MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2014) doi: 10.1111/1755-0998.12308

Discovery and validation of gene-linked diagnostic SNP markers for assessing hybridization between Largemouth bass (*Micropterus salmoides*) and Florida bass (M. *floridanus*)

CHAO LI,*[1] SPENCER GOWAN,*[1] AMMU ANIL,* BENJAMIN H. BECK,† WILAWAN THONGDA,* HUSEYIN KUCUKTAS,* LUDMILLA KALTENBOECK* and ERIC PEATMAN*

**School of Fisheries, Aquaculture and Aquatic Sciences, Auburn University, Auburn, AL 36849, USA, †United States Department of Agriculture, Agricultural Research Service, Stuttgart National Aquaculture Research Center, Stuttgart, AR 72160, USA*

Abstract

Efforts to improve recreational fisheries have included widespread stocking of *Micropterus floridanus* outside its native range of peninsular Florida. Hybridization of Florida bass (*M. floridanus*) with largemouth bass (*Micropterus salmoides*) has now dramatically expanded beyond a naturally occurring intergrade zone in the southeast U.S. In recent years, there has been growing interest in protecting the genetic integrity of native basses and assessing the impact and nature of *M. salmoides*/*M. floridanus* introgression from the standpoint of hatchery and sport-fishery managers, fish biologists, ecologists and evolutionary biologists. Here, we conducted RNA-seq-based sequencing of the transcriptomes of *M. salmoides*, *M. floridanus* and their F1 hybrid and identified a set of 3674 SNP markers with fixed-allelic differences from 2112 unique genes. We then developed a subset of 25 of these markers into a single diagnostic multiplex assay and validated its capacity for assessing integrity and hybridization in hatchery and wild populations of largemouth and Florida bass. The availability of this resource, high-quality transcriptomes and a large set of gene-linked SNPs, should greatly facilitate functional and population genomics studies in these key species and allow the identification of traits and processes under selection during introgressive hybridization.

Keywords: bass, hybridization, introgression, *Micropterus*, single nucleotide polymorphism

Received 30 May 2014; revision received 17 July 2014; accepted 18 July 2014

Introduction

Black basses (*Micropterus*) are ecologically and economically important members of a diverse array of ecosystems across North America, from small ponds and streams to large rivers and lakes (DeVries *et al.* 2014). They are among the most popular freshwater sport fishes in North America and have been widely translocated as part of large-scale stocking efforts (Barthel *et al.* 2010). These stocking efforts have been particularly intense for the largemouth bass, now believed to be represented by two distinct species (Kassler *et al.* 2002; Near *et al.* 2003), the Florida bass (*M. floridanus*) and the (Northern) largemouth bass (*M. salmoides*). Florida bass, endemic to peninsular Florida, have been extensively stocked outside their native range, particularly in the Southern United States, due to superior growth characteristics (Addison & Spencer 1971; Chew 1975) and currently exist in introgressed populations with the more widely occurring largemouth bass (Northeastern Mexico; Southeastern Canada; Midwestern, Southeastern and Eastern US; MacCrimmon & Robbins 1975). Secondary contact between the two taxa during the Pleistocene has resulted in a natural hybrid zone (Nedbal & Philipp 1994; Near *et al.* 2003) in the southeast US although the scope and extent of introgression has been dramatically expanded by stocking (Philipp *et al.* 1983; Johnson & Fulton 1999; Barthel *et al.* 2010). The rate and success of this hybridization appears to be the result of strikingly low selection (in the form of intrinsic genetic incompatibilities) against hybrids within Centrarchidae when compared with other taxonomic groups (Bolnick & Near 2005; Seyoum *et al.* 2013).

The impact of anthropogenically driven hybridization between *M. salmoides* and *M. floridanus* has been the subject of intense debate for over 30 years, with some seeing deleterious breakdown of co-adapted gene complexes and detrimental physiological consequences through outbreeding depression (e.g. Philipp *et al.* 1983; Philipp & Claussen 1995; Cooke & Philipp 2006) and others

Correspondence: Eric Peatman, Fax: +1-334-844-4694; E-mail: peatmer@auburn.edu

[1]These two authors contributed equally.

seeing potential enhancement of managed fisheries, particularly in large reservoirs (e.g. Maceina *et al.* 1988; Maceina & Murphy 1992). Interest in the genetic integrity of native basses and the impact and nature of *M. salmoides*/*M. floridanus* introgression from the standpoint of hatchery and sport-fishery managers, fish biologists, ecologists and evolutionary biologists has driven numerous studies focused on the development and/or utilization of molecular markers for assessing purity and hybridization of the two species (e.g. Philipp *et al.* 1983; Hallerman *et al.* 1986; Norgren *et al.* 1986; Maceina *et al.* 1988; Dunham *et al.* 1992; Nedbal & Philipp 1994; Williams *et al.* 1998; Johnson & Fulton 1999; Lutz-Carrillo *et al.* 2006, 2008; Barthel *et al.* 2010; Dumont & Lutz-Carrillo 2011; Austin *et al.* 2012; Seyoum *et al.* 2013). The early critical study of Florida and largemouth bass genetics (Philipp *et al.* 1983), while wide in scope, relied on allozymes and was hindered in its resolution by the availability of only two diagnostic loci with which to definitively assess species contributions to intergrade individuals. Many of the findings of that study are only beginning to be revisited utilizing microsatellite markers. Lutz-Carrillo *et al.* (2006, 2008) developed sets of 11 and 52 microsatellites that contained several private alleles and, unlike allozymes, did not require the sacrifice of sampled fish. However, these panels were optimized for technologies pre-dating capillary gel electrophoresis (e.g. ABI) and are unsuitable for higher-level multiplexing due to differing cycling and annealing temperature conditions. Extensive work by the Florida Fish and Wildlife Research Institute recently led to development of a panel of 18 additional microsatellites (Barthel *et al.* 2010; Seyoum *et al.* 2013) useful for genetic differentiation. However, this panel has no fixed-allelic differences between species at any locus and requires running of 3-4 ABI multiplexes, increasing the time and cost associated with performing the assay (Seyoum *et al.* 2013).

