Reproductive isolation among endemic pupfishes (Cyprinodon) on San Salvador Island, Bahamas: microsatellite evidence

BRUCE J. TURNER1*, DAVID D. DUVERNELL2, THOMAS. M. BUNT1† and MICHAEL G. BARTON3

1Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
2Department of Biological Sciences, Southern Illinois University, Edwardsville, IL 62026, USA
3Division of Science and Mathematics, Centre College, Danville, KT 40422, USA

Received 11 October 2007; accepted for publication 22 January 2008

Three pupfish (Cyprinodon) morphotypes (two endemic) occur in some of the young (6000 ybp) saline lakes on the Bahamian island of San Salvador. The ‘normal’ morph, a detritivore/omnivore, is not different in its general features from Cyprinodon variegatus from other Bahamian islands. ‘Bulldog’ is a scale-eater/piscivore that preys upon normal pupfish, and ‘bozo’ is a specialized molluskivore. Reproductive isolation among these morphs is not predicted by the evolutionary biology of congeneric species because sympatry of even morphologically and ecologically quite divergent pupfishes has usually resulted in hybridization/introgression. Survey of variation at eight microsatellite loci reveals that sympatric normal and bulldog populations are genetically distinctive by several criteria, and are therefore likely reproductively isolated. The bulldog morph in Crescent Pond is markedly divergent from those in Little Lake and Osprey Lake, a finding consistent with, although it does not prove, separate parallel origins of this morphotype. The data also suggest that the bulldogs in the latter two lakes did not evolve by intralacustrine speciation from the current sympatric normal populations. Some of the genetic data suggest that the bozo morph may also be reproductively isolated from the other two pupfishes, but only a small, pooled sample of this rare morphotype was available, and the issue is not resolved. Isolating mechanisms between bulldog and normal morphs are of special interest because of the possibility that they arose as a consequence of a predator–prey relationship. A strong correlation between reproductive isolation and predator–prey interactions could provide an important example of ecological speciation via direct selection against heterotypic interactions. © 2008 The Linnean Society of London, Biological Journal of the Linnean Society, 2008, 95, 566–582.


INTRODUCTION

‘Ecological speciation’ (Schluter, 1998, 2001) is a prominent paradigm among contemporary ideas about the relationship between natural selection and the advent of reproductive isolation, but the factors that critically influence this relationship are not well understood. Although much attention has been focused on the ‘byproduct model,’ in which reproductive isolation arises as correlated response to selection favouring ecological divergence (Schluter, 2002), isolation could also result from selection against hybrids (classical ‘reinforcement’). In addition, it has been suggested that direct selection against individuals that attempt or participate in heterotypic matings might be particularly effective at generating premating isolation (Servedio, 2001).

An evolving predator–prey interaction might be especially likely to lead to rapid reproductive isolation by direct selection (Schluter, 2002). Premating isolation could be a virtually immediate result of selection for anti-predatory adaptations. For example,
individuals of an ancestral stock that respond positively to courtship by conspecifics who then prey upon them will be selected against, favouring those who avoid such interactions. Fishes provide many examples in which predator and prey are closely related, particularly among the trophically diverse radiations of cichlids in the African Rift Lakes. However, a potential relationship between predator–prey interaction and reproductive isolation has apparently been the subject of only a single published study (Albert & Schluter, 2004).

Three trophically divergent pupfish (Cyprinodon) morphotypes (two endemic) from saline lakes on the island of San Salvador (Bahamas) provide a system in which this relationship can be further explored. One of the morphotypes is a piscivore and scale-eater with morphological features that are rare, and associated behaviours that are very likely novel, among its entire group. Its prey may also be its proximate ancestor. The other morphotype is apparently specialized to exploit a locally abundant food resource and apparently does not interact trophically with the presumptive ancestral form. The geology and hydrological history of the island suggest that the endemic morphotypes have evolved in a span of approximately 6000 years or less.

The San Salvador pupfishes are little known (Holtmeier, 2000, 2001) and undescribed. The three morphotypes (Fig. 1) are normal, bulldog, and bozo (Holtmeier, 2000, 2001). Like most pupfishes, normal is a substrate-browsing omnivore, consuming detritus, algae and other plant material, benthic macroplankton, and fish eggs (Holtmeier, 2000). It is by far the most abundant morph on the island, comprising at least 90% of the pupfish taken in larger lakes and it is the only morph found thus far in ‘blue holes’. In its general appearance, colour patterns and behaviour in the field, this morphotype closely resembles the pupfish that are ubiquitous and usually abundant in coastal mangrove forests, interior saline lakes, and blue holes throughout the Bahamas. These populations have historically been identified as Cyprinodon variegatus (Miller, 1962; Bohlke & Chaplin, 1964), a widespread coastal Atlantic and Caribbean species (or species complex). With the single exception of a New Providence endemic, none of the Bahamian populations are currently recognized as distinct taxa. A recent molecular study of the phylogeography of C. variegatus as a whole (Haney et al., 2007) showed that specimens from Andros Island and San Salvador (normal morphotype only) comprise phylogenetically distinctive matrilines.

Figure 1. Three pupfish morphotypes from Little Lake, San Salvador. Upper, normal; middle, bulldog; lower, bozo.
The bulldog morph has long, recurved mandibular teeth that give the jaws a characteristic wedge shape, with a distinct dorsal gape, and markedly enhance the apparent length of the lower jaw. It consumes scales and other fish parts (e.g. fin clips) in addition to normal’s diet (Holtmeier, 2000). Individuals of this morphotype comprise 10% or fewer of the specimens in our collections and apparently occur in only a few lakes. There is direct evidence that bulldog preys upon other pupfish because identifiable pupfish remains have been observed in the guts of field-collected individuals (Holtmeier, 2000). Circumstantial evidence (see below) argues that pupfishes are its major prey. This implies that its specialized trophic features have been selected in the presence of the normal morphotype and not in allopatry.

The bozo morph has an enlarged, fleshy snout of unknown function (possibly sensory?) It is apparently a molluskivore that preys upon the snails that are abundant in some of the lake habitats. The bozo morph is consistently the rarest in our surveys, usually at less than 1%.

In the present study, we investigate two basic questions about the biology and evolutionary history of the San Salvador pupfishes. First, are the morphotypes reproductively isolated? Second, if reproductive isolation has evolved, does it result from intralacustrine (essentially sympatric) speciation? We place special emphasis on the first question because the evolutionary biology of pupfishes in general predicts that the existence of reproductively isolated, sympatric forms is quite unlikely.

MATERIAL AND METHODS

STUDY AREA AND ITS GEOLOGICAL HISTORY, HYDROLOGY, AND FISH FAUNA

San Salvador is small (approximately 175 km$^2$ in area; 19 km long and 11 km at its widest point), largely flat (highest point is approximately 37.5 m a.s.l.) and is essentially an isolated carbonate ('karst') pillar that has never been connected with other islands. Saline lakes dominate its interior surface features. The existence of these lakes depends on hydrostatic support from the surrounding sea: the karst is permeable and, if the sea level drops, the lakes disappear. The geological record contains evidence of several marked fluctuations in sea level through the Pleistocene and Holocene periods in response to Glacial events, with highstands during interglacial maxima (Hagey & Mylroie, 1995). Some highstands established systems of surface lakes and/or lagoons that have subsequently disappeared. ‘Blue holes’, as defined by Mylroie, Carew & Moore (1995: 231), are also present on the island.

