.0	0.0	4.9	0.6	2.0
		(0.4)	(0.5)	(0.0)	(0.4)	(0.2)	(0.3)
HWGC	10	7.5	4.2	0.0	7.5	0.9	3.9
		(0.7)	(0.4)	(0.0)	(0.7)	(0.2)	(0.6)
RBKR	43	15.3	8.8	0.2	6.4	4.0	8.3
		(1.8)	(1.3)	(0.2)	(0.7)	(0.8)	(1.3)
RC15	31	14.3	8.8	0.2	5.4	4.1	8.3
		(1.6)	(1.2)	(0.1)	(0.5)	(0.9)	(1.3)
RDES	19	12.7	7.3	0.1	6.8	3.4	7.0
		(1.3)	(0.9)	(0.1)	(0.6)	(0.7)	(1.1)
RWWC	59	18.0	9.8	0.2	6.6	4.9	9.9
		(2.0)	(1.3)	(0.1)	(0.7	(1.0	(1.5)
RYAM	44	16.8	9.6	0.8	5.8	4.3	9.2
		(2.0)	1.3	(0.3)	(0.6	(0.9)	(1.5)
RYDT	5	6.1	(5.0	0.1	6.1	1.1	3.1
		(0.5)	(0.5)	(0.1)	(0.5	(0.3)	(0.5)
RLYC	7	7.5	5.9	0.0	7.5	1.9	4.2
		(0.7)	(0.7)	(0.0)	(0.7	(0.4)	(0.6)

Table 8. Private alleles detected in the 21 populations (= Pop) genotyped across the 16 microsatellite loci. For each population, total number of private alleles (= Total) are shown, as well as private alleles found at a particular locus for each population, and total numbers of private alleles for each locus (= All Pop). Zero-values (i.e., no private alleles) are not shown. Population abbreviations are as in Table 1.

Pop	Total	Locus															
		G38	G13	G34	G39	C01	G02	G294	Gel14	C03	C12	G87	G99	Gel16	Gel06	Gel11	Gel09
HBKR	**8**	1				1	1				2	1		1		1	
HDES	**7**	3											2	1	1		
HWWC	**15**	2	2			4				3	1		1			2	
HYAM	**2**					1										1	
H30M																	
HLCR	**1**									1							
HLAH																	
HSHN	**1**												1				
HSAA																	
HRAN	**5**	2								1		1	1				
HHAV																	
HKAN																	
HWGC																	
RBKR	**3**														3		
RC15	**3**				1									1	1		
RDES	**2**	1	1														
RWWC	**3**		1	1												1	
RYAM	**12**	4		1						3			2	2			
RYDT	**2**									1			1				
RLYC																	
BTC	**6**	1			1		2						2				
All Pop.	**70**	**14**	**4**	**2**	**2**	**6**	**3**			**9**	**3**	**2**	**10**	**5**	**5**	**5**	

Table 9. Exact tests for population differentiation (Raymond and Rousset, 1995) performed in software TFPGA. Pair-wise analyses are executed by location, based upon genotypes of individuals sampled from populations. Number of dememorization steps = 2000; Number of batches = 30; Number of permutations per batch = 10,000. Adjusted probability value for each population in pair-wise tests would be $p = 0.0025$. Values not significantly different at the adjusted *p*-value are in bold. Numbers above column names indicate population membership. GP1 through GP8 reflect different genepools (reproductive populations) identified by TFPGA.

	GP1	GP2	GP1	GP3	GP4	GP5	GP5	GP5	GP5	GP5	GP5	GP5	GP5	GP6	GP6	GP2	GP6	GP7	GP7	GP7	GP8
CODE	**HBKR**	**HDES**	**HWWC**	**HYAM**	**H30M**	**HLCR**	**HLAH**	**HSHN**	**HSAA**	**HRAN**	**HHAV**	**HKAN**	**HWGC**	**RBKR**	**RC15**	**RDES**	**RWWC**	**RYAM**	**RYDT**	**RLYC**	**BBTC**
HBKR	X																				
HDES	0	X																			
HWWC	**0.004**	0	X																		
HYAM	0	0	0	X																	
H30M	0	0	0	0	X																
HLCR	0	0	0	0	0	X															
HLAH	0	0	0	0	0	**0.864**	X														
HSHN	0	0	0	0	0	**0.003**	**0.13**	X													
HSAA	0	0	0	0	0	**0.743**	**0.898**	**0.299**	X												
HRAN	0	0	0	0	0	0.0005	**0.665**	0	**0.827**	X											
HHAV	0	0	0	0	0	**0.018**	**0.557**	**0.004**	**0.226**	0.0002	X										
HKAN	0	0	0.0001	0.0003	0	**0.431**	**0.521**	**0.034**	**0.286**	**0.331**	**0.106**	X									
HWGC	0	0	0	0	0	0	**0.012**	0	**0.003**	0	0	0.0006	X								
RBKR	0	0	0	0	0	0	0	0	0	0	0	0	0	X							
RC15	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.184**	X						
RDES	0	**0.4**	0	0	0	0	0	0	0	0	0	0	0	0	0	X					
RWWC	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.696**	**0.821**	0	X				
RYAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X			
RYDT	0	0	0.002	0.0002	0	0	0	0	0	0	0	0.0019	0	0	0	0	0	**0.645**	X		
RLYC	0	0	0	0	0	0	0	0	0	0	0	0.0003	0	0	0	0	0	**0.814**	**0.661**	X	
BBTC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X

Table 10. Designation of clusters using program STRUCTURE at K=4, K=5, and K=6. FST values are shown for each cluster under each K-value. Clusters indicate: GC-H = Grand Canyon *Gila cypha*; UB-H = Upper Basin *G. cypha* ; UB-R = Upper Basin *G. robusta*; DC/BT = *G. cypha* and *G. robusta* in Desolation Canyon (UT) and *G. elegans*; YR-R = *G. robusta* in Yampa River; DC = *G. robusta* and *G. cypha* in Desolation Canyon (UT); BT = *G. elegans*. Probability of assignment to specific clusters identified by STRUCTURE is indicated in color as shown in Figure 7 (K=5) and Figure 8 (K=6); assignment by STRUCTURE at K=4 is not depicted in a Figure.

