Journal of FISH BIOLOGY

fsbi An International Society for Fish Biology

Journal of Fish Biology (2014) **85**, 1766–1776
doi:10.1111/jfb.12506, available online at wileyonlinelibrary.com

Developing major histocompatibility markers in a species of concern: the Sacramento perch *Archoplites interruptus*

P. Lavretsky*†‡, R. S. Schwartz*§, M. R. Baerwald* and B. May*

*Department of Animal Science, University of California–Davis, Meyer Hall, OneShields Avenue, Davis, CA 95616, U.S.A., †Department of Environmental Sciences, Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH 45435, U.S.A. and §Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Arizona State University, Tempe, AZ 85287, U.S.A.

(Received 12 November 2013, Accepted 14 July 2014)

Primers targeting two non-neutral major histocompatibility complex (*mhc*) II β genes were developed and assayed across several disjoint Sacramento perch *Archoplites interruptus* sampling locations. Variability at the two *mhc* loci among sampling stocks strongly correlated to previous estimates with neutral markers, suggesting that the effect of genetic drift was not limited to neutrally evolving regions of the genome. The novel *mhc* primers will help develop admixture schemes in *A. interruptus* captive breeding programmes and will increase the success of future reintroductions of this species of concern.

Key words: captive breeding programme; Centrarchidae; conservation; diversity; management; teleost.

Once endemic to the Sacramento and San Joaquin catchment in California, U.S.A., Sacramento perch *Archoplites interruptus* (Girard 1854) were extirpated from much of their native habitat by the early part of the 20th century due to major environmental disruptions (*e.g.* river channelization, dams and introduction of non-native species; Moyle, 2002). An effort to reestablish this species across its range has suggested the need for a captive breeding programme (Crain & Moyle, 2011). During the decline of *A. interruptus*, several transplanted populations were established (Fig. 1) and are now being considered as potential sources for the captive breeding programme (Moyle, 2002). Unfortunately, all the established populations were derived from one another, and were founded with <50 individuals. Genetic bottlenecks were identified in the majority of populations with microsatellite (Schwartz & May, 2008) and mitochondrial markers (Crain *et al.*, 2011). Due to the lack of genetic diversity, Schwartz & May (2008) suggested drawing from multiple source populations in reintroductions to increase overall genetic diversity and avoid possible problems with inbreeding

‡Author to whom correspondence should be addressed. Tel.: +1 310 770 1758; email: lavretsky.2@wright.edu

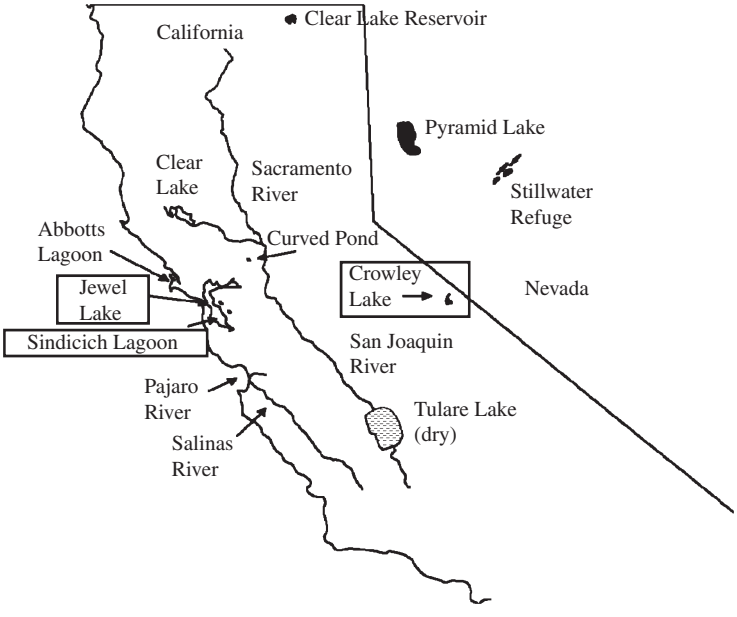

FIG. 1. *Archoplites interruptus* range and translocation sites (modified from Schwartz & May, 2008). Sampling sites are boxed.

depression. They recommended striking a balance between increasing genetic diversity to ensure the potential for the ability to respond to future environmental change, while avoiding breaking up co-adapted gene complexes and loss of local adaptation by breeding all homogenic populations with individuals only from the populations with highest genetic diversity. Microsatellites, being putatively neutral, however, may not directly correspond with a population's adaptive diversity. For instance, although Mikko & Anderson (1995) and Campos *et al.* (2006) found corresponding diversity between neutral makers and those under selection (for brown trout *Salmo trutta* L. 1758 and moose *Alces alces*, respectively), Hansson & Richardson (2005; *Acrocephalus* warblers) as well as van Oosterhout *et al.* (2006; Trinidadian guppy *Poecilia reticulata* Peters 1859) reported contradicting levels of diversity at neutral and non-neutral markers. Thus, a comprehensive assessment of the genetic diversity of a species of concern must include both neutral and non-neutral markers.

