



Final Report

*Microsatellite and MHC Variation to Distinguish Natural Lineages of the Sonoran  
Topminnow*

Arizona Game and Fish Department  
Heritage Grant Award I98002

**Arizona State University Account WRT 2583**

Philip W. Hedrick

*Department of Biology  
Arizona State University  
Tempe, AZ 85287*

**DISCLAIMER**

The findings, opinions, and recommendations in this report are those of the investigators who have received partial or full funding from the Arizona Game and Fish Department Heritage Fund. The findings, opinions, and recommendations do not necessarily reflect those of the Arizona Game and Fish Commission or the Department, or necessarily represent official Department policy or management practice. For further information, please contact the Arizona Game and Fish Department.

## Introduction

For effective conservation of endangered species, it is first important to identify which taxa within closely related groups are different species. Next, identification of evolutionarily significant units (ESUs) within species, i.e., populations evolving independently of other populations (Waples 1995), is important so evolutionary important groups are preserved. Finally, it is important to identify management units (MUs) (Moritz 1994) within ESUs so that conservation actions may be implemented.

Generally, neutral genetic markers, such as microsatellite loci, are generally most appropriate to determine historical distinctiveness of populations and therefore the designation of species, ESUs, and MUs. However, the preservation of adaptive variation is fundamental to the conservation intentions behind designations of both ESUs and MUs. Therefore, here we examine both microsatellite variation and variation at a potentially adaptive locus in the major histocompatibility complex (MHC). Genes in the MHC appears significant in parasite resistance (Hedrick & Kim 2000) and may have a role in mate recognition (Penn & Potts 1999) and maternal-fetal interaction (Ober *et al.* 1998). If variation over groups is consistent for neutral markers and potentially adaptive genes, as those in the MHC, then this lends support for the designations of conservation units based on neutral markers.

The Sonoran topminnow, *Poeciliopsis occidentalis*, is a small, live-bearing fish that occurs in Arizona, United States and Sonora, Mexico. Two putative subspecies, Gila topminnow, *Poeciliopsis o. occidentalis*, and Yaqui topminnow, *P. o. sonoriensis*, both listed as endangered, exist in Arizona. The Gila topminnow was once one of the most abundant fishes in the Gila River drainage, but now exists naturally in the United States in only four isolated Arizona watersheds in eight populations. Others exist as a result of extensive restocking efforts, and at least one site on the Santa Cruz River appears to have been recently naturally re-colonized. The Yaqui topminnow was never widespread in the United States because the Rio Yaqui drainage includes only a small part of southeastern Arizona, now within the San Bernadino National Wildlife Refuge.

We developed molecular techniques to differentiate samples of Gila topminnow from the four major watersheds (Bylas Spring 2, Cienega Creek, Monkey Spring, and Sharp Spring) using microsatellite loci (Parker *et al.* 1999) and a major histocompatibility

complex (MHC) locus (Hedrick & Parker 1998). In the present research, we further examine genetic variation within all natural Gila topminnow populations in Arizona (and several stocked populations) and explore the relationship between Gila and Yaqui topminnows. With this detailed analysis of both neutral and adaptive variation over these populations, we have obtained a comprehensive picture of genetic variation that can be used to define species, ESUs, and MUs in this group.

## **Materials and methods**

### *Samples*

Nine additional Gila topminnow samples are examined in this study and they, with the four populations from our earlier study, include samples from all known natural populations in Arizona. The additional natural populations (collected primarily by David Weedman, Arizona Game and Fish Department) are Bylas Spring 1, Cottonwood Spring, Red Rock at Cott Tank, Red Rock Falls, Coalmine Creek, and Sonoita Creek below Patagonia Lake (see Figure 1). In addition, we examined a sample from a Santa Cruz River site (downstream from the Nogales Water Treatment plant outflow) that has been naturally re-colonized in recent years from an unknown source(s). A sample from Boyce Thompson Arboretum, extensively used for restocking, was also examined. This population was thought to have been established from Monkey Spring fish, and possibly Bylas Spring and/or an extinct Coscio Wash population. In addition, a sample was examined from Watson Wash, which is a warm spring near Safford, Arizona that was stocked from an unknown source. The two samples of Yaqui topminnow were from Tule Spring and North Pond, San Bernadino National Wildlife Refuge (collected by Kevin Cobble, US Fish and Wildlife Service).

[insert Figure 1]

### *Molecular and statistical techniques*

Parker *et al.* (1998) identified 10 microsatellite loci, five polymorphic in Gila topminnows. In addition, three loci monomorphic in Gila topminnows were polymorphic in Yaqui topminnows. One polymorphic Gila topminnow locus did not consistently amplify in Yaqui topminnow. We used primers and techniques, as given by Parker *et al.*

(1998, 1999), to determine the microsatellite genotypes. Molecular techniques (with the internal primer) outlined by Hedrick & Parker (1998a) were used to examine the new class II MHC sequences. At least two subclones containing the new sequences were analyzed on both strands on an Applied Biosystems 377 automated sequencer (division of Perkin-Elmer, Foster City, CA). Sequences were identical in all replicates and from both strands. Genotypes were determined using the SSCP (single-stranded confirmation polymorphism) approach as given by Hedrick & Parker (1998a).