The dramatic decline in sequencing costs associated with next generation sequencing has increased the accessibility of single nucleotide polymorphism (SNP) markers for population genetics and genomics in nonmodel organisms (Hohenlohe *et al.* 2011; Rice *et al.* 2011). SNP markers are valued for their even genome-wide distribution, abundance, ease of multiplexing and low genotyping error rate for high-throughput analyses (Slate *et al.* 2009; Pritchard *et al.* 2012). They are distributed across coding and noncoding regions of the genome, making them particularly useful in studies examining traits and processes under selection during introgressive hybridization (e.g. Fitzpatrick *et al.* 2009; Shen *et al.* 2012). Among teleost fish species, salmonid researchers have been pioneering in efforts to develop diagnostic SNP assays useful in assessing and managing genetic integrity and measuring impacts of hybridization with non-native, introduced populations (e.g. Stephens *et al.* 2009; Hohenlohe *et al.* 2011; Kalinowski *et al.* 2011; Lamaze *et al.* 2012; Pritchard *et al.* 2012, 2013; Lamer *et al.* 2014). SNP development and application in these studies have generally taken one of two approaches: RNA-seq on pooled samples followed by validation in greater numbers of individual samples (Lamaze *et al.* 2012) or reduced-representation sequencing of individual samples using RAD-seq or GBS approaches (e.g. Hohenlohe *et al.* 2011; Li *et al.* 2014). Here, we chose the former approach, utilizing RNA-seq to develop comprehensive transcriptomes for both *M. salmoides* and *M. floridanus* and then identifying a set of gene-based SNPs with fixed-allelic differences between the two species. We then developed a subset of 25 of these markers into a diagnostic single multiplex assay and validated its capacity for assessing integrity and hybridization in hatchery and wild 'largemouth' bass.

Materials and methods

Sample collection for RNA-seq

Micropterus bass were collected from genotyped stocks held by American Sport Fish Hatchery (Montgomery, AL, USA). Genotyping was conducted based on a subset of microsatellite markers from Lutz-Carrillo *et al.* (2006) and Seyoum *et al.* (2013). Sixty micropterids were collected, these included 20 *M. floridanus*, 20 *M. salmoides* and 20 F1 hybrids (*M. salmoides* ♀ × *M. floridanus* ♂), with 10 males and 10 females selected from each group. Tissues including brain, liver, skin, spleen, intestine, gonad, muscle and kidneys were collected from each fish and immediately stored in 5 mL RNA later™ (Ambion, Austin, TX, USA) in separate tubes. Following an overnight incubation at 4 °C, the samples were stored at −80 °C until RNA extraction. Prior to RNA extraction, equal amounts of each tissue from the 20 fish within a group were homogenized into a master pool with mortar and pestle in the presence of liquid nitrogen. Total RNA was extracted using the RNeasy Universal Tissue Kit (Qiagen, Valencia, CA, USA). The three resulting master pools (Florida bass, largemouth bass, and F1) were carried forward for library construction.

Library construction and RNA-seq

Sequencing libraries were prepared with 2.14–3.25 μg of starting total RNA and processed using the Illumina TruSeq RNA Sample Preparation Kit, as dictated by the TruSeq protocol. The libraries were amplified with 15 cycles of PCR and contained TruSeq barcode indices, identifying each of the three groups, within the Illumina adapters. Amplified library yields were 30 μL of

19.8–21.4 ng/μL with an average length of ~270 bp, indicating a concentration of 110–140 nM. After KAPA quantitation and dilution, based on included DNA standards (1–6), the libraries were sequenced in a single lane on an Illumina HiSeq 2000 instrument with 100 bp paired-end (PE) reads at HudsonAlpha Genomic Services Lab (Huntsville, AL, USA). The image analysis, base calling and quality score calibration were processed using ILLUMINA PIPELINE SOFTWARE v1.5. FASTQ files containing the raw sequencing reads, quality scores and paired reads information were exported for the following trimming and assembly process.

De novo assembly and annotation of sequencing reads

Raw reads were processed for initial trimming by CLC Genomics Workbench (version 5.5.2; CLC Bio, Aarhus, Denmark). Before assembly, raw reads were trimmed by removing adapter sequences and ambiguous nucleotides. Reads with quality scores <20 and length below 30 bp were removed. The resulting high-quality sequences were used in the assembly. Assembly methodologies closely followed those described by Luo *et al.* (2014) and An *et al.* (2014), Li *et al.* (2014). Briefly, high-quality reads from the three barcoded pools (*M. salmoides*, *M. floridanus* and F1) were used to perform the *de novo* assembly using the Trinity assembler (v. 2014-04-13; Grabherr *et al.* 2011). This composite assembly was subsequently used for read mapping and SNP identification (below). The reads of each group were also assembled separately using Trinity, following the methodology of Luo *et al.* (2014), and subsequently annotated. The final assembled contigs from *M. salmoides*, *M. floridanus* and their F1 were used as queries against the NCBI nonredundant (NR) protein database and the UniProtKB/SwissProt (Uniprot) database using BLASTX by setting the cut-off Expect value (*E*-value, the likelihood that the matching sequence is obtained by chance) of 1e-20 and score ≥ 100.