As the cornerstone of an individual's adaptive immune system (Klein, 1986), the major histocompatibility complex (*mhc*) is often a focal non-neutral marker in population assessments (Sommer, 2005). MHC genes exhibit a high level of polymorphism (Brown *et al.*, 1993) that correlates with the ability of the immune system to recognize a wide range of antigens (Jardetzky, 1996). Consequently, elevated *mhc* diversity permits populations to resist extirpation from novel diseases or stochastic environmental events (Apanius *et al.*, 1997; Beacham *et al.*, 2004; Miller *et al.*, 2004; Neff *et al.*, 2008). Because *mhc* is under strong balancing selection (Hughes & Nei, 1992), it is believed to counteract demographic limitations and the effects of drift in small populations (Takahata *et al.*, 1992). Thus, greater diversity is predicted to exist in *mhc* genes as compared to neutral markers (van Oosterhout *et al.*, 2006) and lower levels

of differentiation (*i.e.* F_{ST}) between isolated populations (Muirhead, 2001; O'Farrell *et al.*, 2012). Objectives of this study were to isolate and compare *mhcII-β* locus diversity across four *A. interruptus* wild populations and a captive stock, as well as to compare *mhc* variation to previous neutral divergence estimates (Schwartz & May, 2008) to determine the extent to which a genetic bottleneck affected overall diversity.

A total of 52 fin tissue samples were collected from four locations (upper and lower Sindicich Lagoons, Crowley Lake and Jewel Lake; Fig. 1 and Table SI, Supporting Information) chosen based on differentiating levels of bottlenecking established with microsatellite (Schwartz & May, 2008). Fin tissue was used as it did not kill or harm fish and would permit the use of those individuals in admixture schemes. An additional 10 samples comprised second generation (F2) offspring from a preliminary captive breeding stock that was originally created by crossing individuals from Abbotts and Sindicich Lagoons. Unfortunately, Abbotts Lagoon was not sampled; however, increased heterozygosity in F2 individuals (as compared to Sindicich individuals) would demonstrate that the developed markers are suitable to establish admixture regiments.

A single degenerate primer pair was first designed based on conserved regions from an alignment of other teleost species' *mhc* class II *β* exon 2 protein sequences obtained from GenBank (Table SII, Supporting Information). After amplification failed with genomic DNA, RNA was extracted from fin tissue using the RNeasy Extraction kit (Qiagen; www.qiagen.com) and converted to cDNA with the iScript cDNA Synthesis kit (Bio-Rad, Inc.; www.bio-rad.com), using an Oligo dT-primed cDNA reaction for four Crowley Lake adults. Polymerase chain reaction (PCR) amplification was performed using 1 μl template complementary (c)DNA 1× Green GoTaq Flexi Buffer (Promega; www.promega.com), 2 mM $MgCl_2$, 0·8 μM deoxynucleotide triphosphate (dNTP) (Promega), 0·6 μM of each primer and 0·04 U *Taq* DNA polymerase (Promega), in a total volume of 10 μl. PCR was carried out in a Bio-Rad PTC-100 thermocycler under the following conditions: 94° C for 2 min, 30 cycles at 94° C for 30 s, 50° C for 60 s, 72° C for 2 min, followed by 72° C for 10 min, and held at 4° C. Amplification was verified by the presence of a predicted *c.* 550 bp band using 1% agarose gel. PCR products were purified using a QIAquick Gel Extraction kit (Qiagen). PCR products were then cloned using a pGem-T Easy kit (Promega), and subsequently purified and extracted with the Qiagen miniprep cleanup kit.

