Extensive outcrossing and androdioecy in a vertebrate species that otherwise reproduces as a self-fertilizing hermaphrodite

Mark Mackiewicz*1, Andrey Tatarenkov†‡, D. Scott Taylor§, Bruce J. Turner¶ and John C. Avise*†

*Department of Genetics, University of Georgia, Athens, GA 30602; †Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; §Brevard County Environmentally Endangered Lands Program, Melbourne, FL 32940; and ¶Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Contributed by John C. Avise, May 11, 2006

The mangrove killifish (Kryptolebias marmoratus) is the only vertebrate known to be capable of self-fertilization. Its gonad is typically an ovotestis that simultaneously produces eggs and sperm, and fertilization is internal. Although most populations of this species consist primarily or exclusively of hermaphroditic individuals, gonochoristic males occur at ~20% frequency in a natural population at Twin Cays, Belize. Here we use a battery of 36 microsatellite loci to document a striking genetic pattern (high intraspecific heterozygosities and low within-population linkage disequilibria) that differs qualitatively from the highly homozygous (or “clonal”) genetic architecture characteristic of killifish populations previously studied in Florida, where males are much rarer. These findings document that outcrossing (probably between gonochoristic males and hermaphrodites) is common at the Belize site, and, more importantly, they demonstrate the dramatic impact that functional androdioecy can have on the population genetic architecture of this reproductively unique vertebrate species.

Results

Among a total of 112 K. marmoratus captured from Belize in 2005, 19 specimens (17%) were males and the remaining 91 (83%) were hermaphrodites, as judged by external phenotypic appearance. For this study, we assayed 101 of these specimens (11 of which were males). For these fish, plus 127 fish from several other locales (Table 1), diploid genotypes at all surveyed microsatellite loci are provided in Table 2, which is published as supporting information on the PNAS web site. Representative subsets of those data are presented in Fig. 1, which also introduces the dramatic differences between Belize and the other collection sites in overall population genetic architecture.

Population Genetic Variation. Table 1 provides a summary of microsatellite variation within the 12 populations of K. marmoratus surveyed to date. The range of intrapopulation genetic variation was high. For example, the proportion of polymorphic loci (frequency of the most common allele <99%) ranged from 0.31 (Exuma Island, Bahamas) to 0.94 (Twin Cays, Belize). Mean expected heterozygosities [under Hardy-Weinberg equilibrium (HWE)] and numbers of alleles per locus similarly spanned wide and roughly parallel ranges. Observed heterozygosities were typically <0.10 in most collections from Florida and the Bahamas and invariably were far below expected values for random mating populations. In sharp contrast, mean observed heterozygosities were ~0.48 in the Belize collections and were proportionately much closer to their mean HWE expected values of 0.65.

With regard to both the magnitude and the distribution of microsatellite variation among individuals, a drastic difference between the Belize and Florida samples is further illustrated in Fig. 2. For the collection from Charlotte County, FL, 13 of the 17 individuals assayed (76%) were homozygous at all 36 loci examined, and only one specimen displayed a level of intrain-
individual heterozygosity that might be expected (given the observed allele frequencies) for a random-mating deme (Fig. 2A).

By contrast, 29 of the 49 assayed specimens (59%) from Belize had observed heterozygosities >0.50, and the frequency distribution of heterozygosities in this subset of the population closely approximated values expected if the entire population had been random mating (Fig. 2B). Furthermore, according to a mixed mating model (using the selfing \((S)\) and outcrossing \((T)\) parameters empirically estimated for Belize, \(S = 0.424\) and \(T = 0.576\); see Selfing and Outcrossing Rates), the remaining Belize individuals formed additional modes in the heterozygosity histogram corresponding closely to what is anticipated if most such specimens were only one or two selfing generations removed from their most recent outcross events (Fig. 2C).