The current lakes are shallow (maximum depth of approximately 12 m) and many fluctuate seasonally in level. Their bottoms are irregular and often covered with fine grain sand and/or mud. Some contain blue holes at their bottoms. Many of the lakes are connected to one another, and/or the surrounding sea, by fissures and subterranean channels. Most of the lakes are separated by shallow berms that are easily breached by storms (some are composed only of shells) and are likely not effective barriers to movement of biota. However, migration has not established biotic uniformity among the lakes despite their propinquity and underground connections, and several support faunas comprised of unique and characteristic combinations of invasive species and, sometimes, distinct microbial communities (Godfrey et al., 1994). Lakes that connect to the sea generally have salinities of approximately 50 ppt, but those without a marine connection are hypersaline. The flora of all lakes is predominantly benthic algae (Batophora and other species).

The surrounding sea reached its current level approximately 3000 years BP, and some of the lakes probably began filling approximately 3000 years earlier. Multiple lines of evidence, including C-14 analysis (Kwolke, 1985), indicate that the lakes date from between 3500 and 6000 years BP (Pacheco & Foradas, 1986; Godfrey et al., 1994; Hagey & Mylroie, 1995; Teeter, 1995; Carew & Mylroie, 1997). For example, the first red mangrove (Rhizophora sp.) in the subrecent record, indicating connections to the sea by which fish could have entered, is evident only approximately 4000 years BP. (Pacheco & Foradas, 1986; Godfrey et al., 1994). Thus, 6000 years BP places a ceiling on the age of the current fish fauna of the lakes and, presumably, on the span in which any endemics may have evolved. For any components of the fauna to be older than this, they would have had to survive the last sea level drop in refugia that would be approximately 125 m below the present surface of the island. No blue holes are known on the island of more than approximately 25 m in depth. The columnar shape of the island implies that drops in sea level likely did not establish refugia in the form of peripheral lake basins that are now submerged.

The fish fauna of the interior waters is comprised of the pupfishes and only three other species. One, the poeciliid Gambusia hubbsi, occurs in large numbers in almost every surface body of water on the island. Downhower, Brown & Matsui (2000), Abney & Rakociński (2004) and Langerhans, Gifford & Joseph (2007) discuss its ecology on other islands. Detailed studies are lacking, but its general biology suggests that G. hubbsi is not likely to be a major bulldog prey item. It is a surface-dwelling, streamlined, manoeu-
valorous fish that often forms dense shoals, and is more likely to prey upon pupfishes than the reverse (i.e. predation by introduced *Gambusia affinis* has had a negative impact on threatened populations of endemic pupfishes in New Mexico; Rogowski & Stockwell, 2006). Pupfishes are far less manoeuvrable, typically remain close to the substrate, and are not adapted for feeding in open water or for the rapid pursuit of swimming prey. Scale-eating often involves 'stealth and ambush' (Sazima, 1984, 1990; Grubh & Winemiller, 2004) and this tactic is unlikely to be effective against a potential prey species that usually lives higher in the water column, and above most of the plants and substrate irregularities that could provide cover for a lurking pupfish. Horstkotte & Strecker (2005) found that *Gambusia sexradiata*, a similar species, was not a significant component of the diets of all members of the Laguna Chichancanab pupfish species flock except *Cyprinodon maya*, (a very large piscivore without known counterpart in the San Salvador fauna).

The two other fishes on the island are the killifish, *Kryptolebias marmoratus*, which frequents fossorial habitats that pupfishes seldom enter, and a lacustrine population of the marine silverside, *Atherinomorus stipes*, a generally pelagic, benthic macroplanktivore; field observations suggest that pupfishes ignore its presence.

**SAMPLES AND LOCALITIES**

A total of 238 individuals from five localities (Fig. 2), collected in June 1999 (Watling’s Blue Hole, Clear Pond) and 2002 (all other localities), were surveyed, Sample designations (used throughout), localities, and sizes were: LLBD, Little Lake Bulldogs (*N* = 26); LLNO, Little Lake Normals (*N* = 28); OLBD, Osprey Lake Bulldogs (*N* = 16); OLNO, Osprey lake Normals (*N* = 67); CRBD, Crescent Pond Bulldogs (*N* = 16); CRNO, Crescent Pond Normals (*N* = 28); WBNNO, Watling’s Blue Hole Normals (*N* = 15); CLNO, Clear Pond Normals (*N* = 28) (n.b. only the normal morphotype was taken at the latter two localities); and BOZO, Bozo morphotype, pooled sample, Little Lake and Osprey Lake (*N* = 14).

Little Lake and Osprey Lake are typical shallow Bahamian lakes (*sensu* Sealey, 1994) without obvious connections to the sea and usually somewhat saltier; both are continuous with the central Great Lake system. Crescent Pond is proximate to Osprey Lake but is flanked to the west and south by a ridge approximately 13 m in height that probably isolates it.

---

**Figure 2.** Outline map of San Salvador, showing collection localities. 1, Little Lake; 2, Osprey Lake; 3, Crescent Pond; 4, Clear Pond; 5, Watling’s Blue Hole.
from the Great Lake system, except during the most extreme flooding. It is deep, tidally influenced, has some of the features of a ‘blue hole’, and is apparently connected to other waters by underground fissures. Watling’s Blue Hole is a typical ‘blue hole’ sensu Mylroie et al. (1995), Clear Pond is separated from the adjacent sea by a shallow sand berm. Both of the latter appear to be isolated from the Great Lake system. Descriptions of the geology, hydrography, and ecology of all of these localities are included in broader discussions of the San Salvador lakes by Gerace (1973), Gerace et al. (1983; Gerace, Ostrander & Smith 1998), Teeter (1983, 1995), Teeter & Bain (1984), Curran et al. (1985), Carew & Mylroie (1994), Hagey & Mylroie (1995) and Godfrey et al. (1994).

Fish were mostly collected with unbaited ‘Gee’ minnow traps set on the substrate (approximate depth 0.8–1.5 m) for 4–16 h within 5 m of the shore. Trapping was supplemented with seining where feasible. In Crescent Pond, the deep, flocculent substrate made trapping ineffective and dipnets were used almost exclusively. Morphotypes were sorted by eye; a small number of mostly immature individuals that could not be readily classified were discarded. Fish were fixed immediately after sorting in absolute ethanol or 91% isopropanol (two changes within 24 h). Genomic DNA was extracted with the aid of Aquapure Genomic DNA Tissue Kits (Bio-Rad). An ABI 310 Genetic Analyzer (Applied Biosystems) was used to analyze polymerase chain reaction (PCR) products, and GENEMAPPER, version 3.0 (Applied Biosystems) was used to score the raw data.