SNP and microsatellite marker identification

The SNP detection module (CLC Genomics Workbench) and composite reference assembly were used to identify SNPs. The composite reference assembly was used to identify SNPs utilizing the SNP detection module included in CLC Genomics Workbench (CLC Bio). Mapping of reads from each pooled sample to the composite reference assembly sequence was performed with mismatch cost of 2, deletion cost of 3 and insertion cost of 3. The highest scoring matches that shared ≥95% similarity with the reference sequence across ≥90% of their length were included in the alignment. A minimum coverage (read depth) ≥10 was set for each group to assess the quality of reads at positions for SNP detection. Only biallelic SNPs were allowed. Given the use of pooled samples, we focused on the identification of SNPs with fixed-allelic differences between *M. salmoides* and *M. floridanus* (e.g. homozygous 'A' in *M. salmoides*, homozygous 'T' in *M. floridanus* and heterozygous 'A/T' in their F1). SNPs which showed the consensus base (100% allele frequency) in one species, and the alternative allele in the other species, with both alleles present in the F1 hybrid read file (minor allele frequency ≥10%, minimum coverage ≥ 10), were carried forward as putative fixed-allele diagnostic SNPs.

Microsatellite markers were additionally mined from the *M. salmoides* and *M. floridanus* transcriptomes using MSATFINDER version 2.0.9 (Thurston & Field 2005), with a repeat threshold of eight dinucleotide repeats or five tri-, tetra-, penta-, or hexa-nucleotide repeats. The SSR loci with at least 50-bp sequence on both sides of the microsatellite repeats were considered sufficient for primer design and captured from the candidate marker list.

Validation of fixed-allele interspecific SNPs

DNA was extracted from both blood samples and fin clips, with the source depending on scenarios in which different samples were collected (electrofishing/hatchery, etc.). Briefly, approximately 20 mg of fin clip samples or 200 μL of blood were isolated using the Qiagen DNeasy kit following the manufacturer's specifications (Qiagen). DNA concentration and purity were estimated using a NanoDrop ND-2000 UV-VIS Spectrophotometer as well as by electrophoresis on a 1.5% agarose gel.

The Sequenom MassARRAY platform (Sequenom, San Diego, CA, USA) was employed to validate a subset of identified SNPs. Sequenom assays were designed using the MASSARRAY ASSAY DESIGN Software with the goal to maximize multiplexing of 40 SNPs per well. Only SNPs with at least a 100 bp flanking region on either side of the polymorphic site were selected for the assay design. Amplification and extension reactions were performed using 20 (2 μL of 10 ng DNA) ng of DNA per sample and utilizing the iPLEX Gold Reagent Kit according to the manufacturer's protocols. SNP genotypes were called using the SEQUENOM SYSTEM TYPER 4.0 Analysis software. This software uses a three parameter model to calculate the significance of each genotype. A final genotype was called and assigned a particular name (e.g. conservative, moderate, aggressive, user call) based on the relative significance. Noncalls also were noted (e.g. low probability, bad spectrum). Individuals with lower than 90% call rates were removed or rerun. Based on initial screening, a subset of validated SNPs was merged into a single 25-plex panel through redesign of their extension primers. The final multiplex was tested on 227

samples (103 blood samples and 124 fin clip samples) from 11 populations (hatchery and wild). Subsets of each population were previously genotyped using diagnostic microsatellite markers (Lutz-Carrillo *et al.* 2006; Seyoum *et al.* 2013). Samples included 56 individuals from the Florida Bass Conservation Center, Webster, FL (directly and indirectly through American Sport Fish and Alabama Department of Conservation and Natural Resources), 18 individuals from a Georgia Department of Natural Resources hatchery (Perry, GA) including parents and 16 progeny, 34 individuals from American Sport Fish hatchery (Montgomery, AL) originally sourced from an unknown Illinois lake, 19 individuals from Sugar Lake, MN, 10 F1 individuals from American Sport Fish hatchery, 19 individuals from Lake Guntersville, AL (Tennessee River drainage) caught lake-wide during a tournament event, 15 electro-fished individuals from the North Sauty Creek arm of Lake Guntersville, 13 electro-fished individuals from the Spring Creek arm of Lake Guntersville, 16 individuals from Lay Lake, AL (Coosa River drainage) caught lake-wide during a tournament event, 14 electro-fished individuals from the Dry Creek arm of Lay Lake and 13 electro-fished individuals from the Beeswax Creek arm of Lay Lake.

Population summary statistics were calculated for Lake Guntersville and Lay Lake populations. Briefly, ARLEQUIN v.3.5.1.3 was used to calculate the observed (*Ho*) and expected heterozygosity (*He*) as well as carry out Hardy–Weinberg equilibrium testing (HWE; exact test using a Markov chain with chain length = 1 000 000 and dememorization steps = 100 000;Excoffier & Lischer 2010). The degree of linkage disequilibrium (LD) was estimated between all pairs of loci using the genotyped data from 25 SNPs. The exact genotypic disequilibrium test in GENEPOP 4.0 (Rousset 2008) was used to test the significance of LD.

Results and Discussion

Transcriptome sequencing, assembly and annotation

Illumina sequencing on pooled, barcoded multi-tissue RNA samples from *M. salmoides*, *M. floridanus* and their F1 hybrid generated over 273 million 100 bp reads with >84 million reads from each sample pool (Table S1). Raw reads are archived at NCBI Sequence Read Archive (SRA) under Accession SRP042097.