We calculated allele frequencies using the gene-counting approach and used the small sample size correction in calculating the expected heterozygosity (Nei 1987). Observed heterozygosity was tested for differences from the expected using  $\chi^2$  test with the standard Bonferroni correction for multiple comparisons over loci and populations. We calculated  $D_A$ , the minimum genetic distance, between pairs of populations (Nei 1987).  $D_A$  is an appropriate measure to use when, as in our case, populations are highly differentiated from each other for some loci. To align MHC sequences, the sequence editor ESEE version 3.25 was used. DISPAN (Ota 1993) was used to construct the neighbor-joining tree and run 1000 bootstrap replicates for the Gila topminnow population data. MEGA (Kumar *et al.* 1993) was used to construct a neighbor-joining tree using the genetic distance of Jukes & Cantor (1969) for the sequence data. MEGA was also used to calculate the relative rate of nonsynonymous and synonymous substitutions according to Nei & Gojobori (1986), applying the correction of Jukes & Cantor (1969) for multiple hits.

## Results

### *Gila Topminnow*

Table 1 gives the microsatellite and MHC allele frequencies in the 13 populations of Gila topminnow and average observed heterozygosity, expected heterozygosity, and observed number of alleles. Even without (with) a standard Bonferroni correction for multiple tests, only one observed heterozygosity was statistically different from expected. This exception was the highly significant deficiency for the MHC locus in the sample from Sharp Spring, also found earlier to be significant by Hedrick & Parker (1998a). These two samples are the same fish but the genotypes were scored differently because we have

now identified a second duplicated locus in Sharp Spring and five other natural populations (see below). For the primary MHC locus, there is a 43.5% deficiency of heterozygotes compared to expected. Because this deficiency is not present at the microsatellite loci for the same fish, we assume the deficiency is the result of the presence of a null allele (Hedrick & Parker 1998a). Using the approach of Brookfield (1996), the estimated frequency of the null allele is 0.180, somewhat higher than the 0.147 estimated with this approach before (Hedrick & Parker 1998b).

[insert Table 1]

As discussed by Parker *et al.* (1999), the pattern of genetic variation over the four initial populations was similar for the microsatellite loci and the MHC locus. The product-moment correlation coefficient ( $r$ ) between population measures of genetic variation for the microsatellite loci and the MHC locus calculated for the 13 samples in Table 1 also illustrate this; for  $H_O$ ,  $r = 0.75$ ; for  $H_E$ ,  $r = 0.73$ ; and for  $n$ ,  $r = 0.84$  ( $p < 0.01$  in all cases). In other words, it appears that the most significant factors influencing allele frequencies between populations are similar for both microsatellite loci and the MHC locus. This implies that non-selective factors, such as bottlenecks causing low variation in some populations (see below), may be important for the MHC locus as well as the microsatellite loci. In any case, the high correlation provides support to use the MHC locus in addition to the microsatellite loci in our following analysis of the relationships between the different populations.

Figure 2 gives a neighbor-joining tree for the 13 populations based on the data in Table 1. The natural populations appear to fall into five groups as supported by the bootstrap values given (Table 2 gives the genetic distances between the groups). First, the two populations Bylas Spring 1 and 2 are quite similar to each other ( $D_A = 0.093$  and cluster together with a 97% bootstrap probability) but are different from other populations. They also have low genetic variation as measured by both heterozygosity and numbers of alleles. Second, Monkey and Cottonwood Springs are similar to each other ( $D_A = 0.300$  and cluster together with a 73% bootstrap probability when the non-native Boyce-Thompson and Watson Wash sample are not included) but different from all other natural populations. They both have intermediate levels of genetic variation. Third, samples from the Sonoita Creek drainage, Sonoita Creek, Coalmine Canyon, Red

Rock at Cott Tank, and Red Rock Falls, cluster together (the average  $D_A$  between pairs in this group is 0.112 and they cluster together with a 89% bootstrap probability when the recolonized Santa Cruz population is excluded). Except for Red Rock at Cott tank, variable for only one microsatellite locus, these populations are intermediate in genetic variation. Finally, Cienega Creek and Sharp Spring form the fourth and fifth groups. These populations have nearly non-overlapping sets of alleles for loci *C-15* and *LL53* and share no alleles at the MHC locus. Sharp Spring has higher-than-average genetic variation and Cienega Creek has intermediate genetic variation. In other words, the five groups differ significantly from each other based on the bootstrap values, their genetic distances (Table 2), and overall pattern of genetic variation for these genetic markers.

[insert Figure 2]

[insert Table 2]

The samples from Watson Wash and Boyce Thompson are very similar to one another ( $D_A = 0.035$ ), suggesting Boyce Thompson as the source for the Watson Wash population. Further, Boyce Thompson fish are similar to Monkey Spring ( $D_A = 0.134$ ). In particular, because there is no overlap in alleles from Boyce Thompson and the Bylas Spring samples for locus *6-10* and the MHC gene and the very large overall genetic distance, it is unlikely that there is Bylas ancestry in the Boyce Thompson stock. The Boyce Thompson stock does have a new MHC allele, *Pooc-16*, found only as a heterozygote in one individual and not found in any other sample. However, this new allele differs by only one amino acid (one nucleotide) from the most common allele in Monkey Spring (see below).