Initial screening across a total of 80 clones ($n = 20$ clones per individual) yielded two distinct and divergent *mhc* II sequences that were present in each batch of 20 clones per individual. Sequence-specific primers were then designed based on discrete base pair changes between the two divergent sequences to determine if these sequences represented a single locus or multiple loci prior to assigning zygosity. The two forward primers (*mhcII.1 F* and *mhcII.2 F*) were each paired with primer *mhcIIBTel R*, and two reverse primers (*mhcII.1 R* and *mhcII.2 R*) were each paired with primer *mhcIIBTel F* (Table I). As expected, the intron-spanning primers were unable to amplify genomic DNA; thus, only cDNA was used for all sequence analysis. PCR was carried out with the above primer pairs on all 52 individuals with PCR conditions identical to those listed for the initial *mhc* II degenerate amplification. PCR products were purified using AMPure XP beads following Agencourt protocol (Beckman Coulter Co.; www.beckmancoulter.com) and given to University of California Davis College of Biological Science sequencing facility for automated Sanger sequencing. PCR products were sequenced from both directions to ensure sequencing accuracy. PCR and sequencing reactions were duplicated at least twice for half of the individuals to further ensure

TABLE I. Primer sequences deigned to amplify two major histocompatibility complex II β genes

Primer	Sequence (5′-3′)
mhcIIBTel F	TGTGTGTTCAACTCCWCTRA
mhcIIBTel R	CTTGTAGTAGATGARTACCWG
mhcII.1 F	GGTCTTATTATTACAACAATGA
mhcII.2 F	GGTCTGATTATTACAACAAGAA
mhcII.1 R	GGTAGTACCAGTCCCCGTCTG
mhcII.2 R	GGTAGTACCAGTCCCCGTCTT

sequence certainty. Sequences were imported and aligned using Sequencher version 4.7 (Gene Codes, Inc.; www.genecodes.com). From these sequences, locus-specific primer pairs were *mhcII.1 F*/*mhcIIBTel R* and *mhcII.2 F*/*mhcIIBTel R*, while allele-specific primers were *mhcIIBTel F*/*mhcII.1 R* and *mhcIIBTel F*/*mhcII.2 R*.

Locus-specific primers amplified a *c*. 551 bp region of the *mhcII* gene without other, for example, amplicons and pseudogenes, using cDNA. Where allele-specific priming did not work, alleles were resolved for each individual per locus using the programme PHASE (Stephens *et al.*, 2001) which derives the most likely state of each allele algorithmically by comparing all known alleles. An open reading frame was observed for all resolved alleles. Final sequences included the last 62 amino acids of exon 2, spanned through exon 3 (71 amino acids) and exon 4 (23 amino acids), as well as for the first 29 amino acids of exon 5 [Fig. 2; amino acid positions inferred from Ono *et al.* (1993)].

The total number of alleles across primer pairs ranged from two to four per individual (Table SIII, Supporting Information). Discovery of novel alleles in the remaining individuals demonstrated that the developed primers successfully targeted many alleles that were not present in the initial cloning process. While these results suggest that two *mhcII* β loci were successfully isolated, and similar to that found in other teleost MHC studies (McConnell *et al.*, 1998), the initial cloning experiments may have limited the ability to detect other *mhcII* β loci. Phylogenetic relationships were reconstructed in MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) and viewed in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/). The data were separated by codon position and evaluated using a general time reversible (GTR) model and gamma-distributed rates across sites. The analysis was run for 2 million generations, with sampling every 200 generations until the S.D. between sampling events was <0·01. The first 25% of the samples were discarded as burn-in. In general, phylogenetic results identified *A. interruptus* as a monophyletic group (Fig. 3) and delineating the *A. interruptus* loci as more similar to one another than to other teleost *mhc* loci. *Archoplites interruptus mhcII-*β genes appear to have diverged sometime after the formation of the Perciform order and the rise of the Centrarchidae family (between 65 and 33·4 million years ago; Near *et al.*, 2005). Such divergence patterns have been reported previously, suggesting that most species' *mhc* loci diverged at the class or family level (Harstad *et al.*, 2008) and are probably paralogues (Reusch *et al.*, 2004). Although two loci were isolated, phylogenetic analyses failed to identify any orthologous relationships among putative class II loci and as a result assigning these loci to any previous nomenclature is inappropriate and they will continue to be referred to as *mhcII.1* and *mhcII.2*.