To generate a comprehensive reference transcriptome for SNP detection, the reads from *M. salmoides*. *M. floridanus* and their F1 were pooled together to generate a composite assembly using Trinity. A total of 343 632 contigs were generated with average contig size 788.9 bp and N50 size of 1182 bp for the composite assembly (Table S2). Simultaneously, species-specific assemblies were generated using Trinity for *M. salmoides*, *M. floridanus* and their F1 hybrid. Reads were assembled into 166 934 *M. floridanus* contigs, 227 220 *M. salmoides* contigs and 123 503 F1 contigs. Average contig sizes and N50 were 984.4 and 2096 bp, respectively, for *M. floridanus*, 1556 and 914.4 bp, respectively, for *M. salmoides*, and 2176 bp and 1017.1 bp for the F1, respectively (Table S2). Transcriptome assemblies have been deposited to NCBI's Transcriptome Shotgun Assembly (TSA) under Accessions GBFM00000000 (*M. floridanus*), GBGA00000000 (*M. salmoides*), GBFO00000000 (F1).

Annotation was carried out by BLAST against the UniProt and NR databases for *M. salmoides*, *M. floridanus* and their F1. Using the stringent criteria (*E*-value ≤ 1e-20, score ≥ 100), similar results were obtained from all three groups, with between 17 258 and 19 053 annotated unigene matches against UniProt and between 23 468 and 27 244 annotated unigene matches against NR (Table 1). Previous work generated a transcriptome from *M. salmoides* from the liver, gonad and brain tissues using 454 sequencing. This previous effort captured 7395 annotated genes, which, along with un-annotated features, were used to develop a toxicology-focused Agilent microarray (Garcia-Reyero *et al.* 2008; Mehinto *et al.* 2014; Richter *et al.* 2014). Contig sequences from this project are not publicly available and were short in length. Our results provide a more comprehensive transcriptome from the two bass species, encompassing many more genes and benefitting from longer contig lengths. As bass are an important model for aquatic toxicology (Denslow *et al.* 2007), this resource should aid future QPCR, microarray and RNA-seq studies in this field as well as others.

Table 1 Summary of gene identification and annotation of assembled *M. floridanus* and *M. salmoides* and F1 hybrid contigs based on BLAST homology searches against various protein databases (UniProt and NR) as well as statistics of fixed interspecific SNPs identified between *M. floridanus* and *M. salmoides*

Transcriptome and SNP coverage	*M. floridanus*	*M. salmoides*	F1
Unigene matches (UniProt)	19 053	22 412	17 258
Unigene matches (NR)	23 709	27 244	23 468
SNPs with coverage ≤20X	160	140	718
SNPs with coverage 21–50X	1055	1878	1400
SNPs with coverage 51–100X	1148	1114	841
SNPs with coverage 101–500X	1262	519	692
SNPs with coverage >500X	49	23	23
Average coverage	109	68	70
SNP annotation			
Total number of SNPs	3674		
Annotated SNPs (NR)	3445		
Annotated SNPs from unique genes	2112		

Microsatellite marker identification in M. floridanus and M. salmoides transcriptomes

Although not the focus of the current study, we also mined microsatellites from the transcriptome assemblies of Florida bass and largemouth bass. The higher allelic richness of microsatellites makes them superior for some applications in structure and parentage analysis (Lapegue *et al.* 2014). In *M. floridanus*, from a total of 13 354 microsatellites identified by MSATFINDER, 51.71% (6,905) had sufficient flanking regions to allow design of primers. The microsatellite-bearing contigs had 4376 putative gene matches to the NR database from 2576 unique genes (Tables S3 and S4).

Similarly, in *M. salmoides*, from a total of 13 099 microsatellites identified by MSATFINDER, 59.60% (7807) had sufficient flanking regions to allow design of primers. The microsatellite-bearing contigs had 4964 putative gene matches to the NR database from 2249 unique genes (Tables S3 and S4).

SNP identification in M. floridanus and M. salmoides

Given the complexity of determining genotypes from pooled populations, we focused on the identification of SNPs with fixed-allelic differences between species (i.e. homozygous 'A' in *M. salmoides*, homozygous 'T' in *M. floridanus* and heterozygous 'A/T' in F1) similar to the approach of Lamaze *et al.* (2012). We detected a set of 3674 SNPs with fixed-allelic differences using the parameters and cut-off values described in the Methods. These SNP contigs had 3445 putative gene matches to the NR database from 2112 unique genes (Table 1). Details of the 3674 SNPs, with contig ID, genotypes, coverage and annotation information, are provided in Table S5. Average read coverage in *M. floridanus*, *M. salmoides* and F1 was 109 reads/SNP, 68 reads/SNP and 70 reads/SNP, respectively.

Validation of SNPs by Sequenom MassARRAY

To determine the accuracy and usefulness of this resource for the study of genetic integrity and introgression of *M. salmoides* and *M. floridanus*, a subset of the fixed-allelic SNPs was tested on 227 individual bass samples in multiplex panels on the Sequenom MassARRAY. From a total of 37 randomly chosen SNPs, 28 SNPs were validated based on the larger sample set (76%). Failing SNPs, although amplifying, showed allelic patterns deviating from those expected by RNA-seq, likely representing either rare alleles previously uncaptured or SNPs within duplicated genes. Of the 28 high-quality, fixed-allele SNPs, 25 were amenable to remultiplexing through redesign of mass-specific extension primers. The final 25-plex SNP panel was genotyped across the 227 individuals from eleven populations sourced from hatchery and wild populations. Details of the 25 SNP markers are provided in Table 2, including contig ID, species-specific genotypes and coverage based on RNA-seq, and gene annotation. Putative functions of the encoding genes are also given. Future studies examining phenotypic differences between the two bass species and selective pressures on allele usage in hybrid populations may benefit from use of these markers (Redenbach & Taylor 2003; Fitzpatrick *et al.* 2009). Individual genotyping results and multiplex primer information are provided in Table S6. A total of 20 individuals were run on multiple plates with 99.6% of genotypes matching among technical replicates (data not shown).