K	Cluster	F_{ST}	Color
4	GC-H	**0.0785**	not shown
4	UB-H	**0.0129**	not shown
4	UB-R	**0.2900**	not shown
4	DC/BT	**0.0466**	not shown
5	GC-H	**0.0809**	green
5	UB-H	**0.0153**	yellow
5	UB-R	**0.0368**	red
5	YR-R	**0.0291**	braun
5	DC/BT	**0.0500**	blue
6	GC-H	**0.0839**	red
6	UB-H	**0.0171**	blue
6	UB-R	**0.0384**	yellow
6	YR-R	**0.0311**	purple
6	DC	**0.0427**	green
6	BT	**0.1579**	grey

Table 11. Proportion of membership of each pre-defined population (=POP) in each of the designated clusters using program STRUCTURE at K=4, 5 or 6.Population names as per Table 1, and cluster acronyms are as in Table 9; N = sample size; 4-1/ 4-4: K = four clusters (1--4); 5-1/ 5-5: K= five clusters (1--5); 6-1/ 6-6: K = six clusters (1--6). Membership of populations within clusters is indicated with bold type and shading.

		GC-H	UB-H	UB-R	DC/BT	GC-H	UB-H	UB-R	YR-R	DC/BT	GC-H	UB-H	UB-R	YR-R	DC	BT
POP	**N**	**4-1**	**4-2**	**4-3**	**4-4**	**5-1**	**5-2**	**5-3**	**5-4**	**5-5**	**6-1**	**6-2**	**6-3**	**6-4**	**6-5**	**6-6**
HBKR	51	0.012	**0.877**	0.093	0.018	0.011	**0.851**	0.092	0.029	0.016	0.011	**0.825**	0.091	0.028	0.042	0.004
HDES	51	0.024	0.178	0.049	**0.748**	0.024	0.184	0.042	0.028	**0.722**	0.015	0.061	0.022	0.013	**0.877**	0.011
HWWC	77	0.039	**0.828**	0.101	0.033	0.041	**0.748**	0.081	0.096	0.034	0.037	**0.718**	0.076	0.088	0.074	0.006
HYAM	7	0.013	**0.713**	0.224	0.051	0.013	**0.518**	0.121	0.287	0.061	0.012	**0.496**	0.119	0.272	0.095	0.006
H30M	11	0.448	0.439	0.072	0.041	0.461	0.376	0.061	0.064	0.039	0.445	0.361	0.056	0.055	0.071	0.012
HLCR	77	**0.969**	0.009	0.006	0.015	**0.962**	0.009	0.006	0.007	0.015	**0.961**	0.009	0.005	0.006	0.015	0.005
HLAH	8	**0.965**	0.011	0.021	0.003	**0.959**	0.008	0.021	0.007	0.005	**0.958**	0.008	0.020	0.006	0.006	0.003
HSHN	26	**0.973**	0.009	0.007	0.011	**0.959**	0.008	0.021	0.006	0.011	**0.962**	0.008	0.006	0.005	0.012	0.007
HSAA	7	**0.985**	0.006	0.004	0.005	**0.981**	0.006	0.004	0.004	0.005	**0.981**	0.005	0.004	0.004	0.005	0.001
HRAN	80	**0.955**	0.019	0.012	0.014	**0.947**	0.015	0.009	0.014	0.014	**0.945**	0.014	0.008	0.012	0.012	0.007
HHAV	11	**0.956**	0.032	0.007	0.006	**0.947**	0.031	0.005	0.011	0.006	**0.947**	0.026	0.005	0.011	0.009	0.001
HKAN	4	**0.932**	0.023	0.011	0.034	**0.916**	0.021	0.006	0.021	0.035	**0.915**	0.021	0.006	0.016	0.022	0.021
HWGC	10	**0.809**	0.163	0.007	0.021	**0.801**	0.015	0.007	0.006	0.021	**0.794**	0.173	0.007	0.006	0.011	0.009
RBKR	43	0.012	0.043	**0.937**	0.008	0.011	0.041	**0.901**	0.036	0.011	0.011	0.044	**0.899**	0.033	0.009	0.004
RC15	31	0.018	0.040	**0.933**	0.009	0.019	0.033	**0.863**	0.075	0.009	0.018	0.033	**0.862**	0.071	0.012	0.004
RDES	19	0.023	0.103	0.027	**0.847**	0.021	0.071	0.011	0.073	**0.823**	0.015	0.019	0.009	0.062	**0.834**	0.061
RWWC	59	0.013	0.143	**0.823**	0.021	0.014	0.133	**0.778**	0.053	0.021	0.013	0.132	**0.777**	0.051	0.022	0.004
RYAM	44	0.031	0.133	**0.821**	0.006	0.031	0.017	0.141	**0.804**	0.007	0.029	0.018	0.139	**0.802**	0.009	0.003
RYDT	5	0.024	0.349	**0.621**	0.006	0.011	0.016	0.018	**0.950**	0.006	0.009	0.016	0.017	**0.947**	0.007	0.004
RLYC	7	0.021	0.196	**0.774**	0.009	0.011	0.034	0.035	**0.913**	0.006	0.011	0.038	0.032	**0.908**	0.011	0.002
BBTC	15	0.003	0.003	0.002	**0.992**	0.003	0.003	0.002	0.003	**0.989**	0.003	0.004	0.002	0.002	0.005	**0.985**

Table 12. Grouping of 21 populations into six clusters as suggested by STRUCTURE analysis. Listed for each cluster is cluster number (as per Table 10), descriptive acronym and populations grouped into this cluster. Population abbreviations are in Table 1, and probability of assignment for each individual is depicted in Figure 8.