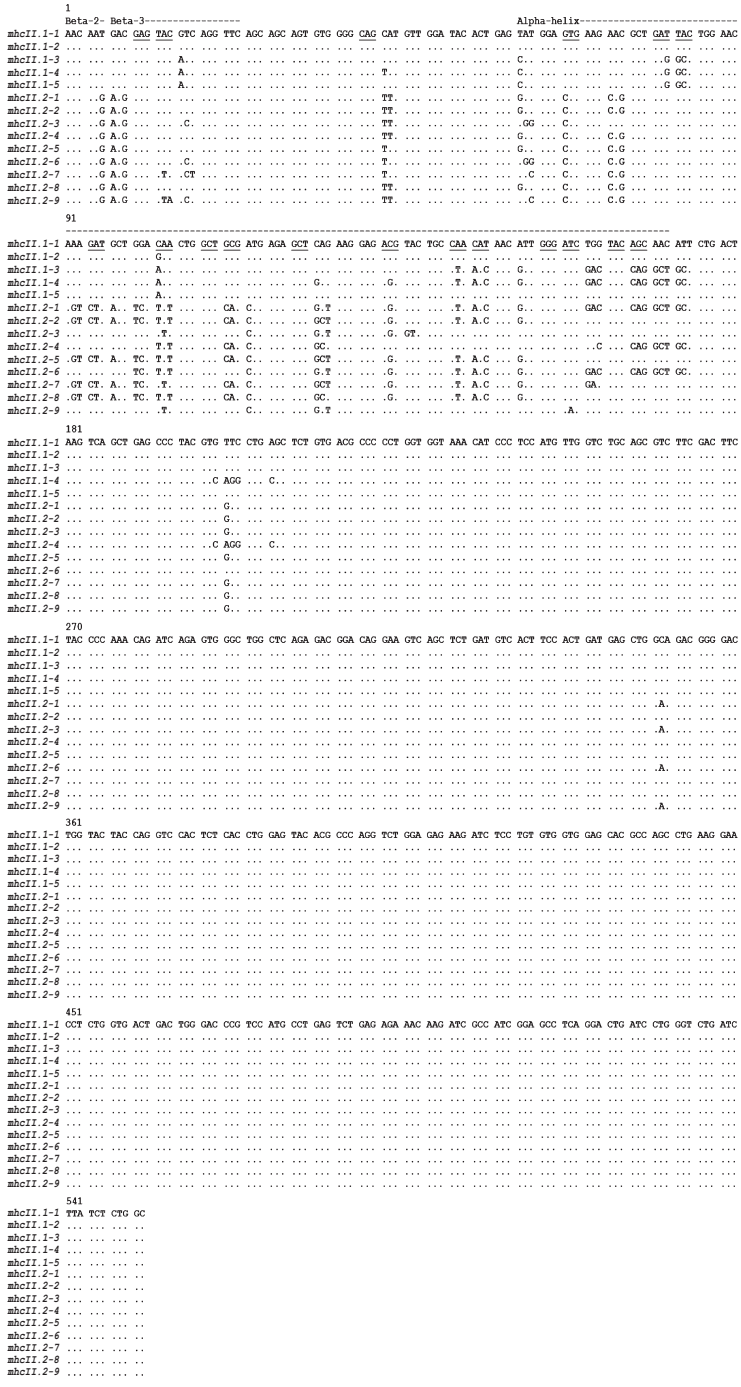

FIG. 2. Alignment of five *mhcII.1* and nine *mhcII.2* allelic sequences separated by codons. … , indicate identity with the top sequence. The G-domain β-pleated sheets (β-2 and β-3) and the α helix are indicated by name and extension of the dashes. Codons involved in protein-binding regions (PBRs) inferred from Cohen (2002) are underlined.