As summarized in Table 3, 'pure' Florida bass (n = 74), based on previous microsatellite genotyping, had the *M. floridanus* allele in 99.5% of genotypes, while 'pure' largemouth bass (n = 53) had the *M. salmoides* allele in 100% of genotypes on average. F1 hybrids were heterozygous (50% *M. floridanus*, 50% *M. salmoides* allele frequencies) at all 25 loci. To avoid ascertainment bias, in all cases, we genotyped additional 'pure' individuals from two or more populations not present in the original RNA-seq pools. In our attempt to avoid natural or anthropogenic 'intergrade' individuals, diagnostic markers were validated on fish from the more isolated edges of largemouth bass and Florida bass ranges. It is likely therefore that additional genetic variation will be revealed at some of the loci through future genotyping of fish closer to the still disputed intergrade zone (Bailey & Hubbs 1949; Philipp *et al.* 1983). In many cases, however, widespread stocking of *M. floridanus* in the southeastern US presents a significant obstacle to distinguishing between natural intraspecific variation in *M. salmoides* and historical signatures of hybridization with introduced Florida bass.

To test performance of these markers in wild populations, we also examined three groups of fish from two large Alabama reservoirs, Lake Guntersville in North Alabama, Tennessee River Drainage, and Lay Lake in Central Alabama, Coosa River Drainage. Summary statistics based on genotyping in these populations are provided in Table S7. From each reservoir, we genotyped fish caught lake-wide as well as fish caught in specific regions of the lake by electrofishing (nearby Florida bass stocking sites and distant). Both lakes have been the focus of extensive stocking of Florida bass over the last 30 years, with over two times more Florida bass stocked into the considerably smaller Lay Lake (4800 ha) than into Lake Guntersville (28 000 ha; Table 3). Lay Lake, on average (n = 43), showed the greater Florida influence with 51% of alleles coming from *M. floridanus* and 49% of

Table 2 Detailed information of the 25-SNP Sequenom multiplex, with genotype, read coverage in *M. floridanus*, *M. salmoides*, and F1 largemouth bass, and gene annotation. Putative assigned functions are based on functions of orthologous genes in other species, with a supporting work cited

Contig	Position	*M. floridanus* genotype	*M. salmoides* genotype	F1 genotype	Gene name	Function
Contig2930	875	T(232)	A(304)	T/A(171)	Carboxypeptidase D	Immune (1)
Contig26936	5048	T(191)	G(134)	T/G(120)	Splicing factor, proline-and glutamine-rich	Immune/stress (2)
Contig25677	809	C(179)	T(158)	C/T(195)	Vacuolar protein-sorting-protein 25	Endocytosis (3)
Contig17385	4548	A(298)	G(269)	G/A(461)	Clustered mitochondria protein homolog	Mitochondrial
Contig8751	2930	G(212)	A(234)	A/G(236)	Mitochondrial glutamate carrier 1	Glucose homoeostasis (4)
Contig34438	153	T(167)	A(66)	A/T(62)	Protein kinase C and casein kinase substrate	Endocytosis (5)
Contig4716	1720	G(430)	A(285)	A/G(406)	CpG-binding protein-like	Expression regulation (6)
Contig10770	6042	T(119)	A(258)	T/A(87)	Protein VPRBP-like	Immune (7)
Contig25196	1374	A(99)	G(71)	A/G(129)	Putative transferase CAF17 homolog, mitochondrial	Heme biosynthesis (8)
Contig5903	2988	T(652)	A(326)	T/A(480)	Calcium/calmodulin-dependent protein kinase type II	Neural function (9)
Contig33105	2477	C(99)	T(196)	T/C(81)	Acyl-CoA dehydrogenase, very long chain	Fatty acid metabolism (10)
Contig15421	896	T(226)	A(252)	A/T(168)	Serine incorporator 1	Lipid biosynthesis (11)
Contig35139	851	T(400)	C(292)	C/T(207)	Repressor of RNA polymerase III transcription MAF1	Nutrient-dependent growth (12)
Contig2993	1992	T(458)	A(259)	T/A(335)	DNAJ homolog subfamily C member 7	Steroid receptor chaperone (13)
Contig11367	889	A(154)	G(164)	G/A(369)	Kinesin-like protein KIF22	Neural function (14)
Contig11367	748	G(96)	A(120)	A/G(289)	Kinesin-like protein KIF22	Neural function (14)
Contig35112	957	A(277)	C(362)	A/C(103)	Calreticulin	Chaperone (15)
Contig33087	1916	A(466)	G(345)	A/G(355)	Heat-shock protein 60 kDa, mitochondrial	Chaperone (16)
Contig6106	1199	T(111)	C(168)	C/T(232)	Fanconi anaemia group F protein	DNA repair (17)
Contig20911	1554	G(149)	A(215)	A/G(308)	Mitochondrial import receptor subunit TOM40	Mitochondrial (18)
Contig19092	233	G(627)	T(269)	G/T(237)	Nonspecific cytotoxic cell receptor protein-1	Immune (19)
Contig31992	101	C(1686)	T(72)	C/T(383)	Choriogenin L	Reproductive (20)
Contig5885	2325	G(485)	A(149)	G/A(506)	Spermatogenesis associated 2-like	Reproductive (21)
Contig31857	326	C(831)	T(337)	C/T(481)		
Contig32455	234	C(1226)	A(514)	A/C(585)		