Hierarchical Structure		
Number	Acronym	Pop
Cluster 6-1	**GC-H**	H30m
		HLCR
		HLAH
		HSHN
		HSAA
		HRAN
		HHAV
		HKAN
		HWGC
Cluster 6-2	**UB-H**	HBKR
		HWWC
		HYAM
Cluster 6-3	**UB-R**	RBKR
		RC15
		RWWC
Cluster 6-4	**YR-R**	RLYC
		RYDT
		RYAM
Cluster 6-5	**DC**	HDES
		RDES
Cluster 6-6	**BT**	BBTC

Table 13. Hierarchical analysis of molecular variance (AMOVA) based on 16 microsatellite loci genotyped across 21 populations. Listed are test statistics, including variance (= *V*), genotypic correlation at corresponding level (F) and test statistic (Value). Also listed are probabilities of a more extreme variance component than that observed. Hierarchical structure is detailed in Table 11.

Source of Variation	**d.f.**	**Sum of Squares**	**Variance Component**			**Fixation Indices**	
			V	%	F	Value	*p*
Among clusters	5	334.9	Va	0.282	4.16	Fct	0.0416	<0.0000
Among populations/ within clusters	15	144.9	Vb	0.063	0.93	Fsc	0.0097	<0.0000
Within populations	1265	8146.6	Vc	6.440	94.91	Fst	0.0509	<0.0000

Table 14. Pairwise F_{ST} values calculated among the six clusters identified by STRUCTURE analysis. Populations contained in each cluster are detailed in Table 11. All values are significant at $p<0.00001$ (as determined via permutation).

	GC-H	UB-H	UB-R	YR-R	DC	BT
GC-H	0.000					
UB-H	0.044	0.000				
UB-R	0.062	0.016	0.000			
YR-R	0.059	0.026	0.024	0.000		
DC	0.061	0.025	0.045	0.046	0.000	
BT	0.218	0.204	0.223	0.233	0.201	0.000

Table 15. Average pair-wise differences calculated among six clusters as identified by STRUCTURE analysis. Listed are: above diagonal = average number of pair-wise differences between clusters (PiXY); diagonal = average number of pair-wise differences within populations (PiX); below diagonal = corrected average pair-wise differences (PiXY-(PiX+PiY)/2). Populations contained in each cluster are detailed in Table 11. All values are significant at $p<0.00001$ (as determined via permutation).

	GC-H	UB-H	UB-R	YR-R	DC	BT
GC-H	12.693	13.591	13.610	13.466	13.697	15.091
UB-H	0.592	13.306	13.290	13.332	13.521	15.254
UB-R	0.842	0.215	12.843	13.070	13.554	15.285
YR-R	0.790	0.350	0.319	12.659	13.476	15.272
DC	0.823	0.340	0.605	0.619	13.055	14.945
BT	3.534	3.390	3.653	3.732	3.208	10.421

Appendix 1. Distribution of mtDNA ATPase 8 & 6 haplotypes calculated by species/basin and across populations (=CODE; acronyms defined in Table 1). For each population, numbers of individuals per haplotype are listed, as well as total number of individuals in that population (N). Numbers of individuals analyzed per species/basin in parentheses. HBC-UB= *Gila cypha* from upper basin; HBC-GC= *G. cypha* from lower basin Grand Canyon; RTC-UB= *G. robusta* from upper basin. Vertical grey bars designate *G. elegans* haplotypes.

SPECIES/BASIN	CODE	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	N
HBC-UB	**HBKR**	7	7	1	1														16
(62)	**HDES**	19			1	1													21
	HWWC	10	5	1	3		1												20
	HYAM	4	1																5
HBC-GC	**H30M**	2	3			1		2	1										9
(152)	**HLCR**	2	23			2		19											46
	HLAH		4					3		1									8
	HSHN		12					10			2								24
	HSAA		2					4											6
	HRAN	1	26			1		15				1							44
	HKAN		1					2											3
	HHAV		3					5		1									9
	HWGC	1	1					1											3
RTC-UB	**RBKR**	13	3		3														19
(117)	**RC15**	11	3		2		1						3						20
	RDES	17			2									1	1	1			22
	RWWC	12	6		3														21
	RYAM	9	10		1														20
	RYDT	1	2																3
	RLYC	1	4							1									7
	WYO	4																1	5
	Total	114	116	2	16	5	2	61	1	3	2	1	3	1	1	1	0	1	331

Appendix 2. Distribution of mtDNA ND2 haplotypes calculated by species/basin and across populations (=CODE; acronyms defined in Table 1). For each population, numbers of individuals per haplotype are listed, as well as total number of individuals in that population (N). Numbers of individuals analyzed per species/basin in parentheses. HBC-UB= *Gila cypha* from upper basin; HBC-GC= *G. cypha* from lower basin Grand Canyon; RTC-UB= *G. robusta* from upper basin. Vertical grey bars designate *G. elegans* haplotypes.