Finally, the sample of the naturally recolonized population from the Santa Cruz River clusters with the four nearby Sonoita Creek samples (average  $D_A = 0.176$ ). The Santa Cruz River sample is, however, the most variable analyzed with variation at all microsatellite loci and the MHC locus and has both the highest heterozygosity and number of alleles of all samples. It has several alleles also found in Sharp Spring but it is not clear how much ancestry is from upper Santa Cruz stocks. In addition, it has two MHC alleles (*Pooc-15* and *Pooc-17*) not found in other samples. These are very similar, however, to already described alleles, *Pooc-15* is one amino acid (two nucleotides)

different from *Pooc-5*, and *Pooc-17* differs by one amino acid (one nucleotide) from *Pooc-10* (see below).

In our earlier study in Gila topminnow (Hedrick & Parker 1998a), we described 12 different MHC sequences, indicated as *Pooc-1* through *Pooc-12* in Tables 1 and 4. In this expanded study, we found five more MHC sequences, three of which were just mentioned (these five new complete sequences, and 12 more found in Yaqui topminnow, have been deposited in Genbank, accessions xxx – xxx). However, for this population analysis we assume that sequences *Pooc-9*, -12, and -13 are from a second duplicated locus (see discussion below and in Appendix) and have calculated frequencies in Table 1 based only on the other alleles.

#### *Yaqui Topminnow*

Table 3 gives allele frequencies and average heterozygosity for the microsatellite loci and the MHC locus in Yaqui topminnow. Note there are three additional loci, *4-44*, *Acc*, and *G53*, variable in Yaqui topminnow but invariant in Gila topminnow. Locus *6-10* did not amplify reliably in the Yaqui topminnow and therefore was not used. Only one comparison of observed to expected heterozygosities was statistically significant (a deficiency of observed MHC heterozygosity in North Pond) but it became non-significant after a standard Bonferroni correction. Twelve new MHC alleles were in the two Yaqui topminnow samples, none of which was in any Gila topminnow populations. The two Yaqui samples were quite similar, sharing alleles or very similar alleles at all loci. Genetic variation in the two samples was also substantial and similar.

[insert Table 3]

There was a substantial divergence in these markers between Yaqui and Gila topminnows. For six loci, there was no overlap in alleles. In fact, for three microsatellite loci, the sizes of the alleles in the two species was greatly divergent. The largest difference was for locus *OO56*, which in the Gila topminnow alleles varied from 143 to 153 base pairs, but fixed in Yaqui topminnow for allele 256, more than 100 base pairs larger. However, this locus is a complicated repeat (Parker *et al.* 1998) so it is not clear how many mutational steps are responsible for the difference in size between species. However, the other two loci that have disjunct size distributions, *C-15* and *LL53*, are

simple GT dinucleotide repeats. Because the most common mutation at perfect repeat microsatellite loci is thought to differ by only a single repeat (for a direct examination in a fish, see Jones *et al.* 1999), it appears that isolation between the taxa may been long and complete for so many differences to accumulate.

#### *Further analysis of MHC variation*

In this study of MHC variation, expanded from Hedrick & Parker (1998a), it appears we are amplifying sequences from two MHC loci in some individual Gila topminnows from some populations. For example, we have found from SSCP that 67 individuals have banding patterns indicating three sequences, and in two individuals, four sequences. This has led us to reevaluate our earlier conclusions and suggest that sequences *Pooc-9* and *Pooc-12*, and probably *Pooc-13*, are at a second locus (see Appendix for rationale)

Overall, *Pooc-9* is common in all natural populations except Cottonwood Spring, the two Bylas Springs populations, and Cienega Creek, i.e., all populations with *Pooc-6* except Cottonwood Spring (Table 1). Interestingly, *Pooc-9* was not found in the samples from Watson Wash, suggesting that the source of this stock did not have (or has lost) *Pooc-9*. *Pooc-12* was found in low frequency only in two populations, Red Rock Falls and Santa Cruz River, both of which had *Pooc-11*. *Pooc-13* was only found in Red Rock Falls.

The number of amino acid (nucleotide) differences ranged from 1 to 15 (1 to 30) for the Gila topminnow MHC sequences and ranged from 1 to 18 (1 to 32) for the Yaqui topminnow MHC sequences (Table 4). The phylogenetic tree for all the sequences (Fig. 3) shows a pattern of intermixture for the Gila and Yaqui (indicated by arrows) sequences. This pattern is typical for MHC sequences of related taxa and is thought to be the result of transspecies polymorphism (Klein 1987; Klein *et al.* 1993), in which polymorphisms are older than the species because of balancing selection. Notice, however, that the 12 sequences of Yaqui topminnow are on three major branches that do not include any Gila topminnow sequences (all of which show some bootstrap support). In other words, even though the tree has sequences from the two taxa intermixed, the nearest sequence for each sequence is another sequence in the same taxon unlike that often found for other MHC studies (e.g. Hedrick *et al.* 2000).

[insert Table 4]

[insert Figure 3]

Of the 62 amino acids, 26 (41.9%) were variable and of the 190 nucleotides sequenced, 55 (28.9%) were variable. At two amino acid positions, 73 and 74, there were five different amino acids present over the 29 sequences while there were four different amino acids observed at position 40. There were only four codon positions, positions 39, 42, 59, and 71 that had only synonymous variation. Interestingly, the silent differences at sites 39 and 41 were only found in *Pooc-9* and *Pooc-12*, sequences at the hypothesized second locus, and the silent difference at position 59 was only found in these two sequences and *Pooc-13*.