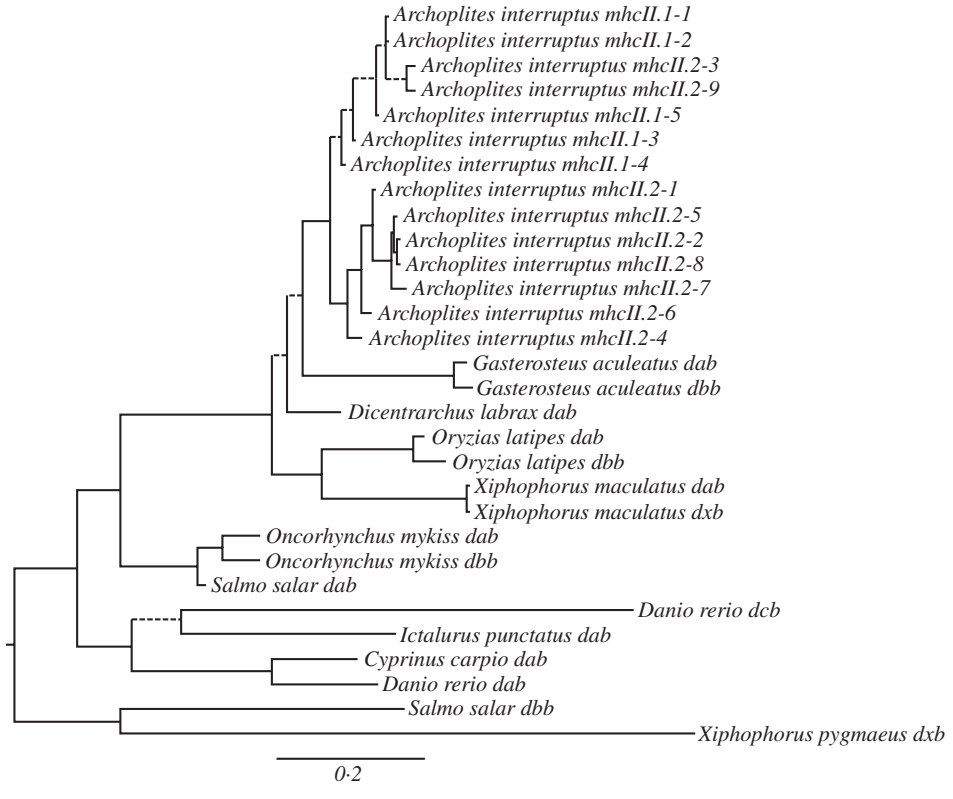

FIG. 3. Phylogeny of major histocompatibility complex (MHC) samples from *Archoplites interruptus*. The phylogeny was constructed in a Bayesian framework. All nodes with posterior support <90% are noted with dashed branches. Numbers adjacent to *A. interruptus* correspond to alleles (Table SII, Supporting Information).

Deviations from neutrality among alleles were based on per codon Tajima's *D* estimated using a sliding window approach implemented in DnaSP 5 (Librado & Rozas, 2009) (Fig. 4). Additionally, the two loci were independently analysed for the presence of positive selection (ω) based on relative frequencies of non-synonymous (dN) and synonymous (dS) base substitutions (*i.e.* dN:dS > 1) obtained from omegaMap 0.5 (Wilson & McVean, 2006) (Fig. 4). Positive Tajima's *D* estimates were obtained for *mhcII.1* codon positions 6, 21, 27–28, 48–49, 51, 54 and 56–58 and negative estimates for codon positions 15, 35, 42, 45, 67–68 and 70. Positive Tajima's *D* estimates were obtained for *mhcII.2* codon positions 6, 15, 21, 31–35, 38, 42, 45, 48–49, 51, 54, 56–58 and 117 and negative estimates for codon positions 5, 25, 46, 53, 67–68 and 70. Significant posterior probabilities for positive selection {[p (dN:dS) > 1] > 0·95}were obtained for *mhcII.1* codon positions 6, 21, 27–28 and 35 and for *mhcII.2* codon positions 6, 15, 21, 25, 31–35, 38, 42, 48–49, 51, 54, 56–58 and 117 (Fig. 4). Codons 27–28, 32, 35, 38, 45, 48–49, 53 and 56 are part of the *mhcII* protein-binding region (PBR) (Cohen, 2002) (Fig. 2) and were found to deviate from neutrality. Although many of the codons that showed similar patterns were not identified as part of the PBR, all but codon 117 (for *mhcII.2* only) were located immediately adjacent to the putative