SPECIES/ BASIN	CODE	N 1	N 2	N 3	N 4	N 5	N 6	N 7	N 8	N 9	N 10	N 11	N 12	N 13	N 14	N 15	N 16	N 17	N 18	N 19	N 20	N 21	N 22	N 23	N
HBC-UB	**HBKR**	13	1	1	1																				16
(62)	**HDES**	20				1																			21
	HWWC	19	1																						20
	HYAM	3					1	1																	5
HBC-LB	**H30M**	7							1	1															9
(152)	**HLCR**	32								2	11	1													46
	HLAH	3									4		1												8
	HSHN	18									6														24
	HSAA	5									1														6
	HRAN	26								1	16			1											44
	HKAN	1									2														3
	HHAV	6									2		1												9
	HWGC	3																							3
RTC-UB	**RBKR**	14													2	3									19
(117)	**RC15**	17			1												1	1							20
	RDES	21												1											22
	RWWC	14													1	2	2	1	1						21
	RYAM	18																				1	1		20
	RYDT	2																						1	3
	RLYC	5					1						1												7
	RWYO	3																		1	1				5
	Total	250	2	1	2	1	2	1	1	4	42	1	3	2	3	5	3	2	1	1	1	1	1	1	331

Appendix 3. Distribution of mtDNA Dloop haplotypes calculated by species/basin and across populations (=CODE; acronyms defined in Table 1). For each population, numbers of individuals per haplotype are listed, as well as total number of individuals in that population (N). Numbers of individuals analyzed per species/basin in parentheses. HBC-UB= *Gila cypha* from upper basin; HBC-GC= *G. cypha* from lower basin Grand Canyon; RTC-UB= *G. robusta* from upper basin. Vertical grey bars designate *G. elegans* haplotypes.

SPECIES/ BASIN	CODE	D 1	D 2	D 3	D 4	D 5	D 6	D 7	D 8	D 9	D 10	D 11	D 12	D 13	D 14	D 15	D 16	D 17	D 18	D 19	D 20	D 21	D 22	D 23	N
HBC-UB	**HBKR**	2	11	2	1																				16
(62)	**HDES**		17			3	1																		21
	HWWC		17	1	1			1																	20
	HYAM		4						1																5
HBC-GC	**H30M**		6							1	1	1													9
(152)	**HLCR**		28	1						7		2	6	1	1										46
	HLAH		4										4												8
	HSHN		13							4			5			1	1								24
	HSAA		6																						6
	HRAN		29									1	13					1							44
	HKAN		1										2												3
	HHAV		8										1												9
	HWGC		3																						3
RTC-UB	**RBKR**		13																2	3	1				19
(117)	**RC15**		20																						20
	RDES		20			1																1			22
	RWWC		16	1															1	3					21
	RYAM		12	2	3											1							1	1	20
	RYDT		3																						3
	RLYC		4	1	1		1																		7
	RWYO		5																						5
	Total	2	240	8	6	4	2	1	1	12	1	4	31	1	1	2	1	1	3	6	1	1	1	1	331

Appendix 4. Distribution of composite haplotypes for combined mtDNA regions (calculated by species/basin and across populations (=CODE; acronyms defined in Table 1). For each population, numbers of individuals per haplotype are listed, as well as total number of individuals in that population (N). Upper panel shows *Gila cypha* from upper basin; middle panel *G. cypha* from lower basin Grand Canyon; and lower panel *G. robusta* from upper basin. Vertical grey bars designate *G. elegans* haplotypes.

CODE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
HBKR	2	5	1	4	1	1	1	1																							
HDES		16							3	1	1																				
HWWC		10	1	3	1	1				2		1	1																		
HYAM		2												1	1	1															
H30M		2		2													2	1	1	1											
HLCR			1	4													18	7		2	5	6	1	1	1						
HLAH																	3					4				1					
HSHN				2													10	4			1	5					1	1			
HSAA				1													4				1										
HRAN		1		8													15			1	5	11							2	1	
HKAN																	1					1									1
HHAV				1													5				1	1				1					
HWGC		1		1													1														
RBKR		10		3																											
RC15		9		3				1		1			1																		
RDES		16							1	2																					
RWWC		9	1	4																											
RYAM		6	2	5																											
RYDT		1		1																											
RLYC		1	1	1											1																
RWYO		2																													
Total	2	91	7	43	2	2	1	2	4	6	1	1	2	1	2	1	59	12	1	4	13	28	1	1	1	2	1	1	2	1	1

Appendix 4. continued

CODE	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	N
HBKR																								**16**
HDES																								**21**
HWWC																								**20**
HYAM																								**5**
H30M																								**9**
HLCR																								**46**
HLAH																								**8**
HSHN																								**24**
HSAA																								**6**
HRAN																								**44**
HKAN																								**3**
HHAV																								**9**
HWGC																								**3**
RBKR	2	3	1																					**19**
RC15				3	1	1																		**20**
RDES							1	1	1															**22**
RWWC	1	2			1	1							1	1										**21**
RYAM										2								1	1	1	1	1		**20**
RYDT																							1	**3**
RLYC										1	1	1												**7**
RWYO															1	1	1							**5**
Total	3	5	1	3	2	2	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	**331**

Appendix 5. Allelic diversity for 21 chub populations (= Pop) genotyped at 16 microsatellite loci. Listed for each population and locus are number of alleles (= Na), size of smallest (= Min) and largest (= Max) allele (in base-pairs), and variance in repeat number (= Var_{RP}). Sample sizes for each population are given in parentheses. Population acronyms are as in Table 1. Values for BBTC (*G. elegans*) are for comparative purpose only, since the sample is hatchery-derived.