Microsatellite survey

Six of the eight dinucleotide loci surveyed (prefix ‘CnAc’ ) were among those isolated by Burg, Wilcox & Martin (2002) from Cyprinodon nevadensis mionectes and found by them to be scorable and polymorphic in C. variegatus. Two (‘WSP’) were winnowed from those developed in the White Sands Pupfish (Cyprinodon tularosa) by Iyengar et al. (2004) by an initial screen of four of their primer pairs with C. variegatus templates. Primer sequences and amplification conditions were as specified by the original authors.

Statistical analysis

CONVERT, version 1.31 (Glaubitz, 2004) was used to interconvert data file formats. Unless otherwise noted, the Bonferroni procedure (Rice, 1989) was used to establish the threshold for statistical significance of P-values generated by multiple comparisons. However, this correction was not used when testing joint null hypotheses in multiple locus comparisons (Ryman & Jorde, 2001).

Null alleles

An individual was scored as a homozygote for a presumptive ‘null’ allele if two attempts at PCR amplification with a given primer pair were unproductive. Null alleles were apparently present at significant frequency at some loci in some of the populations surveyed. Because null alleles can cause FST values to be seriously overestimated in population comparisons (Chapuis & Estoup, 2007), several methods of minimizing their effects were employed. These are discussed below under the specific comparisons in which they were used.

Intrapopulation variation

Observed and expected levels of heterozygosity per locus were obtained with FSTAT, version 2.9.3.2 (Goudet, 1995; http://www2.unil.ch/popgen/softwares/fstat.htm). FST values were calculated with the same software and their significance evaluated with 7200 randomizations. Comparisons with expected Hardy–Weinberg equilibrium (HWE) heterozygosities were also made with ARLEQUIN, version 3.11 (Excoffier, Laval & Schneider, 2005; http://cmpg.unibe.ch/software/arlequin3) using 500 dememorization steps and Markov chains with 100 000 steps.

Interpopulation divergence

Population comparisons were made using measures based only on allele or genotype identity (exact tests, FST). Copy number variation (e.g. RST) was not considered. This is a conservative approach because Gaggiotti et al. (1999) have shown that FST is always the better estimator of gene flow/population divergence when the number of loci is low and sample sizes are small to moderate. Allele designations used in the tables accordingly reflect relative electrophoretic mobility only.

Divergence was assessed with both ‘global’ and pairwise comparisons. Allele frequencies were compared with the exact contingency table test of Raymond & Rouset (1995a) as implemented in GENEPOL, version 3.4 (Raymond & Rouset, 1995b) or ‘GENEPOP on the Web’ (http://genepop.curtin.edu.au/) with one million iterations (1000 batches with 1000 iterations per batch), and also with conventional chi-square tests and Fisher’s exact tests as implemented in CHIFISH (Ryman, 2006) with 3 750 000 iterations (500 batches with 7500 iterations per batch) used for the latter. For a discussion of the relative strengths of these tests, see Ryman et al. (2006). Multilocus null hypotheses were tested with and without data from CnAc1. Genotype frequency comparisons were made with the exact G-test of Goudet et al. (1996) as implemented in GENEPOP (1 000 000 iterations) and with the permutation test (not assuming HWE) in FSTAT, version 2.9.3.2.
(Goudet, 1995; http://www2.unil.ch/popgen/softwares/fstat.htm), using default parameters. Genotype frequency comparisons should not be affected by null alleles.

$F_{ST}$ values over all populations were computed as per Weir (1996) in FSTAT, version 2.9.3.2. Their confidence intervals per locus were obtained by jackknifing over all populations, and per population by bootstrapping over loci. Pairwise $F_{ST}$ values were computed in ARLEQUIN, version 3.11 using pairwise distances, and their $P$-values calculated with 5040 permutations. FREENA (http://www.montpellier.inra.fr/URLB/) was used to estimate null allele frequencies for each locus using the expectation maximum algorithm of Dempster, Laird & Rubin (1977) and to calculate unbiased $F_{ST}$ values (Weir, 1996) with the ‘ENA’ method of Chapuis & Estoup (2007). These methods assume that there is only a single null allele per locus.

Analysis of molecular variance (AMOVA; Excoffier, Smouse & Quattro, 1992) was conducted in ARLEQUIN, version 3.11. Three models of hierarchical population structure were examined: no groups; groups = morphotypes; and groups = lakes.

The ‘INA’ output file of corrected $F_{ST}$ estimates furnished by FREENA was used to compute Cavalli-Sforza & Edwards (1967) genetic distances between populations, as suggested by (Chapuis & Estoup, 2007), with GENDIST in the PHYLIP software package (http://evolution.genetics.washington.edu/phylip.html). SEQBOOT and CONSENSE from the same package were used to bootstrap those distances, and their $P$-values were computed in ARLEQUIN, version 3.11 using pairwise bootstrap over loci. Pairwise $F_{ST}$ values across all populations are given in Table 1. A maximum of eight alleles per locus (WSP26) and a minimum of three ($CnAc1$ and $CnAc25$) were

Population structure

Inferences of population structure were made using STRUCTURE, version 2.2. (Falush, Stephens & Pritchard, 2007; http://pritch.bsd.uchicago.edu/software), which can treat null alleles as recessives. For each locus, all missing data were scored as the result of same null allele in every population, introducing an assumption roughly equivalent to that of the ENA method for $F_{ST}$ calculations (see above). For all analyses, the hyperparameter lambda was inferred at $K=1$, as suggested in the software documentation (Pritchard, Wen & Falush, 2007), with a burnin of 20 000 and a Markov chain of 100 000 iterations. In most cases, the average value of lambda was inferred to be less than 0.5. The inferred lambda value (rather than the default value of 1.0) was used in runs at $K=2$ to $K=10$ (five replicates per $K$-value, where $K$ is the number of assumed clusters), with burnins of 30 000 and 300 000 iterations. Initially, the ‘admixture’ ancestry model, assuming correlated allele frequencies, was used and population information was not included in the analysis. In a second set of analyses, population information was utilized (setting ‘USEPOP’), with ‘GENSBACK’ set at 3 and ‘MIGPRIOR’ at 0.001. These parameter values were used by Falush et al. (2007) to demonstrate hybridization between sympatric normal and dwarf forms of lake whitefish. In a third set of analyses, ‘learning samples’, as described by Pritchard et al. (2007), were used to further explore relationships among the normal and bulldog morphs in different lakes. The software was set-up to make use of population information for both morphotypes in one lake, but not the other two, so that it could potentially group the morphs in the other lakes, and the other populations, with the appropriate original clusters or establish new groupings, depending on which arrangement was more probable. Bulldogs and normals from Little Lake, Osprey Lake, and Crescent Pond were used conjointly as learning samples in separate runs.