### Selfing and Outcrossing Rates

Inbreeding coefficients and deduced outcrossing rates are summarized in Fig. 3. Maximum coefficients of inbreeding [fixation indices \((F_{IS}) = 1; i.e., complete selfing] were found at the Everglades, Lostman’s River, and Shark River locales in Monroe County at the southwestern tip of Florida. Essentially the same can be said for both collections from the Bahamas (Exuma and San Salvador Islands), as well as from Charlotte County, FL (where \(F_{IS} = 0.92\)). Almost all individuals from these sites were homozygous across 36 loci, despite moderate to high levels of clonal diversity in each population (as evidenced, for example, by the expected heterozygosity values in Table 1). The other Florida populations generally exhibited somewhat lower inbreeding coefficients (0.65 < \(F_{IS} < 0.85\)) presumably evidencing modest outcrossing.

### Table 1. Summary of genetic variation within 12 *K. marmoratus* collections based on 36 microsatellite loci

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample size</th>
<th>Proportion of polymorphic loci (99% criterion)</th>
<th>Mean number of alleles per locus</th>
<th>Expected heterozygosity*</th>
<th>Observed heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exuma Island, Bahamas</td>
<td>12</td>
<td>0.31</td>
<td>1.3</td>
<td>0.12</td>
<td>0</td>
</tr>
<tr>
<td>Everglades National Park, FL</td>
<td>4</td>
<td>0.69</td>
<td>2.0</td>
<td>0.41</td>
<td>0</td>
</tr>
<tr>
<td>Shark River, FL</td>
<td>30</td>
<td>0.86</td>
<td>4.3</td>
<td>0.42</td>
<td>0</td>
</tr>
<tr>
<td>San Salvador Island, Bahamas</td>
<td>5</td>
<td>0.56</td>
<td>1.9</td>
<td>0.30</td>
<td>0.01</td>
</tr>
<tr>
<td>Lostman’s River, FL</td>
<td>6</td>
<td>0.53</td>
<td>2.2</td>
<td>0.33</td>
<td>0.01</td>
</tr>
<tr>
<td>Charlotte County, FL</td>
<td>17</td>
<td>0.61</td>
<td>2.0</td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td>Marco Island, FL</td>
<td>8</td>
<td>0.78</td>
<td>2.6</td>
<td>0.36</td>
<td>0.06</td>
</tr>
<tr>
<td>Long Key, FL</td>
<td>7</td>
<td>0.58</td>
<td>1.6</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>St. Lucie County, FL</td>
<td>12</td>
<td>0.86</td>
<td>3.5</td>
<td>0.47</td>
<td>0.08</td>
</tr>
<tr>
<td>No Name Key, FL</td>
<td>5</td>
<td>0.83</td>
<td>2.4</td>
<td>0.46</td>
<td>0.16</td>
</tr>
<tr>
<td>Twin Cays, Belize, 1991</td>
<td>21</td>
<td>0.94</td>
<td>6.5</td>
<td>0.63</td>
<td>0.45</td>
</tr>
<tr>
<td>Twin Cays, Belize, 2005</td>
<td>101</td>
<td>0.94</td>
<td>9.8</td>
<td>0.66</td>
<td>0.48</td>
</tr>
</tbody>
</table>

*Given the empirical allele frequencies and assuming HWE.

---

Fig. 1. Striking differences in patterns of variation at 36 microsatellite loci in three populations of *K. marmoratus*. In the interest of space, only 10 representative individuals from each population are shown here. Individuals are arranged in rows, and loci are in columns. Different colors correspond to different alleles. Hatched boxes indicate missing data. Boxes with a diagonal line and two colors are heterozygous specimens.
rates (0.1 < T < 0.2). By contrast, the populations from Twin Cays, Belize, displayed much lower inbreeding coefficients (FIS = 0.27 – 0.29) and therefore had much higher deduced rates of outcrossing (T = 0.55 – 0.58). Interestingly, the two Belize samples, collected 14 years apart, showed essentially identical population genetic architectures.