Pop.		G38	G13	G34	G39	C01	G02	G294	Gel14	C03	C12	G87	G99	Gel16	Gel06	Gel11	Gel09
HBKR	**Na**	19	17	24	12	26	10	8	2	24	17	21	26	24	25	15	20
(n=51)	**Min**	299	219	208	214	128	213	213	74	245	210	221	337	160	138	163	200
	Max	447	303	324	282	212	293	253	86	393	278	329	521	292	290	247	284
	Var_{RP}	33.7	19.6	57.2	9.2	64.0	5.7	4.7	0.6	32.2	13.0	31.9	70.8	32.5	46.9	15.2	20.3
HDES	**Na**	23	20	23	12	18	10	4	2	20	14	19	28	21	24	15	19
(n=51)	**Min**	295	219	204	214	118	213	221	74	253	210	233	341	168	122	155	200
	Max	419	319	316	278	212	285	237	86	349	266	309	553	288	242	235	284
	Var_{RP}	51.5	23.7	43.9	5.9	50.2	12.3	0.9	1.4	31.8	7.9	21.8	147.0	33.1	56.0	11.8	12.2
HWWC	**Na**	23	19	28	12	29	11	9	2	27	16	24	31	25	29	19	18
(n=77)	**Min**	331	215	204	214	118	213	213	74	245	206	221	337	196	130	163	200
	Max	479	311	320	274	212	285	257	86	405	274	333	529	300	254	263	284
	Var_{RP}	32.6	14.0	54.4	9.4	82.9	5.3	3.7	0.3	44.7	14.5	29.5	70.7	26.4	39.0	19.5	15.3
HYAM	**Na**	8	8	10	7	9	6	2	2	9	9	6	8	9	6	8	9
(n=7)	**Min**	351	219	208	226	136	213	221	74	257	206	241	357	200	170	171	216
	Max	415	287	304	274	206	241	225	86	357	254	285	433	300	226	267	280
	Var_{RP}	15.7	31.0	70.6	16.5	91.2	3.9	0.1	0.6	56.3	14.9	21.3	35.6	38.0	18.2	34.4	35.2
H30M	**Na**	13	10	11	5	8	8	5	3	12	11	10	12	12	9	3	7
(n=11)	**Min**	303	219	208	214	118	221	221	74	253	202	249	349	220	142	183	212
	Max	447	299	264	238	186	285	249	90	341	250	321	513	272	202	195	264
	Var_{RP}	77.8	19.0	25.8	2.2	100.2	14.5	4.2	2.0	48.3	12.7	17.7	123.3	25.1	12.3	1.3	8.6

Appendix 5 continued

HLCR	**Na**	18	12	19	10	24	12	5	3	18	11	19	23	15	18	10	11
(n=77)	**Min**	303	219	204	214	118	217	221	74	257	202	229	349	212	126	167	212
	Max	451	323	292	278	200	285	253	90	369	254	337	533	276	202	219	272
	Var_{RP}	81.4	22.3	35.3	7.1	102.3	20.3	4.6	1.7	35.3	7.7	25.6	115.0	18.7	15.2	7.3	9.6
HLAH	**Na**	7	7	9	5	11	9	2	2	10	5	4	8	6	8	6	8
(n=8)	**Min**	303	219	228	226	130	221	221	74	261	210	253	357	224	134	175	212
	Max	415	327	288	254	186	285	241	86	325	234	321	537	276	202	219	264
	Var_{RP}	67.7	37.2	21.9	5.0	128.8	27.0	1.6	1.5	29.7	6.3	17.4	176.8	22.6	15.3	10.0	11.3
HSHN	**Na**	14	10	18	7	18	8	3	3	13	10	10	14	12	12	8	9
(n=26)	**Min**	303	223	204	214	118	221	221	74	257	206	229	349	212	130	167	212
	Max	447	323	296	254	196	261	249	90	333	254	321	525	276	190	219	264
	Var_{RP}	95.2	30.6	39.8	5.5	139.6	11.9	6.0	2.6	18.7	10.2	26.4	116.5	15.1	10.1	7.3	9.4
HSAA	**Na**	9	6	9	5	9	7	3	1	8	6	7	8	8	8	4	5
(n=7)	**Min**	303	235	208	226	122	221	221	74	261	210	249	357	216	134	175	220
	Max	415	323	288	278	186	261	241	74	325	234	321	505	272	190	195	264
	Var_{RP}	63.2	32.4	55.3	13.3	137.0	8.3	1.8	0.0	22.2	4.5	21.7	90.7	15.8	18.6	1.6	9.2
HRAN	**Na**	17	13	21	11	20	11	5	3	20	12	18	23	17	20	10	10
(n=80)	**Min**	303	219	204	214	118	221	221	74	245	202	201	341	212	130	167	212
	Max	467	327	296	278	200	285	253	90	381	254	333	533	304	218	219	264
	Var_{RP}	87.6	15.6	34.9	8.0	140.6	16.4	4.8	1.3	27.6	10.5	56.5	119.3	23.6	22.7	7.3	6.4
HHAV	**Na**	10	5	12	6	9	9	3	3	9	7	9	10	8	7	8	9
(n=11)	**Min**	303	223	208	222	118	221	221	74	261	206	249	357	228	138	167	212
	Max	447	259	292	278	186	285	249	90	313	234	321	497	272	194	219	272
	Var_{RP}	51.8	6.6	44.2	13.3	136.0	18.5	5.5	2.9	18.1	5.0	27.1	61.5	15.4	8.3	6.8	19.3
HKAN	**Na**	6	4	5	5	5	8	2	1	7	5	4	6	5	6	3	6
(n=4)	**Min**	335	235	224	226	138	221	221	74	257	210	241	357	220	154	167	220
	Max	447	251	292	254	186	265	253	74	305	234	305	501	276	194	187	260
	Var_{RP}	98.8	2.1	28.2	4.2	88.0	18.3	8.0	0.0	16.6	5.8	42.0	232.8	24.8	9.4	6.3	11.7