The Kolomogorov–Smirnov comparison test was used to assess the statistical significance of populations that appeared to have different levels of admixture by visual inspection of the STRUCTURE results. The version of the test that we used (http://www.physics.csbsju.edu/cgi-bin/stats) can also identify individuals as statistical outliers by nonparametric ('Tukey') criterion.

RESULTS

Allele frequency data, heterozygosities, and $F_{ST}$ values across all populations are given in Table 1. A maximum of eight alleles per locus (WSP26) and a minimum of three ($CnAc1$ and $CnAc25$) were
<table>
<thead>
<tr>
<th>Locus:</th>
<th>CnAC1</th>
<th>CnAC10</th>
<th>CnAC14</th>
<th>CnAC17</th>
<th>CnAC25</th>
</tr>
</thead>
<tbody>
<tr>
<td>N:</td>
<td>18</td>
<td>25</td>
<td>26</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>Alleles:</td>
<td>94% 0.52</td>
<td>96% 0.93</td>
<td>81% 0.39</td>
<td>80% 0.52</td>
<td>80% 0.52</td>
</tr>
<tr>
<td>Ho:</td>
<td>0.53</td>
<td>0.51</td>
<td>0.54</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>N:</td>
<td>24</td>
<td>23</td>
<td>15</td>
<td>65</td>
<td>63</td>
</tr>
<tr>
<td>Alleles:</td>
<td>94% 0.52</td>
<td>96% 0.93</td>
<td>81% 0.39</td>
<td>80% 0.52</td>
<td>80% 0.52</td>
</tr>
<tr>
<td>Ho:</td>
<td>0.53</td>
<td>0.51</td>
<td>0.54</td>
<td>0.52</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 1. Allele frequency comparisons at eight microsatellite loci for nine samples of San Salvador pupfishes
<table>
<thead>
<tr>
<th>Locus</th>
<th>Alleles</th>
<th>Ho</th>
<th>N</th>
<th>26</th>
<th>28</th>
<th>16</th>
<th>59</th>
<th>16</th>
<th>28</th>
<th>15</th>
<th>28</th>
<th>14</th>
<th>230</th>
<th>0.192 ±0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>CnAC29</td>
<td>137:</td>
<td>0.923</td>
<td>0.768</td>
<td>0.969</td>
<td>0.712</td>
<td>1.000</td>
<td>0.911</td>
<td>1.000</td>
<td>0.661</td>
<td>0.964</td>
<td>0.833</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>139:</td>
<td>0.019</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>143:</td>
<td>0.058</td>
<td>0.232</td>
<td>0.031</td>
<td>0.280</td>
<td>&lt; 0.001</td>
<td>0.089</td>
<td>&lt; 0.001</td>
<td>0.339</td>
<td>0.036</td>
<td>0.163</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>145:</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.009</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho:</td>
<td>0.147</td>
<td>0.362</td>
<td>0.063</td>
<td>0.419</td>
<td>&lt; 0.001</td>
<td>0.165</td>
<td>&lt; 0.001</td>
<td>0.458</td>
<td>0.071</td>
<td>0.189</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus:</td>
<td>WSP24</td>
<td>[&lt; 0.001]</td>
<td>[&lt; 0.001]</td>
<td>[&lt; 0.001]</td>
<td>[0.257]</td>
<td>[0.001]</td>
<td>[&lt; 0.001]</td>
<td>[0.001]</td>
<td>[0.411]</td>
<td>[0.203]</td>
<td>[0.083]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alleles:</td>
<td>101:</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.268</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.032</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>107:</td>
<td>0.673</td>
<td>0.768</td>
<td>0.625</td>
<td>0.746</td>
<td>0.219</td>
<td>0.482</td>
<td>0.867</td>
<td>0.946</td>
<td>0.577</td>
<td>0.688</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>111:</td>
<td>0.115</td>
<td>&lt; 0.001</td>
<td>0.063</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>137:</td>
<td>&lt; 0.001</td>
<td>0.018</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.054</td>
<td>0.001</td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>193:</td>
<td>0.212</td>
<td>0.214</td>
<td>0.313</td>
<td>0.254</td>
<td>0.781</td>
<td>0.250</td>
<td>0.133</td>
<td>0.001</td>
<td>0.423</td>
<td>0.255</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho:</td>
<td>0.497</td>
<td>0.373</td>
<td>0.531</td>
<td>0.382</td>
<td>0.358</td>
<td>0.646</td>
<td>0.238</td>
<td>0.103</td>
<td>0.506</td>
<td>0.404</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus:</td>
<td>WSP26</td>
<td>[&lt; 0.001]</td>
<td>[0.098]</td>
<td>[0.128]</td>
<td>[0.018]</td>
<td>[0.134]</td>
<td>[0.027]</td>
<td>[&lt; 0.001]</td>
<td>[&lt; 0.001]</td>
<td>[0.218]</td>
<td>[0.049]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alleles:</td>
<td>176:</td>
<td>0.019</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.016</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.009</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>178:</td>
<td>&lt; 0.001</td>
<td>0.056</td>
<td>&lt; 0.001</td>
<td>0.023</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.115</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>180:</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.031</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.037</td>
<td>0.039</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>182:</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.008</td>
<td>0.031</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.039</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>184:</td>
<td>0.096</td>
<td>0.407</td>
<td>0.188</td>
<td>0.422</td>
<td>0.813</td>
<td>0.839</td>
<td>0.650</td>
<td>0.389</td>
<td>0.615</td>
<td>0.463</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>186:</td>
<td>0.808</td>
<td>0.370</td>
<td>0.750</td>
<td>0.406</td>
<td>0.156</td>
<td>0.143</td>
<td>0.350</td>
<td>0.389</td>
<td>0.115</td>
<td>0.401</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>188:</td>
<td>0.077</td>
<td>0.056</td>
<td>&lt; 0.001</td>
<td>0.031</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.037</td>
<td>&lt; 0.001</td>
<td>0.029</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>190:</td>
<td>&lt; 0.001</td>
<td>0.111</td>
<td>0.031</td>
<td>0.094</td>
<td>&lt; 0.001</td>
<td>0.180</td>
<td>&lt; 0.001</td>
<td>0.148</td>
<td>0.039</td>
<td>0.064</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho:</td>
<td>0.340</td>
<td>0.689</td>
<td>0.415</td>
<td>0.652</td>
<td>0.327</td>
<td>0.280</td>
<td>0.500</td>
<td>0.687</td>
<td>0.609</td>
<td>0.499</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Null:</td>
<td>0.051</td>
<td>[0.057]</td>
<td>[0.022]</td>
<td>[0.109]</td>
<td>[0.065]</td>
<td>[0.029]</td>
<td>[0.585]</td>
<td>[0.119]</td>
<td>[0.117]</td>
<td>[0.085]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho:</td>
<td>0.236</td>
<td>0.415</td>
<td>0.230</td>
<td>0.388</td>
<td>0.101</td>
<td>0.307</td>
<td>0.182</td>
<td>0.272</td>
<td>0.386</td>
<td>0.246</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Upper value as defined by Weir (1996); lower value with ENA correction for null alleles (Chapuis & Estoup, 2007).
†Exact P-values for genic (upper) and genotypic (lower) comparisons (GenePop 3.4).
‡Estimates of presumptive null allele frequencies using the EM algorithm of Dempster et al. (1977) as per Chapuis & Estoup (2007).
§Individual genic χ² values: CnAc1, 116 (d.f. = 16), P < 0.001; CnAc10, 210 (d.f. = 40), P < 0.001; CnAc14, 60.0 (d.f. = 40), P = 0.002; CnAc17, 86.9 (d.f. = 32), P < 0.001; CnAc25, 19.9 (d.f. = 16), P = 0.223; CnAc29, 63.3 (d.f. = 24), P < 0.001; WSP24, 247 (d.f. = 32), P < 0.001; WSP26, 178 (d.f. = 56), P < 0.001.
detected, of which a maximum of seven and a minimum of one are found in a single population. Eight apparent ‘private’ alleles occur among five (of nine) populations, of which two are present in LLNO, two in OLNO, two in BOZO, and one each in LLBD and CRNO. Observed heterozygosities per locus are in the range 0–0.689 when all samples are considered. If loci are considered singly, the lowest range of heterozygosity values is 0.000–0.071 (CnAc25) and the highest is 0.000–0.624 (CnAc17). Average observed heterozygosity per locus range from 0.014–0.499 (mean value = 0.329) and per population from 0.182 to 0.388 (mean value = 0.246). In general, departures from heterozygosity levels predicted by HWE were not significant. Eight $F_{IS}$ values are individually significant by simulation in FSTAT (data not shown), with a deficiency of heterozygotes in each case, but were not significant with the Bonferroni correction for multiple comparisons. Five of these cases were associated with high frequencies of putative null alleles (see below).