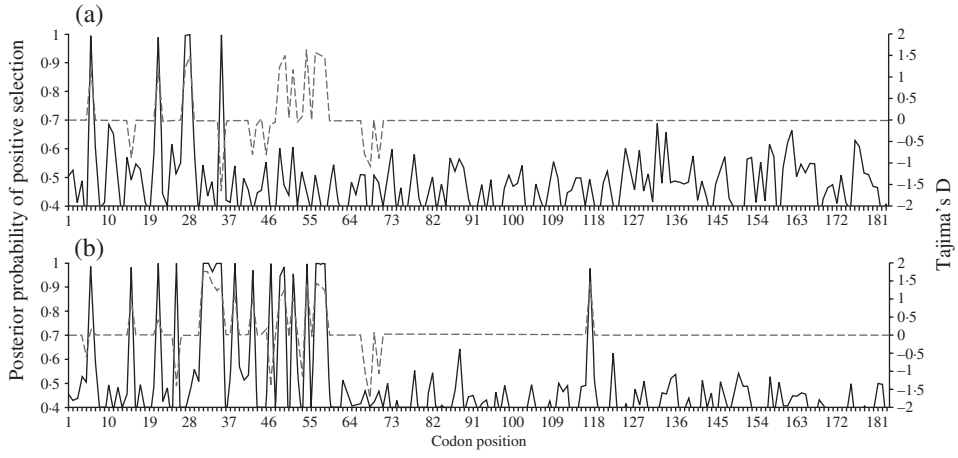

FIG. 4. Overlap of posterior probability of positive selection [=P(dN:dS > 1); ___] and Tajima's D (Tajima, 1983; _ _) across codons for (a) *mhcII.1* and (b) *mhcII.2*. Presence of positive selection was independently analysed for alleles of both loci in omegaMap 0.5 (Wilson & McVean, 2006). Without prior knowledge, all codons were assumed to have an equal equilibrium frequency and all priors were set as recommended by Wilson & McVean (2006), with the exception of ω and ρ models that were allowed to independently vary across codons. Analyses were run twice for 2 million Markov chain Monte-Carlo iterations with a thinning interval of 200. Burn-in dependent on maximizing the effective sample size for all parameters across paired runs and were examined using the *trace.omegaMap* command provided with the omegaMap package in R (www.r-project.org). Tajima's D was analysed with a sliding window approach in DnaSP 5 (Librado & Rozas, 2009). Positions departing from neutrality were based on the presence of a positive (indicating a decrease in population size and balancing selection) or negative (population size expansion and purifying selection) Tajima's D.

PBRs. Elevated levels of selection in codons within verse outside the PBR are typical of *mhc* molecules (Sommer, 2005).

Pair-wise estimates of genetic differentiation (F_{ST}) between each sampling location was calculated in FSTAT 2.9.3.2 (Goudet, 1995; Goudet, 2001) with $F_{ST} > 0{\cdot}10$ considered as indicative of differentiation (Table II). Specifically, Crowley, upper Sindicich and lower Sindicich locations were indistinguishable, while Jewel Lake samples were differentiated at the *mhcII.1* locus. Conversely, with the exception of the two Sindicich populations, all others were significantly differentiated at the *mhcII.2* locus. The overall pattern of differentiation across *mhc* loci significantly correlated ($r^2 = 0{\cdot}83$) to previous F_{ST} estimates obtained for each of the sampling locations with neutral microsatellite markers (Schwartz & May, 2008) (Table II). The correlation between the two marker types suggests that demographic pressures [*e.g.* bottlenecks due to small founder populations (Moyle, 2002)] adversely affected neutral and non-neutral diversity in *A. interruptus* and contradicts previous studies showing that balancing selection on *mhc* counteracts the effects of drift (Takahata *et al.*, 1992).

Establishing the number of genetically unique clusters based on *mhc* variability was evaluated using Structure (Pritchard *et al.*, 2000). Simulations were run 10 times for each K cluster, ranging from $K = 1-5$, with 500 000 Markov chain Monte-Carlo (MCMC) steps following a burn-in of 100 000 steps. The optimum K was determined by calculating ΔK in the programme Structure Harvester (Earl & VonHoldt, 2012) (Fig. 5). Final assignment probabilities were based on the optimal clustering

TABLE II. F_{ST} estimates between pairs of sampling locations ($P < 0{\cdot}05$ for all comparisons) for the *mhcII.1* locus, *mhcII.2* locus and microsatellites (obtained from Schwartz & May, 2008)

	mhcII.1	*mhcII.2*	Microsatellite
Crowley–upper Sindicich	0·00	0·13	0·05
Crowley–lower Sindicich	0·00	0·4	0·05
Crowley–Jewel Lake	0·77	0·74	0·34
Upper Sindicich–lower Sindicich	0·00	0·06	0·00
Upper Sindicich–Jewel Lake	0·71	0·49	0·32
Lower Sindicich–Jewel Lake	0·71	0·38	0·32

alignment across all 10 replicates of the optimum *K* using a FullSearch algorithm as implemented in the programme CLUMPP (Jakobsson & Rosenberg, 2007). The optimum population model was two (Fig. 4) and closely corresponded to population structure established with microsatellites for the four sampling locations (Schwartz & May, 2008). Specifically, all Sindicich and Crowley individuals comprised one cluster, while the other was composed of Jewel Lake individuals. Genetic differentiation among the stocking locations did not correspond to geographic locality. Although Jewel Lake is geographically closer to Sindicich Lagoon (Fig. 1), Sindicich and Crowley Lake individuals comprised a single genetic cluster. Such results demonstrate the importance of quantifying and understanding the number of genetically distinct clusters present without the assumption that these correspond to geographic relationships.