We tried to determine the amino acid positions thought to be important in the antigen-binding site (ABS) from examination of the structure of an analogous class II molecule in humans (Brown *et al.* 1993), an Atlantic salmon sequence, a guppy sequence, and our topminnow sequences. We are confident of our alignment of these sequences and the homologous putative ABS positions are underlined in Table 4. Note that these differ by three from that we indicated in Hedrick and Parker, 1998), i.e., position  $x$  here was identified as  $x + 3$  there. Of the 20 ABS positions, 11 (55.0%) are variable for different amino acids over the 29 alleles, while for the remaining 42 positions that are not thought to interact with the ABS, 15 (35.7%) are polymorphic for different amino acids.

In addition, the estimated rate of nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions, for antigen-binding and non-antigen-binding amino-acid positions are in Table 5. For all categories of sites in both taxa,  $d_N$  (0.238) is greater than  $d_S$  (0.100) and the overall ratio of  $d_N/d_S$  of 2.38 is statistically significant. Generally for MHC genes, the ratio for the ABS is higher than for the non-ABS. This is true for Gila topminnow but not for Yaqui topminnow. In previous studies in a number of mammals, a  $d_N/d_S$  ratio higher than unity for ABS positions and a ratio near unity for non-ABS positions has been found. In other words, the  $d_N/d_S$  value higher than unity for non-ABS positions in Yaqui topminnows is quite surprising. It is possible that this observation indicates that the exact identity of the ABS sites, and non-ABS sites by extension, in topminnows (and maybe other fish) may differ from that in humans.

[insert Table 5]

## Discussion

Molecular genetic variants are being widely used to identify species, ESUs, and MUs in a number of different endangered taxa. From examination of molecular variants here we show that the Gila and Yaqui topminnows are quite divergent from each other and that status as separate species (not subspecies) is appropriate. In addition, the Gila topminnow populations form five distinct groups. Below we recommend that these groups be separated into two ESUs and that one ESU be composed of four MUs.

These recommendations are based on the present data from microsatellite loci and the MHC locus, as well as other geographic, ecological, and life-history information (see below; Parker *et al.* 1999; Minckley 1999). Neutral genetic markers are assumed to be reflective of the historical relationships between groups and other ecological or life-history information reflective of the potential for adaptive differences between groups. The MHC gene, because of its known contribution to pathogen resistance would be a prime candidate for a genetically adaptive gene. In fact, the higher rate of non-synonymous to synonymous substitutions observed here suggests a selective advantage for variants that have amino-acid differences. However, the time scale of the selection resulting in the higher rate of non-synonymous to synonymous substitutions may be many thousands of generations and may not reflect demographic or stochastic events of recent decades, or even centuries.

In the Gila topminnow, it appears that non-selective factors such as differences in population size and the natural extinction-recolonization cycles that have occurred as the result of drought-flood cycles are the primary influence on the distribution of genetic variation within populations (Parker *et al.* 1999). For example, low variation in the Bylas populations for both the microsatellite loci and the MHC gene is most probably the result of small habitat size (the site can support at most a few hundred adults) and the recurring bottlenecks observed since the site was discovered (Marsh & Minckley 1990; Minckley *et al.* 1991). In other words, even though there is evidence of longterm selective effects on the MHC sequence variation, the immediate factors influencing the amount of MHC genetic variation in Gila topminnows are non-selective. The high correlation of neutral

and potentially adaptive variation observed here actually may be more likely when non-selective factors dominant the distribution of genetic variation within and between populations.

### *Gila Topminnow*

Parker *et al.* (1999) suggested, based on physical habitat, associated biota, life-history, and molecular genetics, that the Monkey Spring population be managed as a separate ESU. In particular, Monkey Spring is a constant, warm spring that has been isolated by a 10-m high travertine dam for perhaps 10,000 years from other populations and was historically occupied by a now extinct species of pupfish and a distinct form of Gila chub (Minckley 1999). In addition, it was highly differentiated genetically and had 50% longer male development than other samples in a common laboratory environment (Cardwell *et al.* 1998). We find nearby Cottonwood Spring is similar genetically and suggest it as part of a Monkey Spring ESU. Monkey Spring fish may migrate downstream to Sonoita Creek and then upstream to the smaller Cottonwood Spring, a total distance of only five stream km, hence their similarity genetically. Monkey Spring is the unusual and isolated primary habitat for this ESU, while Cottonwood Spring is a secondary site. Monkey Spring ESU is quite differentiated from the other geographically separate groups (average  $D_A$  value with Bylas, Cienega, and Sharp is 0.656). The Sonoita Creek group, which is directly downstream from Monkey and Cottonwood Springs, and the Monkey Spring ESU have an average genetic distance of 0.422.