Presumptive null alleles are present in nearly all samples. Estimates of mean null allele frequencies using the EM algorithm in FREENA range from 0.041 (CnAc17) to 0.325 (CnAc1), and, with the latter locus, frequency estimates in some individual samples reach very high values (e.g. 0.775 in WBNO, 0.674 in CRNO). Accordingly, some allelic comparisons were made with this locus excluded.

Considerable genetic divergence among populations is evident in Table 1 from both direct tests and $F_{ST}$ values. Exact tests and conventional chi-square comparisons reveal highly significant variation ($P = 0.000$) in allelic frequencies for seven of the eight loci surveyed; only CnAc25 is not significantly divergent. Both the level and the pattern of overall divergence is independent of locus CnAc1, the locus in which a putative null allele is at highest average frequency, because six other loci are approximately equally divergent. This locus contributes only 12% of the overall genic chi-square value, a contribution exceeded by three other loci. The pattern of genotypic variation is completely congruent with the genomic pattern.

Values of $F_{ST}$, uncorrected for null alleles were in the range 0.073–0.291 (mean value = 0.177); values ‘corrected’ for null alleles were in the range 0.087–0.252 (mean = 0.164). The corrected $F_{ST}$ values for four loci are lower than the uncorrected ones (greatest change: −21% of original value for locus CnAc1) as generally predicted by the simulations of Chapuis & Estoup (2007). However, corrected $F_{ST}$ values for four of the loci actually increased (greatest change: +49% of original value for locus CnAc29). In all cases with increases in $F_{ST}$, the presumptive null allele was in relatively high frequency in one or two populations and in much lower frequencies in the others. With all $F_{ST}$ values, the 95% confidence interval (CI) computed in FREENA did not include zero (data not shown, except for global average). Twenty-nine of the 36 multilocus pairwise comparisons (Table 2) are significantly divergent by all criteria: genotypic frequencies, gene frequencies, and $F_{ST}$ values. In all cases, the significant multilocus comparisons comprise at least two, and sometimes as many as four, statistically significant single-locus comparisons.

Among the significant pairwise comparisons are sympatric samples of bulldog and normal morphotypes from Little Lake, Osprey Lake, and Crescent Pond. In these, the gene frequency comparisons generate $F_{ST}$ values in the range 0.104–0.310 (theta of Weir (1996), corrected for null alleles (Table 2); note that the 95% CI of these measures do not include zero). All three of the sympatric morphotype pairs remain highly divergent in gene frequencies if locus CnAc1 is omitted because three or four other loci also generate significant comparisons.

Evidence of genetic divergence between the single, pooled sample of the bozo morphotype and sympatric populations of other morphs is not as straightforward as in bulldog/normal comparisons. It is significantly different by all criteria when compared to Osprey Lake normals ($F_{ST} = 0.074$), and to Little Lake bulldogs ($F_{ST} = 0.157$). It is divergent from Little Lake normals in gene frequencies (Fisher’s exact test: $P = 0.002$ with CnAc1, 0.014 without CnAc1; $F_{ST} = 0.041$ when corrected for null alleles), but genotype frequencies are not significantly different. Compared to Osprey Lake bulldogs, both its genotypic and gene frequencies are significantly divergent (gene frequencies, Fisher’s exact test, $P = 0.0006$; $\chi^2 = 48.6$; d.f. = 18, $P = 0.0001$), but the $F_{ST}$ value of 0.110 (conventional calculation), although substantial in magnitude, is not significant when tested by permutation in Arlequin, possibly because its sampling variance across loci is large (theta corrected for null alleles is 0.148 with a 95% CI of ±0.129).

Samples of both the bulldog and normal morphs from Little Lake are not significantly divergent from their respective counterparts in Osprey Lake, but each morph from both of these lakes differs markedly in gene frequencies from its counterpart in Crescent Pond. For the bulldog morphs, this inter-lake (but within-morph) comparison generates $F_{ST}$ values of approximately 0.308; for the normal morphs, they are in the range 0.116–0.125.

The largest pairwise $F_{ST}$ values, approximately 0.48, result from comparisons of Crescent Pond bulldogs with normals from Clear Pond or Watling’s Blue Hole. Many of the $F_{ST}$ values resulting from pairwise comparisons among normals are less striking but still substantial: Watling’s Blue Hole versus Clear Pond...
Table 2. Pairwise population comparisons over all loci
Above diagonal: Upper entries: Exact tests of genotypic frequencies (GenePop 3.4): χ² (d.f.), P. Note: all significant genotypic comparisons are also significant in FSTAT unless otherwise noted. They are also significant at the same level in GENEPOP when allelic frequencies are compared. Lower entries: F_{ST} values above, P (by permutation in Arlequin 3.1) below.
Below diagonal: Pairwise comparisons correcting for presumptive null alleles. Upper entries: Uncorrected F_{ST} values (Weir, 1996). Lower entries: F_{ST} values with ENA correction of Chapuis & Estoup (2007). Unless otherwise noted, all 95% CIs of these values do not include zero.
Bold type indicates statistically significant differences between sympatric samples of normal and bulldog.