Appendix 5 continued

HWGC	**Na**	8	6	11	7	12	8	5	2	11	6	5	9	9	8	5	8
(n=10)	**Min**	303	223	208	214	122	217	221	74	257	210	237	353	216	154	167	208
	Max	403	275	292	258	186	257	249	86	329	250	289	489	272	190	191	260
	Var_{RP}	94.7	9.5	35.3	8.7	76.8	15.4	7.7	1.3	15.0	7.2	7.2	64.4	15.7	5.4	2.6	8.1
RBKR	**Na**	14	15	21	10	12	8	5	2	18	13	20	20	22	31	17	17
(n=43)	**Min**	303	223	204	222	130	213	221	74	245	214	233	357	196	146	167	204
	Max	403	319	312	270	180	285	257	86	341	262	337	509	304	298	239	288
	Var_{RP}	13.0	12.5	95.9	7.0	12.5	4.4	1.8	0.7	15.2	9.6	31.0	43.0	29.9	85.2	19.0	23.5
RC15	**Na**	17	11	23	10	13	7	3	2	14	14	18	19	21	25	15	16
(n=31)	**Min**	303	219	204	218	128	213	221	74	253	206	233	357	212	154	167	204
	Max	419	275	312	274	154	241	249	86	329	262	321	497	308	302	239	280
	Var_{RP}	23.6	7.0	77.5	6.5	8.6	1.7	2.6	0.3	9.9	10.8	28.3	35.7	31.8	57.6	18.3	20.3
RDES	**Na**	17	15	19	6	13	11	6	2	13	9	14	20	17	17	11	13
(n=19)	**Min**	303	223	208	226	118	213	221	74	257	214	237	341	168	130	155	200
	Max	411	307	308	274	158	281	241	86	321	254	317	537	280	218	227	292
	Var_{RP}	43.8	23.5	38.9	4.7	27.6	14.6	2.1	1.5	15.3	7.1	24.7	203.8	44.5	44.0	10.1	17.9
RWWC	**Na**	20	20	28	10	17	11	5	2	17	16	22	25	23	32	20	20
(n=59)	**Min**	303	223	204	222	118	213	221	74	257	206	221	345	196	138	159	208
	Max	419	319	328	270	202	265	249	86	341	266	325	517	304	278	243	292
	Var_{RP}	22.9	18.6	95.0	5.3	23.0	4.0	2.3	0.8	15.5	10.2	28.4	59.6	31.9	61.3	19.5	32.2
RYAM	**Na**	21	15	25	7	15	11	2	3	27	17	19	23	25	28	15	15
(n=44)	**Min**	347	223	208	222	122	213	221	74	249	198	209	357	200	138	175	208
	Max	455	303	332	262	166	257	225	90	385	262	325	505	320	290	235	276
	Var_{RP}	34.3	16.7	85.3	2.6	20.5	5.6	0.0	0.3	45.0	14.1	24.6	38.7	35.3	70.4	9.8	15.4
RYDT	**Na**	6	8	7	5	6	5	1	2	7	8	7	7	8	8	6	7
(n=5)	**Min**	351	219	208	226	128	213	221	74	265	198	237	365	200	162	191	212
	Max	379	263	304	242	154	241	221	86	353	242	285	461	280	262	239	256
	Var_{RP}	5.8	11.9	92.9	2.5	23.5	4.5	0.0	0.9	76.5	17.2	16.3	69.6	32.0	74.9	13.9	15.4

Appendix 5 continued

RLYC	**Na**	9	7	8	5	7	8	1	2	11	8	10	10	8	11	8	7
(n=7)	**Min**	347	219	208	226	122	213	221	74	257	202	209	345	232	162	175	212
	Max	399	259	300	262	146	245	221	86	349	262	305	433	304	238	211	248
	Var_{RP}	13.0	7.3	64.8	4.8	15.0	5.0	0.0	0.6	58.6	27.3	82.7	38.5	21.7	36.9	11.4	7.3
BBTC	**Na**	6	7	1	6	4	6	5	1	8	6	7	13	6	7	4	8
(n=15)	**Min**	303	251	204	238	118	245	221	86	257	202	249	381	168	126	183	228
	Max	347	299	204	286	138	281	241	86	329	230	317	561	268	194	195	292
	Var_{RP}	8.2	20.1	0.0	25.2	4.0	12.1	2.8	0.0	23.5	2.8	46.6	175.3	72.6	31.1	1.6	15.3

Appendix 6. Observed (*Ho*) and expected (*He*) heterozygosity and fixation index (F) averaged across loci (= Ave) for 21 chub populations (= Pop) genotyped at 16 microsatellite loci. Significant deviations from Hardy-Weinberg equilibrium are indicated in bold type. Population acronyms are provided in Table 1. Sample sizes are in parenthesis below population acronym.