(1) Hadkar & Skidgel (2001), (2) Imamura *et al.* (2014), (3) Yorikawa *et al.* (2005), (4) Casimir *et al.* (2009), (5) Goh *et al.* (2012), (6) Ansari *et al.* (2008), (7) Kassmeier *et al.* (2012), (8) Mandilaras & Missirlis (2012), (9) Rodrigues *et al.* (2004), (10) Tucci *et al.* (2010), (11) Inuzuka *et al.* (2005), (12) Rideout *et al.* (2012), (13) Moffatt *et al.* (2008), (14) Blaker-Lee *et al.* (2012), (15) Wang *et al.* (2012), (16) An *et al.* (2014), (17) Zhao *et al.* (2014), (18) Bender *et al.* (2013), (19) Cai *et al.* (2013), (20) Bugel *et al.* (2014), (21) Onisto *et al.* (2001).

alleles from *M. salmoides* when compared with Lake Guntersville (n = 47; 32% *M. floridanus* /68% *M. salmoides*). There was also potential evidence of site-specific impacts of stocking as the two sampling areas nearby Florida bass release sites showed a greater presence of Florida alleles in both reservoirs (North Sauty Creek, Guntersville and Beeswax Creek, Lay). Averages, however, masked considerable variation among individual fish, with fish in Lay ranging from 94% *M. floridanus* to 77% *M. salmoides* and fish in Guntersville ranging from 63% *M. floridanus* to 84% *M. salmoides*. No sampled individual in either population was fixed for Florida or largemouth bass alleles at all loci, and no F1 hybrid individuals were re-covered, indicating that these populations likely exist as hybrid swarms (Allendorf *et al.* 2001; Hasselman *et al.* 2014), as observed previously in

Table 3 Genotyping summary of the bass multiplex averaged across individuals within each population. Stocking refers to historical stocking records of *M. floridanus* by ADNR in 1981–1999 and 1983–2000 for Guntersville and Lay Lakes, respectively

Population	*N*	*M. floridanus* allele frequency	*M. salmoides* allele frequency	*M. floridanus* stocking
Florida Bass Conservation Center, FL	56	1.00	0.00	
State Fish Hatchery, GA	18	0.99	0.01	
Total/Average	74	0.995	0.005	
American Sport Fish Hatchery (Illinois)	34	0.00	1.00	
Sugar Lake, Minnesota	19	0.00	1.00	
Total/Average	53	0.00	1.00	
F1 (American Sport Fish Hatchery)	10	0.5	0.5	
Lake Guntersville, AL Lake-wide	19	0.35	0.65	571 000
Lake Guntersville, AL North Sauty Creek*	15	0.35	0.65	
Lake Guntersville, AL Spring Creek†	13	0.32	0.68	
Total/Average	47	0.34	0.66	
Lay Lake, AL Lake-wide	16	0.55	0.45	1.35M‡
Lay Lake, AL Dry Creek†	14	0.45	0.55	
Lay Lake, AL Beeswax Creek*	13	0.54	0.46	
Totals/Average	43	0.51	0.49	
Total	227			

*Sampling nearby the site of *M. floridanus* stocking.
†A sampling location distant relative to the stocking site.
‡Does not include additional private tournament sponsor stocking of unknown quantities of *M. floridanus* prior to restrictions.

sampling of Florida intergrade populations (Barthel *et al.* 2010).

Although not the primary purpose of this study, a direct comparison of these two systems is complicated by the fact that Lay Lake and the Coosa River Drainage was likely in the centre of the natural intergrade range, while Guntersville (Tennessee River Drainage) was at the northern edge of this range. This is supported by allozyme results obtained prior to much of the state-directed stocking of Florida bass (Philipp *et al.* 1983). That study reported a 92.5% *M. salmoides* contribution in Guntersville and an 87.5% *M. salmoides* contribution in Lake Mitchell (just below Lay Lake in the Coosa system; Lay Lake was not sampled by Philipp *et al.* 1983). A subsequent study by Norgren *et al.* (1986), however, found that Jordan Lake, south of Lake Mitchell on the Coosa River, was 99% *M. salmoides* despite previous *M. floridanus* stocking. Ancestral genetic makeup, river drainage, latitude, lake characteristics, etc. are all likely impacting the success of introgressive hybridization (Norgren *et al.* 1986). Regardless, based on our results, it is clear that persistent *M. floridanus* stocking in these systems have had a sizeable impact on populations within these drainages within recent decades.

While this study focuses on SNP resource development and initial validation of a subset of fixed-allele SNP markers, future studies should expand testing of these and additional SNP panels (developed from Table S5) to better define the distributions of pure *M. floridanus* and pure *M. salmoides* and update the intergrade hybrid zone, given the improvements in resolution offered by these additional markers and the continued widespread stocking of *M. floridanus* since the last range-wide study (Philipp *et al.* 1983). The lower cost and higher throughput nature of SNPs relative to other marker types would greatly improve the feasibility of such a study. Also of great interest would be use of this resource to examine allelic usage patterns in key functional genes in naturally introgressed populations when compared with populations with anthropogenic impacts on hybridization through stocking. This should be performed across the wide range of environments where bass thrive (Jiggins & Mallet 2000; Martinsen *et al.* 2001; Gompert *et al.* 2006; Payseur 2010; Carson *et al.* 2012; DeVries *et al.* 2014).