<table>
<thead>
<tr>
<th></th>
<th>LLBD</th>
<th>LLNO</th>
<th>OLBD</th>
<th>OLNO</th>
<th>WBNO</th>
<th>CLNO</th>
<th>CRBD</th>
<th>CRNO</th>
<th>BOZO</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLBD</td>
<td>× (16)</td>
<td>8.178 (12)</td>
<td>× (16)</td>
<td>53.01 (12)</td>
<td>× (12)</td>
<td>× (12)</td>
<td>× (14)</td>
<td>× (16)</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>0.151</td>
<td>0.771</td>
<td>0.260</td>
<td>0.371</td>
<td>0.407</td>
<td>0.345</td>
<td>0.161</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.000</td>
<td>0.828 (NS)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.67 (16)</td>
<td>× (16)</td>
<td>× (16)</td>
<td>× (16)</td>
<td>35.30 (16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLBD</td>
<td>0.159</td>
<td>0.236</td>
<td>0.000</td>
<td>0.110</td>
<td>0.004 (NS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.140</td>
<td>0.000</td>
<td>0.185</td>
<td>0.363</td>
<td>0.318</td>
<td>0.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLNO</td>
<td>0.116</td>
<td>0.128</td>
<td>0.000</td>
<td>43.53 (16)</td>
<td>× (16)</td>
<td>× (16)</td>
<td>× (16)</td>
<td>48.76 (16)</td>
<td>0.000</td>
</tr>
<tr>
<td>0.159</td>
<td>0.139</td>
<td>0.000</td>
<td>0.110</td>
<td>0.297</td>
<td>0.116</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.128</td>
<td>0.020</td>
<td>0.141</td>
<td>0.175 (NS)</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>WNO</td>
<td>0.275</td>
<td>0.267</td>
<td>0.070</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLNO</td>
<td>0.159</td>
<td>0.103</td>
<td>0.154</td>
<td>0.101</td>
<td>0.240</td>
<td>0.215</td>
<td>0.097</td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>0.382</td>
<td>0.095</td>
<td>0.388</td>
<td>0.098×</td>
<td>0.322</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.285</td>
<td>0.091</td>
<td>0.327</td>
<td>0.075</td>
<td>0.263</td>
<td>0.531</td>
<td>0.178</td>
<td>0.244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRBD</td>
<td>0.388</td>
<td>0.306</td>
<td>0.347</td>
<td>0.273</td>
<td>0.504</td>
<td>0.525</td>
<td>0.185</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.381</td>
<td>0.307</td>
<td>0.281</td>
<td>0.481</td>
<td>0.484</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRNO</td>
<td>0.329</td>
<td>0.117</td>
<td>0.297</td>
<td>0.119</td>
<td>0.244</td>
<td>0.190</td>
<td>0.250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.285</td>
<td>0.116</td>
<td>0.282</td>
<td>0.125</td>
<td>0.208</td>
<td>0.196</td>
<td>0.310</td>
<td></td>
<td></td>
<td>0.068</td>
</tr>
<tr>
<td>BOZO</td>
<td>0.186</td>
<td>0.060</td>
<td>0.142</td>
<td>0.087</td>
<td>0.151</td>
<td>0.172</td>
<td>0.080</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>0.157</td>
<td>0.041</td>
<td>0.148</td>
<td>0.074</td>
<td>0.158</td>
<td>0.198</td>
<td>0.223</td>
<td>0.103</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Critical P-value with Bonferroni correction = 0.0014.
†Negative F_{ST} values (undefined) reported as ‘0.’
§HS = ‘Highly significant’. Note: GENEPOP evaluates the joint null hypothesis for multiple loci with Fisher’s combined probability test: χ²C = −2 Σ ln(Pi), where Pi is the probability of the null hypothesis for each individual locus. When that probability is zero for one or more loci, the software reports an infinite χ²C.
¶Genotypic comparison not significant in FSTAT but gene frequency comparison significant in GENEPOP.
**Comparison without locus CnAc1 (CHIFISH); Fisher’s exact test, overall P = 0.000; four loci with significant P-values. Overall χ² = 80.6; d.f. = 22; P = 0.000; four loci with significant P-values.
††Comparison without locus CnAc1 (CHIFISH); Fisher’s exact test, overall P = 0.000; four loci with significant P-values. Overall χ² = 62.4; d.f. = 21; P = 0.000; four loci with significant P-values.
‡‡Comparison without locus CnAc1 (CHIFISH); six loci only; Fisher’s exact test, overall P = 0.000; three loci with significant P-values. Overall χ² = 66.3; d.f. = 10; P = 0.000; three loci with significant P-values.
The Bozo sample appears to be differentiated from most allopatric populations. AMOVA indicated that variance among and within morphotypes explains nearly all of the variation among populations. Variation within and among morphotypes is far more trenchant than variation among lakes. Approximately 80% of the total variation occurs among individuals within populations. If the ‘groups’ are lakes, then more than 80% of the non-individual variation is within groups (i.e. among morphotypes; different populations within the same lake are different morphs); the fraction among lakes per se appears inconsequential. If the ‘groups’ are morphotypes, the non-individual variation is divided roughly evenly between among-group and within-group components (i.e. there is approximately the same fraction of total variation among and within morphotypes). In other words, from the perspective of analysis of variance, variation among lakes is discernible only when it is measured as variation within morphotypes.

The unrooted UPGMA cladogram of Cavalli-Sforza & Edwards (1967) chord distances (Fig. 3) does not resolve relationships among most of the samples. However, one very well supported node separates bulldog morphotypes from Little Lake and Osprey Lake from all other samples, including sympatric normal and bozo morphotypes and including both morphs from Crescent Pond.

Four instances of apparent linkage disequilibrium (i.e. nonrandom allelic association) were detected in three population samples: CnAc10/CnAc17 and CnAc10/Wsp26 in Osprey Lake bulldogs ($P = 0.001$ in each case), CnAc17/CnAc29 in Little Lake normals ($P = 0.005$) and Wsp24/Wsp26 in Crescent Pond bulldogs ($P = 0.001$). In each case, the apparent nonrandom association was evident in only one member of a sympatric morphotype pair (e.g. Osprey Lake bulldogs but not Osprey Lake normals).

The first two axes of the factorial correspondence analysis (Fig. 4) account for 54% of the total individual genotype variation. Bulldog and normal morphs are well separated along axis 1, which also separates the Clear Pond normal population from the others. Sympatric morphotype pairs from Little Lake, Osprey Lake, and Crescent Pond tend to separate along axis 2. Individuals of the bozo morph are not well separated from the other samples.

Model-based cluster analyses in STRUCTURE (Fig. 5A) reach maximum likelihood at $K = 2$ ($K$ = number of ‘ideal’ clusters or subpopulations posited) when the default parameters are used (admixture model, correlated allele frequencies). (As used here, the term ‘admixture’ refers to the extent that the samples can be assigned membership in conceptually idealized subpopulations derived from the data; it does not necessarily indicate actual
hybridization.) In this analysis, every individual is 'admixed' but, in Little Lake, Osprey Lake and Crescent Pond, bulldog, and normal samples are clearly distinct in admixture proportions, although there is considerable variation among individuals. This distinction is evident by Kolmogorov–Smirnov comparison of admixture proportions ($P = 0.000$). The three bulldog populations are indistinguishable, as are Little Lake, Osprey Lake, Crescent Pond, and Clear Pond normal samples. The Watling’s Blue Hole normal sample is divergent from the other normal samples (Kolmogorov–Smirnov: $P = 0.001$ in all four pairwise comparisons), but is not distinct from the bulldog samples. The bozo sample is not distinct from either bulldog or normal samples.