The other ESU is composed of four different MUs. First, the Bylas Springs populations are the only samples from the mainstem Gila River and are quite differentiated from the other populations except for Cienega. The Bylas Spring 1 sample is so similar (and so invariant) to the Bylas Spring 2 sample, that we have combined our captive stocks of these two populations. The Bylas populations and the Cienega Creek sample have a mean  $D_A$  values of 0.287 but a substantial proportion of this genetic distance appears to have been generated by loss of genetic variation in the Bylas populations (Hedrick 1999). However, the two sites are separated by approximately 580 kilometers of stream channel, much of which is dry now except during flooding, so that suggesting different MUs for the Bylas and Cienega populations is reasonable. In

addition, the sample from Sharp Spring is highly divergent on a molecular basis and so physically isolated to be classified as a third MU. The other three natural populations, Red Rock (both sites), Coalmine, and Sonoita, all in the Sonoita Creek drainage, form another, highly variable group, considered a fourth MU.

Two caveats are important to mention. First, bottlenecks and other factors that result in low population size may result in a large increase in genetic distance in a short time period (Hedrick 1999). Although we do not have an appropriate comparison, the extent of genetic variation in these populations for the loci examined here does not appear to be large as compared to similar loci in other species. Further, the extent of variation for allozymes (Vrijenhoek *et al.* 1985) and mtDNA (Quattro *et al.* 1996) is low in these populations. This suggests that non-selective factors, perhaps in recent decades or centuries, have reduced genetic variation and may have resulted in an increase in genetic distance between populations. Second, a study by Sheffer *et al.* (1999) demonstrated that crosses between all pairs of Bylas Spring 2, Cienega Creek, Monkey Spring, and Sharp Spring were successful and produced both  $F_1$  and  $F_2$  progeny. Further, here was no evidence that the  $F_1$  and  $F_2$  progeny had reduced fitness as indicated by growth rate, fecundity, survival, and bilateral asymmetry. Therefore, it appears that although there is substantial differentiation between these populations for molecular markers, it does not appear to have influenced mating ability between the groups or fitness in subsequent  $F_1$  and  $F_2$  progeny.

The Boyce Thompson stock is very similar to Monkey Spring. We have recently been informed (J. Brooks, pers. comm.) that the Boyce Thompson pond was drained in the early 1980s to remove non-native western mosquito fish (*Gambusia affinis*). It was then restocked with fish from Dexter National Fish Hatchery, at the time a stock from Monkey Spring. Also from our analysis, we conclude the Watson Wash population was similar to the Monkey Spring-Boyce Thompson stock. We also were recently informed (J. Brooks, pers. comm.) that Watson Wash was stocked from Boyce Thompson in the late 1980s by personnel from the US Bureau of Land Management. Our findings, based on molecular genetics data, are thus consistent with these anecdotal accounts.

### *Yaqui Topminnow*

Yaqui topminnow differs substantially from Gila topminnow for both microsatellite and MHC markers we examined. Only 2 of 25 microsatellite alleles in the Yaqui topminnow are found in all of our Gila topminnow samples and a number of Yaqui topminnow microsatellite alleles differ dramatically in size from Gila topminnow alleles. All MHC sequences differ between the two taxa, unlike the sharing of MHC sequences sometimes observed between species. These findings are consistent with the observation that Gila and Yaqui topminnow (from the San Bernadino National Wildlife Refuge) differ by seven restriction sites at mitochondrial DNA (Quattro *et al.* 1996).

Given this evidence, and a past history of identification of the two taxa as separate species (see synopsis of their status over time in Quattro *et al.* 1996), it appears that subspecific status for Gila and Yaqui topminnows needs reevaluation. Minckley (1969, 1971) considered that the two taxa subspecies because of subtle morphological differences, but the recent molecular findings make species status for Gila and Yaqui topminnows appropriate (see also Minckley 1999).

In an earlier study, Quattro & Vrijenhoek (1989) compared several fitness correlates in the laboratory for samples that included both putative subspecies of Sonoran topminnows. They used Gila topminnows from Monkey Spring (as an example of low allozyme heterozygosity), Yaqui topminnows from Tule Spring (as an example of intermediate allozyme heterozygosity), and Gila topminnows from Sharp Spring (as an example of high allozyme heterozygosity). From our results here and those of Quattro *et al.* (1996), it appears that use of Yaqui topminnows from Tule Spring was inappropriate as it appears to be a different species. Further, using the seven microsatellite loci in common to the two taxa, Monkey Spring has  $H_E = 0.138$  and  $n = 2.0$ , Tule Spring has  $H_E = 0.443$  and  $n = 2.9$ , and Sharp Spring has  $H_E = 0.202$  and  $n = 3.0$ . Therefore, the ranking of genetic variation based on allozyme loci is inconsistent with that found for more variable molecular markers.

### *MHC Considerations*

We used both microsatellite and MHC data to distinguish different groups of Gila topminnows. The use of MHC data was justified by the very high correlation between the

extent of variation for microsatellite loci and the MHC gene within populations (see discussion above). However, there are at least two sources of evidence that selection has played an important role in the pattern of MHC variation observed in the Sonoran topminnows. First, the transpecies polymorphism observed in Gila and Yaqui topminnows is unlikely under neutrality (Klein *et al.* 1993). The generally accepted explanation for this pattern is that balancing selection maintains lineages, even over speciation events. Second, the higher observed ratio of non-synonymous to synonymous substitutions is generally explained by positive Darwinian selection, a general term which includes balancing selection. However, we found a high  $dN/dS$  ratio for putative non-ABS positions in the Yaqui topminnows, a result contrary to that in mammals. This may suggest that the ABS positions, identified by X-ray crystallography for a human class II molecule, may not be identical in topminnows or other fishes. A new MHC finding from the present investigation is the polymorphic for a MHC duplication in the Gila topminnow populations. Further investigations of this duplication, including crosses of individuals with and without the duplication and crosses of different haplotypes with the duplicated gene and examination of gamete sequences, are planned.