**Linkage Disequilibrium.** We also evaluated linkage disequilibrium (LD) in populations from which at least 17 individuals were sampled, with results summarized in Fig. 4. Each bar in the figure represents the percentage of variable microsatellite loci in significant LD with each locus specified. For example, in the Charlotte County population (Fig. 4B), locus R93 is in LD with a total of 71.4% (15 among 21) of the other polymorphic loci screened, whereas, at the Twin Cays site, it is in LD with only 6.1% (two of 33) of the other variable loci examined (Fig. 4A).

These statistics warrant elaboration. In the Twin Cays 2005 collection, for example, each locus was in LD with 0–5 (mean = 2.5, or 7.5%) among the 34 loci screened (again excluding invariant loci), but most of these face-value associations (28 of 42, or 66.7%) were significant only at the 5% level, and all but two became nonsignificant after Bonferroni correction for multiple tests. In sharp contrast, a variable locus in the Charlotte County population was in linkage disequilibrium with 0–14 (mean 8.3, or 39.5%) of 21 possible loci. Most of these face-value associations (72 of 91, or 79.1%) were significant at the 0.1% level (hatched bar), whereas <1 is expected to be so by chance (with 231 comparisons at this level of significance), and 58 associations remained significant after sequential Bonferroni correction. Finally, for the Shark River sample, the level of LD was intermediate to the Twin Cays and Charlotte County populations. Each variable locus was in LD with 0–14 others (mean 5.5 of 30 possible loci, or 18.3%). Eighty-five allelic associations were detected among 465 possible di-locus comparisons, and 9 of these associations remained significant after Bonferroni corrections (but many other associations must be genuine because their number exceeds that expected by chance).

**Discussion**

The mangrove killifish is the only hermaphroditic vertebrate known to reproduce by self-fertilization. However, previous molecular genetic surveys (multilocus DNA fingerprinting and microsatellite genotyping) have also unveiled occasional individuals displaying concentrations of multilocus heterozygosity reflective of recent outcross events, presumably mediated by males that typically are rare. Here we have shown that at a Belize site, where gonochoristic males are much more common, outcrossing rates appear to be far higher than previously suspected for this species. Such outcrossing in turn has had a dramatic impact on the genetic architecture of the Belize population when compared with other surveyed populations of this species from Florida and the Bahamas. These genetic footprints (all interrelated) of extensive outcrossing include higher within-individual heterozygosities, lower inbreeding coefficients, closer approximations of single-locus heterozygosities to expected values under random mating, and greatly diminished levels of linkage dis-
equilibria. The estimated outcrossing rate and the resulting population genetic architecture at the Twin Cays site in Belize also appear to have remained constant over a 14-year period.

It should be noted that estimates of $T$ calculated from $F_{IS}$ are based on assumptions that genotypic frequencies have reached equilibrium proportions expected under neutrality for a given level of selfing (23). Such assumptions are not always fulfilled in nature. Rates of selfing may vary from generation to generation, and natural selection may sometimes favor heterozygotes (thus biasing estimates of $T$ upwardly). On the other hand, mating between relatives (inbreeding) will increase $F_{IS}$, and accordingly decrease estimates of $T$. Similarly, significant population substructure can also cause an underestimation of $T$. Although estimates of outcrossing obtained from $F_{IS}$ are thus provisional, our qualitative conclusion about the magnitude of differences in outcrossing rates between Twin Cays, Belize, and Florida should remain secure.

Sources of Linkage Disequilibrium. Although the genetic markers we have used are presumably selectively neutral per se, the effect of inbreeding on multilocus associations can extend throughout the genome to include physically unlinked, as well as linked, loci, and neutral hitchhiking loci, as well as those that are under direct selection (24). We have shown that nonrandom allelic associations are rampant in $K$. marmoratus, especially in populations with high rates of selfing as opposed to outcrossing. Questions thus arise as to how many of these nonrandom associations involve unlinked genes and how many might be augmented by virtue of physical linkages of loci along particular chromosomes.