Kolmogorov–Smirnov comparison identifies two individuals in the Little Lake bulldog sample and one individual in the Osprey Lake bulldog sample as outliers. The original morph assignments of these individuals were re-examined visually, and no basis for a change of assignment was noted. Note that the ‘outliers’ were included in the Kolmogorov–Smirnov comparisons of the population admixture data.

When STRUCTURE uses population identity as a ‘prior’ in its analyses, maximum likelihood is achieved at $K = 9$ or 10. At $K = 9$, all population samples are clearly distinguished (Fig. 5B). A single individual with probable ‘mixed’ ancestry is identified in the Osprey Lake Bulldog sample. This is the same individual that was identified in this sample as an outlier.
in the Kolmogorov–Smirnov analysis of admixture proportions. At MIGPRIOR = 0.01, a second individual has significant genetic contributions from other populations (not shown).

In all three of the analyses with ‘learning samples’, likelihood maxima are attained at $K = 5$ when the admixture model is used for the ‘nonspecified’ populations. In each, the two nonfixed morphotype pairs are distinct from each other, but the divergence between bulldog and normal morphotypes is retained as well. In the example shown (where the Little Lake bulldogs and normals are the fixed samples; Fig. 5C) the software makes use of the Little Lake bulldog and normal components in partitioning the Osprey Lake samples, but it introduces admixture with three other components as well. Crescent pond bulldogs and normals are quite distinct from each other, and each sample is divergent from its counterparts in the other two lakes and is an admixture of components that are rare to absent in the Little Lake morphotype pair. The Watling’s Blue Hole normal sample contains more of the major Little Lake Bulldog component than do the Crescent pond bulldogs, and the Clear Pond normals contain a major component that is otherwise found mostly in the Osprey Lake normals. The bozo sample is very similar to the Osprey Lake normal population. (Note that if the ‘no admixture’ option is used, the maximum log likelihood is at $K = 3$, but most of the divergences noted above are still visible.) In all three ‘learning sample’ analyses, the same single individual in the Osprey Lake bulldog sample was identified as an outlier by Kolmogorov–Smirnov as was ascribed to mixed ancestry previously.

**DISCUSSION**

The endemic San Salvador pupfish morphotypes, bulldog and bozo, differ markedly from each other and from the sympatric, more common normal morph in traits that many ichthyologists routinely use, often without genetic data, to delineate congeneric species. Holtmeier (2000) has shown that the divergent trophic features of both morphotypes are not environmentally-induced, ecophenotypic modifications but have demonstrable genetic bases: They breed true through the second laboratory generation, and progeny of inter-morphotype matings have intermediate phenotypes.

However, in this case, the presence of reproductive isolation must be assessed directly because the evolutionary biology of pupfishes in general predicts *a priori* that its emergence among the San Salvador morphs is unlikely, despite their marked morphological and ecological divergences. The recent evolutionary history of the approximately 50 taxa that comprise the genus *Cyprinodon* is one of divergence in almost complete allopatry. Postmating isolation is seldom evident (Turner & Liu, 1977). Premating isolating mechanisms have not been studied in detail, but there is no reason to suppose that they are generally well developed, or even that they exist at all, because sympatry of divergent forms (anthropogenic in many cases) usually leads to the rapid formation of hybrid swarms and/or complete introgression (Echelle et al., 2005). Mate preferences on the part of females of one species actually enhance its hybridization with males of introduced *C. variegatus* (Rosenfeld & Kodric-Brown, 2003). Nor is there any reason to argue that premating isolation might be more likely/effective among forms that are highly divergent ecologically. For example, two endemic pupfishes in the Quatro Ciéncegas basin (Mexico), *Cyprinodon atrorus* and *Cyprinodon bifasciatus*, are more ecologically and morphologically divergent than almost any two of their congeners (Miller, 1968) and are markedly divergent behaviourally as well (Arnold, 1972). Yet the species hybridize extensively in intermediate habitats (Minckley, 1969; Carson & Dowling, 2006) and there is compelling phyleogeographic evidence of ancient hybridizations as well (Carson & Dowling, 2006). The five specialized components of the endemic pupfish species flock in Laguna Chichancanab, Yucatan, are the most prominent exception to the ‘no sympathy without introgression rule’ in this group (Strecker, 2006; for a second exception, but one that involves postmating isolation, see Tech, 2006). Even in this case, however, genetic divergence between any of the novel forms and *Cyprinodon beltrani*, their presumptive immediate ancestor, could not be demonstrated with either microsatellite data or mitochondrial DNA control region haplotypes, so the possibility of indirect genetic contact among the other endemics cannot be excluded, and ‘unidentifiable specimens’ (i.e. presumptive hybrids) occur in the lake in fair numbers.

For the San Salvador pupfishes, intraspecific trophic polymorphism would be the most likely alternative to speciation. Such polymorphisms are well known among fishes (Skulason & Smith, 1995; Smith & Skulason, 1996; for some more recent cases, see Dimmick, Berendzen & Golubtsov, 2001; Ruzzante et al., 2003) and they include several examples (lake charr: Blackie, Weese & Noakes, 2003; artic charr: McCarthy et al., 2004; Cuatro Cienegas cichlid: Swanson et al., 2003) in which one of the morphotypes is piscivorous and can prey upon the other(s). The distinctive alternative phenotypes frequently have substantial genetic as well as environmental and/or developmental components (Skulason et al., 1993; Klemetsen et al., 2006).

The microsatellite data falsify the intraspecific polymorphism hypothesis for the bulldog and normal...
morphotypes. In three lakes in which these morphs are sympatric, they are genetically divergent. That is, the data indicate that some level of reproductive isolation exists between the two morphotypes in at least three instances. The model-based analysis in STRUCTURE provides perhaps the most compelling evidence of genetic distinctiveness. Even when the software uses no information on population identification, it divides each of the three lake samples into two components that have different levels of putative admixture relative to two idealized subpopulations. With very little overlap between them, these components correspond in each case to the bulldog and normal morphs in each lake. The factorial correspondence analysis also shows well-marked distinctions between the multilocus genotypes of two morphotypes in each of the three lakes in which they were sampled. Traditional allele frequency comparisons of normal and bulldog generate multilocus $F_{ST}$ values that are statistically significant by permutation, and whose 95% CI do not include zero, even when corrected for the potentially exaggerating effects of null alleles. Gene frequencies are also divergent when compared directly. In both comparisons, the significant differences are the result of at least three loci with individually significant allelic frequency differences, two of which have only low frequencies of null alleles. Genotype frequency differences are also evident, an important observation because genotype frequency comparisons are not affected by null alleles. Finally, linkage disequilibrium for at least one allelic combination has been detected in three cases, and in each case, it was found in only one member of a sympatric morphotype pair, and this observation cannot be easily explained if the lake populations were panmictic. It is likely that reproductively isolated isolation between bulldog and normal morphotypes involves mostly premating isolating mechanisms because all three are interfertile at least to $F_2$ in laboratory crosses (Holtmeier, 2000, 2001). Field observations of normals and bulldogs in Little Lake (M. G. Barton & C. Barton, unpublished data) support this suggestion.