### **Acknowledgements**

We thank ... for their comments on the manuscript, W. L. Minckley for constructing Figure 1, and Dan Garrigan for constructing Figure 3. We appreciate the efforts of David Weedman, Kevin Cobble, and Ruby Sheffer in collecting the samples. This research was supported by the NSF, Heritage Fund of Arizona, Bureau of Reclamation, and Ullman Distinguished Professorship.

## References

- Bodmer JG, Marsh SG, Albert ED, Bodmer WF, Dupont B, Erlich HA, Mach B, Mayr WR, Parham P, Sasuzuki T. 1992. Nomenclature for factors of the HLA system, 1991. WHO Nomenclature Committee for factors of the HLA system. *Tissue Antigens* **39**, 161-173.
- Brown JH, Jardetzky TS, Gorg, JC, Stern LJ, Uban RG, et al. (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33-39.
- Cardwell TN, Sheffer RJ, Hedrick PW (1998) Differences in male development time among populations of the endangered Gila topminnow. *Journal of Heredity* **89**, 353-355.
- Edwards, S., Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology and Evolution* **13**, 305-311.
- Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **82**, 313-318.
- Hedrick PW, Kim KJ (2000) Genetics of complex polymorphisms: Parasites and maintenance of the major histocompatibility complex variation. In: Singh, R. S., and Krimbas, C. B. (eds.) *Evolutionary Genetics: From Molecules to Morphology*, pp. 204-234. Cambridge Univ. Press.
- Hedrick PW, Parker KM, Gutierrez-Espeleta GA, Rattink A, Lievers K (2000) Major histocompatibility complex (MHC) variation in the Arabian oryx. *Evolution* (in press).
- Hedrick PW, Parker KM (1998) MHC variation in the endangered Gila topminnow. *Evolution* **52**, 194-199.
- Jones AG, Rosenqvist G, Berglund A, Avise JC (1999) Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics* **152**, 1057-1063.
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro, H. N. (ed) *Mammalian Protein Metabolism*, pp. 21-132. Academic Press, New York.
- Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunology* **19**, 155-162.

- Klein J, Satta Y, Takahata N, O'Huigin C (1993). Transpecific MHC polymorphism and the origin of species in primates. *Journal of Medical Primatology* **22**, 57-64.
- Kumar S, Tamura K, Nei M (1993) MEGA: Molecular Evolutionary Genetics Analysis, Version 1.01. Pennsylvania State University, University Park, PA.
- Minckley WL (1969) Native Arizona fishes, part I. Livebearers. *Wildlife Views* **16**, 6-8.
- Minckley WL (1973) Keys to native and introduced fishes of Arizona. *Journal of the Arizona Academy of Science* **6**, 183-188.
- Minckley WL (1999) Ecological review and management recommendation for recovery of the endangered Gila topminnow. *Great Basin Naturalist* **59**, 230-244.
- Moritz C (1994) Defining "evolutionarily significant units" for conservation. *Trends in Ecology and Evolution* **9**, 373-375.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418-426.
- Parker KM, Hughes K, Kim TJ, Hedrick PW (1998) Isolation and characterization of microsatellite loci from the Gila topminnow (*Poeciliopsis o. occidentalis*) and their utility in guppies (*Poecilia reticulata*). *Molecular Ecology* **7**, 357-363.
- Parker KM, Sheffer RJ, Hedrick PW (1999) Molecular variation and evolutionarily significant units in the endangered Gila topminnow. *Conservation Biology* **13**, 108-116.
- Ober C, Hyslop T, Elias W, Weitkamp LR, Hauch WW (1998) HLA matching and fetal loss: results of a 10-year prospective study. *Human Reproduction* **13**, 33-38.
- Ota T (1993) *DISPAN: Genetic Distance and Phylogenetic Analysis*. Pennsylvania State University, University Park, PA.
- Penn DJ, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *American Naturalist* **153**, 145-164.
- Quattro JM, Leberg PL, Douglas ME, Vrijenhoek RC (1996) Molecular evidence for a unique evolutionary lineage of endangered Sonoran desert fish (genus *Poeciliopsis*). *Conservation Biology* **10**, 128-135.

- Quattro JM, Vrijenhoek RC (1989) Fitness differences among remnant populations of the endangered Sonoran topminnow. *Science* **245**, 976-978.
- Sheffer RJ, Hedrick PW, Minckley WL, Velasco AL (1997) Fitness in the endangered Gila topminnow. *Conservation Biology* **11**, 162-171.
- Sheffer RJ, Hedrick PW, Velasco AL (1999). Testing for inbreeding and outbreeding depression in the endangered Gila topminnow. *Animal Conservation* **2**, 121-129.
- Vrijenhoek RC, Douglas ME, Meffe GK. 1985. Conservation genetic of endangered fish populations in Arizona. *Science* **229**, 400-402.
- Waples RS (1995) Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *American Fisheries Society Symposium* **17**. 8-27.