We have scrutinized our data for possible consistencies across multiple populations in the particular loci that displayed nonrandom allelic associations. The task was complicated by the fact that not all loci were variable in all populations, and that sample sizes at some locales were small. Nonetheless, by this approach, we did infer and then subsequently confirm two cases of actual “physical linkage.” Alleles at two pairs of loci (R9/R24 and R7/R45) showed significant associations in all populations where they were variable. By reexamining the original sequences of the microsatellite clones, we discovered that the primers we had developed for R9 and R24 cover exactly the same genomic region, with the forward primer for one “locus” being a reverse compliment of the reverse primer of the other and vice versa. In another similar case, locus R45 encompassed locus R7, such that alleles at the former were 66 bp longer than alleles at the latter.

We also addressed whether LD existed between particular microsatellite loci and the sex of the specimens in the Belize collection. No significant LD was detected, as would probably be expected given the environment-influenced nature of male formation in this species (see Androdioecy).

Surprisingly, we found strong LD in the Charlotte County population compared with the Shark River population, despite a higher outcrossing rate in the former. Even occasional outcrossing (e.g., $\approx$4.3% in the Charlotte County population) should lead eventually to a decay of LD between unlinked loci. It appears that the high level of LD in the Charlotte County population is caused by a presence of two distinct lineages (Fig. 1), which may indicate extensive population substructure, recent admixture, or a recent population bottleneck perhaps associated with small numbers of founders. Conversely, the relatively low level of LD in the Shark River population may indicate past outcrossing events (even though the heterozygote deficit in the available sample of 17 individuals resulted in a low estimate of outcrossing rate; $T = 0.2\%$).

Androdioecy. Mangrove killifish live in mangrove habitats characterized by challenging conditions such as high H2S concentrations, variable salinities, and rapidly changing water levels. Evolutionary theory predicts that one adaptive benefit of self-fertilization is the intact propagation and preservation of highly fit and environment-tested multilocus genotypes (25). Conversely, evolutionary theory also predicts that outcrossing should be of special adaptive significance in variable environments by promoting recombinational variation upon which natural selection might act with greatest efficacy. Either possibility or both might apply to $K$. marmoratus, perhaps depending on environmental circumstances, and our data provide no direct way to distinguish between these competing scenarios. Our data do indicate, however, that outcrossing rates and associated LD are highly variable across geographic populations of this species, and that much of this variation seems to be associated with differing frequencies of gonochoristic males, i.e., with the extent that a population exhibits androdioecy as opposed to strict hermaphroditism.

Androdioecy is exceedingly rare in the biotic world, and $K$. marmoratus is the only vertebrate known to display this reproductive mode. Variation in the frequency of males among

Fig. 4. Histograms depicting levels of linkage disequilibrium among microsatellite loci in three populations of $K$. marmoratus. Each bar represents the proportion of variable loci (among L-1 possible) found to be in association with the given locus at the significance levels of $p < 0.05$ (white), $p < 0.01$ (black), and $p < 0.001$ (hatched). Crosses indicate invariable loci.
geographical localities is probably due at least in part to ecological conditions. Gonochoristic males can readily be produced in the laboratory by altering the fishes’ rearing temperature and photoperiod regime (26–28), but the ecological variables in nature that favor male production remain uncertain. Recent “common-garden” experiments in aquaria have suggested that male development also has an appreciable genetic component that may operate in conjunction with environmental conditions (B.J.T., unpublished work).

Regardless of the underlying mechanisms promoting androdioecy (as opposed to strict hermaphroditism), our findings demonstrate that the extent of outcrossing, which is at least correlated with variation in the frequencies of gonochoristic males, can have a profound effect on the multifaceted genetic architectures of various populations of this peculiar vertebrate species.