In all three localities with both bulldog and normal morphs (total $N = 181$), only a single specimen (0.06%), a bulldog from Osprey Lake, was clearly identified as of mixed or hybrid ancestry when the criteria of Falush et al. (2007) were used in the STRUCTURE analysis, and this number increased only to two when the parameter MIGPRIOR (i.e. the proportion of individuals who are migrants from all other populations) was increased ten-fold to 0.01. This suggests that the rate of hybridization between sympatric normal and bulldog morphs is low. This conclusion is obviously tentative because we have no idea of the real rates of mixed mating between the morphs (the equivalent of MIGPRIOR for sympatric populations). Also, we do not consider as possible hybrids the individuals that were termed ‘outliers’ (i.e. in effect, assigned to the ‘other’ population), when USEPOP was not specified. We attribute their anomalous admixture scores to discrepancies between real data and conceptualized ‘ideal’ populations. This judgment is subjective and it may be overly conservative, although the presumptive frequency of hybridization is still quite low even if these specimens are also considered as hybrids. We emphasize that none of these individuals, including the one that met the criteria of Falush et al. (2007), had an intermediate phenotype (as do laboratory $F_1$ and $F_2$ inter-morph hybrids) or was obviously misclassified by its morphology. The apparently low level of inter-morphotype hybridization in the two San Salvador lakes appears to be rather different from the situation described by Streeker (2006) in the Laguna Chichancanab pupfish flock. It is worth noting that this analysis was influenced by the presence of null alleles. In runs of STRUCTURE in which presumptive null homoygotes are simply treated as missing data, the criteria used by Falush et al. (2007) suggest that four or five additional individuals are of hybrid ancestry.

Gene and genotype frequency comparisons, the results of the factorial correspondence analysis and of the learning samples in STRUCTURE, as well as the topology of the UPGMA tree of chord distances, suggest that the bulldog morphs from Crescent Pond may be distinct from their counterparts in Little Lake and Osprey Lake. These observations are arguably consistent with a possible separate evolutionary origin for the Crescent Pond bulldogs, but they do not eliminate other potential scenarios. Detailed morphological and/or behavioural comparisons of these populations have not yet been made.

The microsatellite data are not readily compatible with recent intralacustrine origin(s) of bulldog morphs in Little Lake and Osprey Lakes from their respective current sympatric normal populations. On the UPGMA tree, the bulldogs in these lakes are separated from their sympatric normal populations by a very well supported node, suggesting that the morphotype pairs are unlikely to be considered sister taxa when better developed phylogenetic analyses become available. That relationship is crucial to any hypothesis of intralacustrine speciation in the current lakes. AMOVA suggests that division into morphotypes explains far more of the variation in the data than does division into lakes, a pattern opposite to that expected with intralacustrine/sympatric origin. If, as we have argued earlier, the bulldog morphotype necessarily evolved in contact with normal pupfish, this finding suggests those in Little Lake/Osprey Lake may have evolved by an earlier intralacustrine divergence from a different normal stock, perhaps one...
in another lake. Alternatively, they evolved as a consequence of secondary contact with a hypothetical, genetically different, less specialized direct ancestor (i.e. their trophic features are a form of character displacement).

Most of the traditional allelic comparisons suggest that the bozo morphotype is genetically divergent from the other two in both of the lakes in which it was collected. However, neither factorial correspondence analysis, nor model-based analyses in STRUCTURE support this conclusion. Better samples will clearly be needed to address the question more fully. We present data from a marginal sample in the present study because the consistent rarity of the morphotype in our collections leads us to fear that better samples cannot be readily obtained without negative impact on the natural populations.

The pairwise $F_{ST}$ values among all of our samples are rather higher than those reported by Streeker (2006) for components of the Laguna Chichancanab pupfish species flock (San Salvador data: 0.020–0.484; Chichancanab data: 0.0002–0.0380; only two values overlap between the two sets). This comparison should be interpreted cautiously because $F_{ST}$ values for microsatellite loci are constrained by population heterozygosity (Hedrick, 1999), and the average heterozygosities of the Chichancanab populations (0.725–0.789) are much higher than those of the San Salvador samples (0.101–0.388). Moreover, the comparisons are based on small numbers of loci (five in Chichancanab, eight in San Salvador) with no overlap, and could be subject to sampling bias. However, comparison of the faunas clearly merits further study because each has supposedly evolved within a similar time frame (8000 years for Chichancanab, 6000 for San Salvador) with no overlap, and could be subject to sampling bias. However, comparison of the faunas clearly merits further study because each has supposedly evolved within a similar time frame (8000 years for Chichancanab, 6000 for San Salvador) with no overlap, and could be subject to sampling bias.

Our data strongly suggest that more detailed studies of the biology of the San Salvador pupfishes will prove informative. The bulldog pupfish on this island are an example of the evolution of apparently efficient reproductive isolation in a group in which it is otherwise quite rare. It will be important to identify the isolating mechanisms between the bulldog and normal morphs, and especially to determine whether the behaviours that might be involved are related to predator–prey interactions. It will also be important to learn if the bulldogs from Crescent Pond indeed had a separate origin from those in Little Lake and/or Osprey Lake because, if that were the case, they would represent a striking example of parallelism that could provide much insight into the relationships between ecological divergence and the evolution of reproductive isolation.

**ACKNOWLEDGEMENTS**

Grant Gilmore, the original discoverer of the San Salvador pupfish endemics, was an early source of information and encouragement, although he does not necessarily accept all of our conclusions. The Gerace Marine Laboratory, San Salvador, provided hospitality and field support and arranged for collecting permits from the Bahamian government. Cami Holtmeier provided some of the specimens used in our surveys, and graciously shared her own knowledge of the San Salvador pupfishes, and Michael M. Stevenson generously shared his data on their diets. John E. Mylroie helped us understand the geology and paleohydrography of San Salvador. Marie-Pierre Chapuis kindly provided an advance copy of a version of her FreeNA software that could compute pairwise comparisons. Jason A. Fonoti assisted with extraction and analyses of DNA samples. We thank William P. Davis, and D. Scott Taylor for assistance in the field, and Chris Barton and Michael A. Bell for stimulating discussions. Supported by grants from the Jeffress Memorial Trust to B.J.T., various VPISU research monies, and National Science Foundation grant OCE-0221879 to D.D.D.

**REFERENCES**