## Appendix

The conclusion that sequences *Pooc-9*, *Pooc-12*, and *Pooc-13* are from a second MHC locus is based on the following observations. First, we have not observed any individuals that have only *Pooc-9* or only *Pooc-12*. Because *Pooc-9* is very common in several populations the likelihood of not observing homozygotes for *Pooc-9* if it were allelic to our other sequences is very small. For example, in the sample of 20 for Red Rock at Cott Tank all individuals have both *Pooc-6* and *Pooc-9*. If *Pooc-9* were allelic to *Pooc-6* so that they would both have a frequency of 0.5 in this sample, then the probability of not seeing any *Pooc-9* homozygotes would be  $(0.75)^{20} = 0.003$ .

Second, in all individuals with a third sequence, *Pooc-9* or *Pooc-12* is present. (For one individual with four sequences, both *Pooc-9* and *Pooc-12* are present. The other individual with four sequences is discussed below.) Further, *Pooc-9* always appears to be associated with allele *Pooc-6* and *Pooc-12* always associated with allele *Pooc-11*. This suggests that *Pooc-9* arose as a duplication on a chromosome carrying *Pooc-6* or that *Pooc-12* arose as a duplication on a chromosome carrying *Pooc-11*. At this point, presence of the duplication is polymorphic in the Gila topminnow, a common observation for class II genes in other species, including humans which may have from two to five *DRB* genes on a chromosome (Bodmer *et al.* 1992). Because *Pooc-6* and *Pooc-11* only differ by two amino acids (five nucleotides), there actually may have been only one origin of the second gene and the difference in *Pooc-6* and *Pooc-11* as well as *Pooc-9* and *Pooc-12* (three amino acids and six nucleotides different) evolved since then.

Third, further support comes from sequence data that places *Pooc-9* and *Pooc-12* together and very different from the other sequences (Table 4 and Fig. 3). These two sequences are identical through position 73 and share an identical sequence at positions 39 to 41 not found in other sequences. An individual from Red Rock Falls has four sequences, *Pooc-11*, -12, -13, and -14. Because *Pooc-13* and *Pooc-14* were not in other individuals, it appears that one of these should be a sequence from the hypothesized duplication. Interestingly, sequence *Pooc-13* shares sequence at positions 56, 59, and 60 with *Pooc-9* and *Pooc-12* that is not found in any other sequences. In addition, *Pooc-14* is very similar to *Pooc-6* and -11. In other words, a reasonable hypothesis is that *Pooc-9*,

-12, and -13 are all from the second locus and that *Pooc*-14 is allelic to the primary MHC locus.

## Figure legends

**Fig. 1.** Map showing the locations of the Bylas and Yaqui samples (above) and the other natural samples (below).

**Fig. 2.** Unrooted neighbor-joining tree for the 13 samples of Gila topminnows based on allelic frequencies at the five polymorphic microsatellite loci and the MHC locus sequences where the numbers are bootstrap values.

**Fig. 3.** Neighbor-joining tree for the MHC sequences where the numbers are bootstrap values and the sequences from Yaqui topminnows are indicated with a triangle and the Gila topminnow sequences are not.





MHC	<i>n</i>	1.2	1.2	2.0	2.8	2.2	2.2	1.8	1.2	2.2	1.8	2.0	4.4	3.8
	1	1.000	1.000	0.500	---	---	---	---	---	---	---	---	---	---
	2	---	---	---	---	---	---	---	---	---	---	---	0.350	---
	3	---	---	---	---	---	---	---	---	---	---	---	0.200	---
	4	---	---	---	0.025	---	---	0.833	---	---	---	---	---	---
	5	---	---	0.500	---	---	---	---	---	---	---	---	---	---
	6	---	---	---	0.938	1.000	0.975	0.167	1.000	0.575	0.600	0.630	0.425	0.088
	7	---	---	---	0.038	---	---	---	---	---	---	---	---	---
	8	---	---	---	---	---	---	---	---	0.200	0.400	0.111	0.175	0.362
	9	---	---	---	(0.425)	(0.750)	---	---	(1.000)	(0.850)	(0.900)	(0.926)	(0.750)	(0.400)
	10	---	---	---	---	---	---	---	---	---	---	0.019	---	---
	11	---	---	---	---	---	---	---	---	0.200	---	0.241	0.325	---
	12	---	---	---	---	---	---	---	---	(0.150)	---	---	(0.050)	---
	13	---	---	---	---	---	---	---	---	(0.050)	---	---	---	---
	14	---	---	---	---	---	---	---	---	0.025	---	---	---	---
	15	---	---	---	---	---	---	---	---	---	---	---	0.050	---
	16	---	---	---	---	---	0.025	---	---	---	---	---	---	---
	17	---	---	---	---	---	---	---	---	---	---	---	0.025	---
$H_O$		---	---	0.600	0.125	0.050	0.050	0.333	---	0.600	0.600	0.667	0.800	0.400*
$H_E$		---	---	0.513	0.119	0.050	0.050	0.197	---	0.604	0.492	0.542	0.697	0.708
<i>n</i>		1	1	2	3	2	2	2	1	3	2	4	5	4

**Table 2** The genetic distance ( $D_A$ ) between pairs of ESUs and MUs. The values on the diagonal are mean  $D_A$  values within group where there are more than one population.