The developed markers successfully captured *mhc* variation in *A. interruptus* and delineated two genetically distinct groups among the four wild sampling locations.

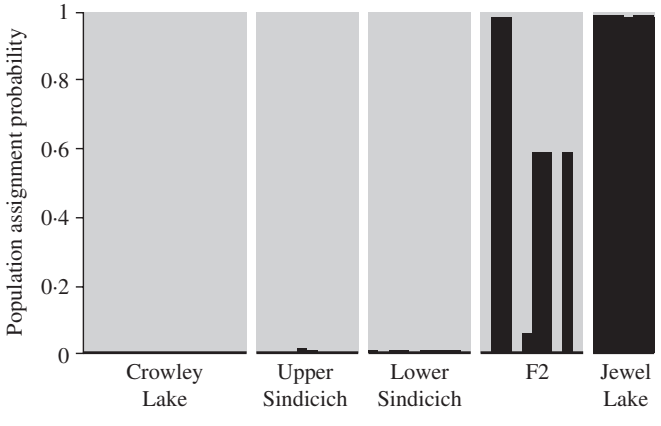

FIG. 5. Bayesian population assignment probabilities per *Archoplites interruptus* individual based on *mhcII.1* and *mhcII.2* variability as implemented in the programme Structure (Pritchard *et al.*, 2000). After analysing population models (*K*) of 1–5, an optimum *K* of two was determined by calculating ΔK in the programme Structure Harvester (Earl & VonHoldt, 2012). Final assignments were based on the optimal clustering alignment across all 10 replicates using a FullSearch algorithm as implemented in the programme CLUMPP (Jakobsson & Rosenberg, 2007), with assignment probabilities to either population represented by ■ and ■.

Moreover, Structure assigned 50% of captive-bred individuals (F2) to cluster one and 20% to cluster two, with 30% being of admixed ancestry. These results demonstrate that although admixture is not complete, a single generation of breeding has already increased overall diversity within the stock as compared to the Sindicich parental. Nevertheless, increasing sample sizes within sampling locations as well as from other locations will be needed for a more comprehensive population-level study.

In conclusion, the diversity of microsatellite and *mhc* markers in *A. interruptus* was strongly correlated, suggesting that genetic drift affects *mhc* diversity despite the potential for balancing selection at these loci. Microsatellite markers may serve as a reasonable estimate of diversity for admixture and captive breeding attempts, although usage of multiple neutral and non-neutral markers that provide a genome prospective in sampled individuals is advised. Thus, these novel *mhc* primers will be instrumental in the reintroduction and conservation of the *A. interruptus*, as well as for other closely related taxa.

We thank the following for help in securing samples: C. Woodley, P. Alexander, L. Ochikubo-Chan and C. Miller. This project was funded by a CALFED grant (Award # ERP02P34) to P. Moyle, J. Cech and B.M. Thanks to J. Peters for insight, discussion and comments on previous drafts. We thank the anonymous reviewers for their insightful comments on previous drafts.

Supporting Information

Supporting Information may be found in the online version of this paper:

Table SI. Sample size and location of *Archoplites interruptus* populations used in this study.

Table SII. GenBank accession numbers for all species used for primer design and included in the phylogenetic tree.

Table SIII. *mhcII.1* and *mhcII.2* allelic frequencies and genotypes from four sampling locations and a captive-bred stock (F2).

References

Apanius, V., Penn, D., Slev, P. R., Ruff, L. R. & Potts, W. K. (1997). The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* **17,** 79–224.

Beacham, T. D., Lapointe, M., Candy, J. R., Miller, K. M. & Withler, R. E. (2004). DNA in action: rapid application of DNA variation to sockeye salmon fisheries management. *Conservation Genetics* **5,** 411–416.

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364,** 33–39.

Campos, J. L., Posada, D. & Morán, P. (2006). Genetic variation at MHC, mitochondrial and microsatellite loci in isolated populations of Brown trout (*Salmo trutta*). *Conservation Genetics* **7,** 515–530.

Cohen, S. (2002). Strong positive selection and habitat-specific amino acid substitution patterns in MHC from an estuarine fish under intense pollution stress. *Molecular Biology and Evolution* **19,** 1870–1880.