Materials and Methods

Sample Collections. Most Belize samples were collected during May 2005 at Papa Gabriel (eastern coast of the largest island in the Twin Cays) at 17°N × 88°W, or ~12 km off the Dangriga coast (22, 29). Fish were captured from temporary pools or the burrows of land crabs (Ucides cordatus) by using cup traps, wire minnow traps, hook and line, or dip nets. The gender of each specimen (hermaphrodite versus male) was assessed by using phenotypic characters, especially body coloration and presence/absence of a black ocellus on the caudal fin. Additional samples came from frozen stocks maintained by B.J.T. of either purified DNA or fish tissue representing previous collections from the following sites: (i) a 1991 collection from Twin Cays, Belize; (ii) Exuma and San Salvador Islands, Bahamas; (iii) Everglades National Park near Flamingo in Monroe County, FL; (iv) Shark River and Lostman’s River also in Monroe County, FL; (v) Charlotte County, FL; (vi) St. Lucie County, FL; (vii) Marco Island in Collier County, FL; and (viii) No Name Key and Long Key in the Florida Keys. Data for the Belize and Bahamas samples are new for this study, whereas those for most of the Florida population have been published (16). All sample sizes are given in Table 1.

Microsatellite Genotyping. High molecular weight genomic DNA was extracted and purified by using standard phenol/chloroform protocols. Primers and PCR conditions for the amplification of microsatellite loci will be described elsewhere (M.M, A.T., A. Perry, J. R. Martin, J. F. Elder, Jr., D. L. Bechler, and J.C.A., unpublished work). Fluorescently labeled PCR products were electrophoresed and detected on an ABI 377 DNA sequencer (Applied Biosystems). Alleles were sized by using software packages GENESCAN V.3.1.2 and GENOTYPER V.2.5 (Applied Biosystems).

Analysis of Genetic Diversity. Population genetic parameters (e.g., proportion of polymorphic loci, mean number of alleles per locus, and fixation indices) were calculated by using GENETIC DATA ANALYSIS (GDA) software (30). Confidence intervals for inbreeding coefficients (Fis) were estimated by bootstrapping (10,000 resamplings of loci). Rates of selfing (S) and outcrossing (T = 1 − S) were estimated from empirical fixation indices by using the relation Fis = S/(2 − S) (23). Departures from HWE at each locus and LD between pairs of loci were assessed by using exact tests implemented in GDA. In the LD analyses, observed two-locus genotypes were compared with expectations of free recombination, which was achieved by randomly drawing genotypes from each locus (the significance of each resulting test, which involved 1,000 permutations, is not influenced by single-locus deviations from HWE).

Individual heterozygosities were calculated by using microsatellite analyzer (31). The theoretical distribution of individual heterozygosities under random mating was calculated as a binomial distribution based on the expected heterozygosities (under HWE) calculated from the empirical data across all surveyed loci. The theoretical distribution under mixed mating was obtained by recursively applying the outcrossing rate T to the random mating distribution of individual heterozygosities, so that, in a sample of N specimens, TN individuals have been derived from random mating and SN have been produced by one or more successive generations of self-fertilization (with a halving of heterozygosity in each generation). In the later group, TSN are first-generation products of selfing, TSN are second-generation products, T3N are third-generation products, and so on. The recursive procedure can be repeated as many times (r) as necessary for TN(1 + S + S2 + · · · + Sr−1 + S) to approach N. For the plot shown in Fig. 2C, we used r = 4.

We thank William Davis, William Dunson, John Grizzle, and Carole McVor for their sample collections. Field work on Twin Cays was made possible by assistance and hospitality of the Smithsonian Tropical Research Institute Laboratory on Carrie Bow Cay, Belize. This work was supported by funds from the University of Georgia and the University of California at Irvine (to J.C.A.), 18. Taylor, D. S., Fisher, M. T. & Turner, B. J. (2001) Environ. Biol. Fishes 61, 455–459.