ESU or MU	ESU or MU				
	Bylas	Cienega	Monkey	Sonoita	Sharp
Bylas	0.093				
Cienega	0.287	---			
Monkey	0.707	0.632	0.300		
Sonoita	0.567	0.431	0.422	0.112	
Sharp	0.518	0.458	0.630	0.420	---

**Table 3** The allele frequencies for the microsatellite loci in two Yaqui topminnow samples (samples sizes in parentheses), the average frequency in the nine Gila topminnow samples having the one shared polymorphic allele, G49-161, the size or size range for the other alleles in the Gila topminnow, and the observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity and observed number of alleles ( $n$ ), and the mean over all loci. \* indicates a significant deficiency of  $H_O$  before (but not after) Bonferroni correction

Locus	Allele	Yaqui		
		Tule Spring (20)	North Pond (19)	Gila
<i>G-49</i>	149-159	---	---	0.917
	161	0.550	0.421	0.083
	163	0.450	0.579	---
	$H_O$	0.400	0.526	0.000-0.350
	$H_E$	0.508	0.501	0.000-0.380
	$n$	2	3	1-2
<i>C-15</i>	164	0.475	---	---
	174	0.100	---	---
	180	0.200	0.026	---
	190	0.150	0.526	---
	192	0.075	0.447	---
	202-248	---	---	1.000
	$H_E$	0.650	0.474	0.000-0.850
	$H_O$	0.714	0.537	0.000-0.879
$n$	5	3	1-12	
<i>OO56</i>	143-153	---	---	1.000
	256	1.000	1.000	---
	$H_E$	0.000	0.000	0.000-0.650
	$H_O$	0.000	0.000	0.000-0.581
	$n$	1	1	1-3
<i>LL53</i>	110	--	0.053	---
	112	0.500	0.237	---
	114	0.425	0.105	---
	116	0.075	0.605	---
	136-164	---	---	1.000
	$H_E$	0.600	0.789	0.000-0.850
	$H_O$	0.579	0.579	0.000-0.669
$n$	3	4	1-5	

<i>4-44</i>	106	0.275	0.778	---	
	108	0.050	0.111	1.000	
	114	0.425	0.111	---	
	118	0.250	---	---	
	$H_E$	0.600	0.789	---	
	$H_O$	0.697	0.579	---	
	$n$	4	3	1	
<i>Acc</i>	124	0.825	0.763	---	
	128	---	---	1.000	
	130	0.175	0.237	---	
	$H_E$	0.250	0.474	---	
	$H_O$	0.297	0.371	---	
	$n$	2	2	1	
	<i>G53</i>	96	0.825	0.806	---
100		---	---	1.000	
102		0.050	---	---	
104		0.125	0.194	---	
$H_E$		0.250	0.333	---	
$H_O$		0.309	0.322	---	
$n$		3	2	1	
<i>MHC</i>	1-17	---	---	1.000	
	18	0.125	0.250	---	
	19	0.250	0.056	---	
	20	0.400	---	---	
	21	---	0.028	---	
	22	0.125	0.028	---	
	23	0.062	---	---	
	24	0.025	---	---	
	25	---	0.333	---	
	26	---	0.056	---	
	27	---	0.222	---	
	28	---	0.028	---	
	29	0.025	---	---	
		$H_O$	0.750	0.500*	0.000-0.800
		$H_E$	0.762	0.794	0.000-0.708
	$n$	6	8	5	
<i>Mean</i>	$H_O$	0.443	0.370	0.026-0.296	
	$H_E$	0.393	0.425	0.009-0.325	
	$n$	2.86	2.43	0.90-2.75	



**Table 5** The estimated rates of nonsynonymous ( $d_N$ ) and synonymous substitutions ( $d_S$ ) for antigen (ABS) and non-antigen binding amino-acid sites and their ratio for sequences found in the Gila topminnow, Yaqui topminnow, and both taxa.  $N$  is the number of codons in each category and  $P$  is the probability that  $d_N$  and  $d_S$  are different

Taxa	Positions	$N$	$d_N$	$d_S$	$d_N/d_S$	$P$
Gila	ABS	20	$0.209 \pm 0.050$	$0.064 \pm 0.030$	2.38	0.02
	Non-ABS	42	$0.045 \pm 0.018$	$0.038 \pm 0.019$	1.18	0.40
	All	62	$0.100 \pm 0.022$	$0.046 \pm 0.016$	2.17	0.02
Yaqui	ABS	20	$0.216 \pm 0.054$	$0.072 \pm 0.041$	3.00	0.02
	Non-ABS	42	$0.079 \pm 0.024$	$0.019 \pm 0.013$	4.16	0.01
	All	62	$0.124 \pm 0.026$	$0.035 \pm 0.015$	3.54	0.00
Both	ABS	20	$0.238 \pm 0.053$	$0.100 \pm 0.039$	2.38	0.02
	Non-ABS	42	$0.073 \pm 0.022$	$0.035 \pm 0.015$	2.09	0.08
	All	62	$0.128 \pm 0.025$	$0.055 \pm 0.016$	2.33	0.01