Crain, P. K. & Moyle, P. B. (2011). Biology, history, status and conservation of Sacramento perch, *Archoplites interruptus*. *San Francisco Estuary and Watershed Science* **9,** jmie_sfews_11025.

Crain, P., Moyle, P. B., Woodley, C., Cech, J., Schwartz, R. & May, B. (2011). *Restoration of the Sacramento Perch to the San Francisco Estuary*. Davis, CA: University of California, Davis.

Earl, D. A. & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4,** 359–361.

Goudet, J. (1995). FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* **86,** 485–486.

Hansson, B. & Richardson, D. S. (2005). Genetic variation in two endangered *Acrocephalus* species compared to a widespread congener: estimates based on functional and random loci. *Animal Conservation* **8,** 83–90.

Harstad, H., Lukacs, M. F., Bakke, H. G. & Grimholt, U. (2008). Multiple expressed MHC class II loci in salmonids; details of one non-classical region in Atlantic salmon (*Salmo salar*). *BMC Genomics* **9,** 193.

Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17,** 754–755.

Hughes, A. L. & Nei, M. (1992). Evolution at the molecular level. *Journal of Molecular Evolution* **35,** 90–92.

Jakobsson, M. & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23,** 1801–1806.

Jardetzky, T. (1996). *Crystal Structures of MHC Class I and Class II Molecules*. Oxford: BIOS Scientific Publishers.

Klein, J. (1986). *Natural History of the Major Histocompatibility Complex*. New York, NY: Wiley.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–1452.

McConnell, T. J., Godwin, U. B., Norton, S. F., Nairn, R. S., Kazianis, S. & Morizot, D. C. (1998). Identification and mapping of two divergent, unlinked major histocompatibility complex class II B genes in *Xiphophorus* fishes. *Genetics* **149,** 1921–1934.

Mikko, S. & Anderson, L. (1995). Low major histocompatibility complex class II diversity in European and North American moose. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 4259–4263.

Miller, K. M., Winton, J. R., Schulze, A. D., Purcell, M. K. & Ming, T. J. (2004). Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. *Environmental Biology of Fishes* **69,** 307–316.

Moyle, P. B. (2002). *Inland Fishes of California*. Berkeley, CA: University of California Press.

Muirhead, C. A. (2001). Consequences of population structure on genes under balancing selection. *Evolution* **55,** 1532–1541.

Near, T. J., Bolnick, D. I. & Wainwright, P. C. (2005). Fossil calibrations and molecular divergence time estimates in Centrarchid fishes (Teleostei: *Centrarchidae*). *Evolution* **59,** 1768–1782.

Neff, B. D., Garner, S. R., Heath, J. W. & Heath, D. D. (2008). The MHC and non-random mating in a captive population of Chinook salmon. *Heredity* **101,** 175–185.

O'Farrell, B., Dennis, C., Benzie, J. A., McGinnity, P., Carlsson, J., de Eyto, E., Coughlan, J. P., Igoe, F., Meehan, R. & Cross, T. F. (2012). Balancing selection on MHC class I in wild brown trout *Salmo trutta*. *Journal of Fish Biology* **81,** 1357–1374.

Ono, H., O'HUigin, C., Vincek, V. & Klein, J. (1993). Exon-intron organization of fish major histocompatibility complex class II B genes. *Immunogenetics* **38,** 223–234.

van Oosterhout, C., Joyce, D. A., Cummings, S. M., Blais, J., Barson, N. J., Ramnarine, I. W., Mohammed, R. S., Persad, N. & Cable, J. (2006). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution* **60,** 2562–2574.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959.

Reusch, T. B. H., Schaschl, H. & Wegner, K. M. (2004). Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics* **56,** 427–437.

Ronquist, F. & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574.
Schwartz, R. S. & May, B. (2008). Genetic evaluation of isolated populations for use in reintroductions reveals significant genetic bottlenecks in potential stocks of Sacramento perch. *Transactions of the American Fisheries Society* **137,** 1764–1777.
Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* **2,** 16.
Stephens, M., Smith, N. & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68,** 978–989.
Takahata, N., Satta, Y. & Klein, J. (1992). Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* **130,** 925–938.
Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **123,** 437–460.
Wilson, D. J. & McVean, G. (2006). Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172,** 1411–1425.

Electronic Reference

Goudet, J. (2001). *FSTAT*, *a Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2.9.3).* Available at http://www.unil.ch/izea/softwares/fstat.html/