Pop		G38	G13	G34	G39	C01	G02	G294	Gel14	C03	C12	G87	G99	Gel16	Gel06	Gel11	Gel09	Ave.
HBKR	**Ho**	0.941	0.882	0.863	0.843	0.863	0.824	0.471	0.137	0.922	0.863	0.784	0.961	0.980	0.922	0.902	0.882	0.815
(n=51)	**He**	0.928	0.870	0.942	0.824	0.908	0.813	0.456	0.128	0.931	0.866	0.932	0.938	0.930	0.943	0.866	0.917	0.825
	F	-0.015	-0.014	0.084	-0.023	0.050	-0.013	-0.031	-0.074	0.010	0.004	0.158	-0.024	-0.054	0.023	-0.041	0.038	0.005
HDES	**Ho**	0.902	0.980	0.863	0.765	0.922	0.882	0.235	0.314	0.980	0.863	0.843	0.961	1.000	0.961	0.843	0.863	0.824
(n=51)	**He**	0.900	0.899	0.926	0.796	0.876	0.810	0.249	0.315	0.921	0.865	0.881	0.927	0.931	0.939	0.863	0.800	0.806
	F	-0.002	-0.090	0.068	0.039	-0.052	-0.090	0.055	0.005	-0.064	0.002	0.043	-0.036	-0.074	-0.023	0.023	-0.079	-0.017
HWWC	**Ho**	0.896	0.909	0.857	0.792	0.948	0.831	0.403	0.078	0.896	0.922	0.857	0.948	0.922	0.896	0.870	0.909	0.808
(n=77)	**He**	0.931	0.896	0.944	0.806	0.933	0.822	0.392	0.075	0.932	0.889	0.935	0.939	0.936	0.945	0.895	0.904	0.823
	F	0.037	-0.015	0.092	0.018	-0.016	-0.012	-0.026	-0.041	0.039	-0.038	0.083	-0.010	0.015	0.052	0.028	-0.005	0.013
HYAM	**Ho**	0.857	0.857	1.000	1.000	0.857	1.000	0.286	0.143	0.857	1.000	0.571	1.000	0.714	0.571	1.000	1.000	0.795
(n=7)	**He**	0.827	0.847	0.857	0.796	0.837	0.796	0.245	0.133	0.847	0.837	0.755	0.837	0.816	0.786	0.806	0.857	0.742
	F	-0.037	-0.012	-0.167	-0.256	-0.024	-0.256	-0.167	-0.077	-0.012	-0.195	0.243	-0.195	0.125	0.273	-0.241	-0.167	-0.073
H30M	**Ho**	0.909	0.909	0.727	0.818	0.727	0.909	0.818	0.364	0.818	0.909	0.727	0.818	1.000	0.818	0.727	0.818	0.801
(n=11)	**He**	0.901	0.855	0.835	0.740	0.818	0.756	0.636	0.492	0.901	0.864	0.756	0.888	0.860	0.818	0.616	0.781	0.782
	F	-0.009	-0.063	0.129	-0.106	0.111	-0.202	-0.286	0.261	0.092	-0.053	0.038	0.079	-0.163	0.000	-0.181	-0.048	-0.025
HLCR	**Ho**	0.896	0.831	0.792	0.701	0.909	0.857	0.338	0.286	**0.675**	0.883	0.779	0.896	0.909	0.896	0.779	0.883	0.769
(n=77)	**He**	0.881	0.837	0.876	0.727	0.914	0.867	0.310	0.304	0.902	0.813	0.814	0.845	0.869	0.883	0.818	0.849	0.782
	F	-0.017	0.006	0.096	0.036	0.005	0.012	-0.089	0.059	0.251	-0.086	0.043	-0.061	-0.046	-0.015	0.047	-0.040	0.013
HLAH	**Ho**	0.750	1.000	1.000	0.875	1.000	1.000	0.125	0.375	0.875	0.875	0.500	0.875	0.875	0.875	0.875	0.875	0.797
(n=8)	**He**	0.828	0.789	0.867	0.758	0.852	0.875	0.117	0.305	0.875	0.766	0.609	0.813	0.805	0.852	0.797	0.805	0.732
	F	0.094	-0.267	-0.153	-0.155	-0.174	-0.143	-0.067	-0.231	0.000	-0.143	0.179	-0.077	-0.087	-0.028	-0.098	-0.087	-0.090

Appendix 6 (continued)

HSHN	**Ho**	0.923	0.923	0.846	0.769	0.846	0.923	0.385	0.462	0.692	0.885	0.769	0.923	0.962	0.731	0.846	0.846	0.796
(n=26)	**He**	0.881	0.817	0.897	0.721	0.888	0.754	0.379	0.504	0.888	0.846	0.792	0.846	0.859	0.831	0.757	0.862	0.783
	F	-0.048	-0.129	0.057	-0.067	0.047	-0.224	-0.016	0.084	0.220	-0.045	0.029	-0.091	-0.119	0.121	-0.117	0.019	-0.017
HSAA	**Ho**	1.000	0.857	1.000	0.857	0.714	1.000	0.286	0.000	0.571	0.714	0.571	1.000	0.714	1.000	0.571	0.714	0.723
(n=7)	**He**	0.878	0.755	0.847	0.745	0.847	0.816	0.255	0.000	0.827	0.755	0.796	0.806	0.847	0.837	0.643	0.735	0.712
	F	-0.140	-0.135	-0.181	-0.151	0.157	-0.225	-0.120	-	0.309	0.054	0.282	-0.241	0.157	-0.195	0.111	0.028	-0.019
HRAN	**Ho**	0.888	0.775	**0.725**	0.775	0.863	0.800	0.338	0.200	**0.675**	0.825	**0.738**	0.850	0.913	0.913	0.838	0.775	0.743
(n=80)	**He**	0.893	0.779	0.871	0.755	0.882	0.833	0.317	0.226	0.896	0.845	0.851	0.854	0.893	0.915	0.821	0.804	0.777
	F	0.007	0.006	0.168	-0.026	0.022	0.040	-0.063	0.117	0.247	0.023	0.134	0.004	-0.021	0.003	-0.020	0.036	0.042
HHAV	**Ho**	0.909	0.909	0.909	0.818	1.000	0.727	0.455	0.545	0.727	0.818	0.909	0.909	0.818	0.818	0.909	0.818	0.813
(n=11)	**He**	0.831	0.760	0.901	0.777	0.843	0.822	0.368	0.545	0.847	0.785	0.777	0.826	0.831	0.798	0.822	0.860	0.775
	F	-0.095	-0.196	-0.009	-0.053	-0.186	0.116	-0.236	0.000	0.141	-0.042	-0.170	-0.100	0.015	-0.026	-0.106	0.048	-0.056
HKAN	**Ho**	1.000	0.750	0.750	0.750	1.000	1.000	0.250	0.000	1.000	0.500	0.250	0.750	1.000	1.000	0.250	1.000	0.703
(n=4)	**He**	0.813	0.719	0.688	0.688	0.688	0.875	0.219	0.000	0.844	0.781	0.719	0.781	0.688	0.813	0.594	0.813	0.670
	F	-0.231	-0.043	-0.091	-0.091	-0.455	-0.143	-0.143	-	-0.185	0.360	0.652	0.040	-0.455	-0.231	0.579	-0.231	-0.044
HWGC	**Ho**	1.000	0.900	1.000	0.700	0.800	1.000	0.600	0.300	0.700	0.900	0.700	0.700	0.900	1.000	0.800	0.800	0.800
(n=10)	**He**	0.795	0.710	0.850	0.800	0.835	0.835	0.595	0.555	0.830	0.710	0.620	0.710	0.785	0.805	0.705	0.710	0.741
	F	-0.258	-0.268	-0.176	0.125	0.042	-0.198	-0.008	0.459	0.157	-0.268	-0.129	0.014	-0.146	-0.242	-0.135	-0.127	-0.072
RBKR	**Ho**	0.814	0.907	0.907	0.767	0.791	0.884	0.186	0.163	0.907	0.907	0.907	0.930	0.930	0.930	0.884	0.907	0.795
(n=43)	**He**	0.835	0.850	0.898	0.705	0.840	0.796	0.175	0.150	0.902	0.874	0.923	0.920	0.932	0.951	0.913	0.913	0.786
	F	0.025	-0.067	-0.010	-0.088	0.059	-0.110	-0.065	-0.089	-0.005	-0.038	0.018	-0.011	0.002	0.022	0.032	0.006	-0.020
RC15	**Ho**	0.871	0.774	0.871	0.806	0.806	**0.806**	0.161	0.065	0.871	0.903	0.968	0.968	0.935	0.903	0.806	0.839	0.772
(n=31)	**He**	0.859	0.845	0.930	0.787	0.882	0.759	0.180	0.062	0.869	0.898	0.919	0.927	0.931	0.939	0.902	0.890	0.786
	F	-0.014	0.084	0.064	-0.024	0.086	-0.063	0.101	-0.033	-0.002	-0.006	-0.053	-0.044	-0.005	0.038	0.106	0.057	0.018

Appendix 6 (continued)

RDES	**Ho**	0.842	0.947	0.947	0.737	0.895	0.895	0.421	0.316	0.895	0.947	0.737	1.000	0.947	0.895	0.842	0.842	0.819
(n=19)	**He**	0.882	0.904	0.922	0.760	0.857	0.839	0.439	0.332	0.893	0.816	0.773	0.902	0.927	0.893	0.809	0.831	0.799
	F	0.046	-0.047	-0.027	0.031	-0.044	-0.066	0.041	0.050	-0.002	-0.161	0.047	-0.109	-0.022	-0.002	-0.041	-0.013	-0.020
RWWC	**Ho**	0.915	0.932	0.881	0.864	0.966	0.797	0.169	0.186	0.831	0.864	0.915	0.932	0.898	0.915	0.847	0.932	0.803
(n=59)	**He**	0.902	0.871	0.903	0.785	0.885	0.794	0.160	0.169	0.912	0.888	0.922	0.931	0.936	0.951	0.919	0.921	0.803
	F	-0.015	-0.070	0.024	-0.101	-0.092	-0.003	-0.061	-0.103	0.089	0.026	0.008	-0.002	0.040	0.037	0.078	-0.012	-0.010
RYAM	**Ho**	0.909	0.886	**0.841**	0.818	0.795	**0.977**	0.023	0.045	0.909	0.932	0.818	0.955	0.955	0.886	0.841	0.864	0.778
(n=44)	**He**	0.906	0.906	0.913	0.718	0.844	0.840	0.022	0.045	0.939	0.899	0.906	0.937	0.930	0.949	0.888	0.862	0.781
	F	-0.003	0.021	0.079	-0.140	0.057	-0.164	-0.011	-0.017	0.032	-0.037	0.097	-0.018	-0.026	0.066	0.053	-0.002	-0.001
RYDT	**Ho**	0.800	0.800	0.800	0.600	0.800	0.600	0.000	0.200	0.800	1.000	0.400	0.800	1.000	0.600	1.000	1.000	0.700
(n=5)	**He**	0.780	0.860	0.820	0.680	0.760	0.720	0.000	0.180	0.820	0.840	0.840	0.840	0.860	0.860	0.800	0.840	0.719
	F	-0.026	0.070	0.024	0.118	-0.053	0.167	-	-0.111	0.024	-0.190	0.524	0.048	-0.163	0.302	-0.250	-0.190	0.020
RLYC	**Ho**	0.857	0.714	0.857	0.714	0.857	1.000	0.000	0.143	1.000	1.000	1.000	1.000	0.714	0.714	0.857	0.714	0.759
(n=7)	**He**	0.878	0.816	0.837	0.704	0.755	0.827	0.000	0.133	0.898	0.857	0.878	0.888	0.847	0.888	0.806	0.796	0.738
	F	0.023	0.125	-0.024	-0.014	-0.135	-0.210	-	-0.077	-0.114	-0.167	-0.140	-0.126	0.157	0.195	-0.063	0.103	-0.031
BBTC	**Ho**	0.800	**0.800**	0.000	0.867	0.533	0.933	0.533	0.000	0.800	0.733	0.800	0.733	0.733	0.667	0.733	0.733	0.650
(n=15)	**He**	0.700	0.811	0.000	0.791	0.433	0.742	0.718	0.000	0.664	0.676	0.767	0.893	0.622	0.809	0.722	0.724	0.630
	F	-0.143	0.014	-	-0.096	-0.231	-0.257	0.257	-	-0.204	-0.086	-0.043	0.179	-0.179	0.176	-0.015	-0.012	-0.046