Acknowledgements

We are grateful to public and private collaborators for their assistance in sample collection, particularly Damon Abernathy at the Alabama Department of Conservation and Natural Resources, Don Keller at American Sport Fish and biologists at the Florida Fish and Wildlife Conservation Commission and Georgia Department of Natural Resources. The study was conducted with funding from the Alabama Department of Conservation and Natural Resources as well as federal Dingell-Johnson Sport Fish Restoration dollars. The USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

References

Addison J, Spencer S (1971) Preliminary evaluation of three strains of largemouth bass, *Micropterus salmoides* (Lacepede), stocked in ponds in south Alabama. *Proceedings of the annual conference of the Southeastern Association of Game and Fish Commissioners*, **25**, 366–374.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.

An L-H, K Lei, Zheng B-H (2014) Use of heat shock protein mRNA expressions as biomarkers in wild crucian carp for monitoring water quality. *Environmental Toxicology and Pharmacology*, **37**, 248–255.

Ansari KI, Mishra BP, Mandal SS (2008) Human CpG binding protein interacts with MLL1, MLL2 and hSet1 and regulates Hox gene expression. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, **1779**, 66–73.

Austin JD, Johnson A, Matthews M *et al.* (2012) An assessment of hatchery effects on Florida bass (*Micropterus salmoides floridanus*) microsatellite genetic diversity and sib-ship reconstruction. *Aquaculture Research*, **43**, 628–638.

Bailey RM, Hubbs CL (1949) The black basses (*Micropterus*) of Florida with description of a new species. *University of Michigan Museum of Zoology Occasional Papers*, **516**, 1–40.

Barthel BL, Lutz-Carrillo DJ, Norberg KE *et al.* (2010) Genetic relationships among populations of Florida bass. *Transactions of the American Fisheries Society*, **139**, 1615–1641.

Bender A, Desplats P, Spencer B *et al.* (2013) TOM40 mediates mitochondrial dysfunction induced by α-synuclein accumulation in Parkinson's disease. *PLoS One*, **8**, e62277.

Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H (2012) Zebrafish homologs of genes within 16p11. 2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. *Disease Models & Mechanisms*, **5**, 834–851.

Bolnick DI, Near TJ (2005) Tempo of hybrid inviability in centrarchid fishes (*Teleostei: Centrarchidae*). *Evolution*, **59**, 1754–1767.

Bugel SM, Bonventre JA, White LA, Tanguay RL, Cooper KR (2014) Chronic exposure of killifish to a highly polluted environment desensitizes estrogen-responsive reproductive and biomarker genes. *Aquatic Toxicology*, **152**, 222–231.

Cai J, Wei S, Wang B *et al.* (2013) Cloning and expression analysis of nonspecific cytotoxic cell receptor 1 (Ls-NCCRP1) from red snapper (*Lutjanus sanguineus*). *Marine Genomics*, **11**, 39–44.

Carson EW, Tobler M, Minckley W, Ainsworth RJ, Dowling TE (2012) Relationships between spatio-temporal environmental and genetic variation reveal an important influence of exogenous selection in a pupfish hybrid zone. *Molecular Ecology*, **21**, 1209–1222.

Casimir M, Lasorsa FM, Rubi B *et al.* (2009) Mitochondrial glutamate carrier GC1 as a newly identified player in the control of glucose-stimulated insulin secretion. *Journal of Biological Chemistry*, **284**, 25004–25014.

Chew R (1975) The Florida Largemouth Bass. In: *Black Bass Biology and Management* (ed. Clepper H), pp. 450–458. Sport Fishing Institute, Washington, DC.

Cooke S, Philipp D (2006) Hybridization among divergent stocks of largemouth bass (*Micropterus salmoides*) results in altered cardiovascular performance: the influence of genetic and geographic distance. *Physiological and Biochemical Zoology*, **79**, 400–410.

Denslow ND, Garcia-Reyero N, Barber DS (2007) Fish 'n'chips: the use of microarrays for aquatic toxicology. *Molecular BioSystems*, **3**, 172–177.

DeVries DR, Wright RA, Glover DC *et al.* (2014) Largemouth bass in coastal estuaries: a comprehensive study from the Mobile-Tensaw River Delta, Alabama. *Southern Division of the American Fisheries Society Symposium, Black Bass Diversity: Multidisciplinary Science for Conservation*, in press.

Dumont SC, Lutz-Carrillo DJ (2011) Estimating genetic composition in introgressed largemouth bass populations: does age matter? *North American Journal of Fisheries Management*, **31**, 176–181.

Dunham RA, Turner C, Reeves W (1992) Introgression of the Florida largemouth bass genome into native populations in Alabama public lakes. *North American Journal of Fisheries Management*, **12**, 494–498.

Excoffier L, Lischer HE (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fitzpatrick BM, Johnson JR, Kump DK *et al.* (2009) Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. *BMC Evolutionary Biology*, **9**, 176.

Garcia-Reyero N, Griffitt RJ, Liu L *et al.* (2008) Construction of a robust microarray from a non-model species largemouth bass, *Micropterus salmoides* (Lacèpede), using pyrosequencing technology. *Journal of Fish Biology*, **72**, 2354–2376.

Goh SL, Wang Q, Byrnes LJ, Sondermann H (2012) Versatile membrane deformation potential of activated pacsin. *PLoS One*, **7**, e51628.

Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC (2006) Homoploid hybrid speciation in an extreme habitat. *Science*, **314**, 1923–1925.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Hadkar V, Skidgel RA (2001) Carboxypeptidase D is up-regulated in RAW 264.7 macrophages and stimulates nitric oxide synthesis by cells in arginine-free medium. *Molecular Pharmacology*, **59**, 1324–1332.

Hallerman EM, Smitherman R, Reed RB, Tucker WH, Dunham RA (1986) Biochemical genetics of largemouth bass in mesosaline and freshwater areas of the Alabama River system. *Transactions of the American Fisheries Society*, **115**, 15–20.

Hasselman DJ, Argo EE, McBride MC *et al.* (2014) Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Molecular Ecology*, **23**, 1137–1152.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Imamura K, Imamachi N, Akizuki G *et al.* (2014) Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Molecular Cell*, **53**, 393–406.

Inuzuka M, Hayakawa M, Ingi T (2005) Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *Journal of Biological Chemistry*, **280**, 35776–35783.

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, **15**, 250–255.

Johnson R, Fulton T (1999) Persistence of Florida largemouth bass alleles in a northern Arkansas population of largemouth bass, *Micropterus salmoides Lacepede*. *Ecology of Freshwater Fish*, **8**, 35–42.

Kalinowski S, Novak B, Drinan D, Dem Jennings R, Vu N (2011) Diagnostic single nucleotide polymorphisms for identifying westslope cutthroat trout (*Oncorhynchus clarki lewisi*), Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and rainbow trout (*Oncorhynchus mykiss*). *Molecular Ecology Resources*, **11**, 389–393.

Kassler T, Koppelman J, Near T *et al.* (2002) Molecular and morphological analyses of the black basses: implications for taxonomy and conservation. *American Fisheries Society Symposium*, **31**, 291–322.

Kassmeier MD, Mondal K, Palmer VL *et al.* (2012) VprBP binds full-length RAG1 and is required for B-cell development and V (D) J recombination fidelity. *The EMBO Journal*, **31**, 945–958.

Lamaze FC, Sauvage C, Marie A, Garant D, Bernatchez L (2012) Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, **21**, 2877–2895.

Lamer JT, Sass GG, Boone JQ *et al.* (2014) Restriction site-associated DNA sequencing generates high-quality single nucleotide polymorphisms for assessing hybridization between bighead and silver carp in the United States and China. *Molecular Ecology Resources*, **14**, 79–86.

Lapegue S, Harrang E, Heurtebise S *et al.* (2014) Development of SNP genotyping arrays in two shellfish species. *Molecular Ecology Resources*, **14**, 820–830.

Li C, Waldbieser G, Bosworth B *et al.* (2014) SNP discovery in wild and domesticated populations of blue catfish, *Ictalurus furcatus*, using GBS and subsequent SNP validation. *Molecular Ecology Resources*, in Press.

Luo Y, Li C, Lndis A *et al.* (2014) Transcriptomic profiling of differential responses to drought in two freshwater mussel species, the giant floater *Pyganodon grandis* and the pondhorn *Uniomerus tetralasmus*. *PLoS One*, **9**, e89481.

Lutz-Carrillo DJ, Nice CC, Bonner TH, Forstner MR, Fries LT (2006) Admixture analysis of Florida largemouth bass and northern largemouth bass using microsatellite loci. *Transactions of the American Fisheries Society*, **135**, 779–791.

Lutz-Carrillo DJ, Hagen C, Dueck LA, Glenn TC (2008) Isolation and characterization of microsatellite loci for Florida largemouth bass, *Micropterus salmoides floridanus*, and other micropterids. *Molecular Ecology Resources*, **8**, 178–184.

MacCrimmon HR, Robbins WH (1975) Distribution of the black basses in North America. In: *Black Bass Biology and Management* (ed. Clepper H), pp. 56–66. Sport Fishing Institute, Washington, DC.

Maceina MJ, Murphy BR (1992) Stocking Florida largemouth bass outside its native range. *Transactions of the American Fisheries Society*, **121**, 686–691.

Maceina MJ, Murphy BR, Isely JJ (1988) Factors regulating Florida largemouth bass stocking success and hybridization with northern largemouth bass in Aquilla Lake, Texas. *Transactions of the American Fisheries Society*, **117**, 221–231.

Mandilaras K, Missirlis F (2012) Genes for iron metabolism influence circadian rhythms in *Drosophilia melanogaster*. *Metallomics*, **4**, 928–936.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.

Mehinto AC, Prucha MS, Colli-Dula RC *et al.* (2014) Gene networks and toxicity pathways induced by acute cadmium exposure in adult largemouth bass (*Micropterus salmoides*). *Aquatic Toxicology*, **152**, 186–194.

Moffatt NSC, Bruinsma E, Uhl C, Obermann WM, Toft D (2008) Role of the cochaperone Tpr2 in Hsp90 chaperoning†. *Biochemistry*, **47**, 8203–8213.

Near TJ, Kassler TW, Koppelman JB, Dillman CB, Philipp DP (2003) Speciation in North American black basses, Micropterus (*Actinopterygii: Centrarchidae*). *Evolution*, **57**, 1610–1621.

Nedbal MA, Philipp DP (1994) Differentiation of mitochondrial DNA in largemouth bass. *Transactions of the American Fisheries Society*, **123**, 460–468.

Norgren K, Dunham R, Smitherman R, Reeves W (1986) Biochemical genetics of largemouth bass populations in Alabama. *Annual Conference of Southeast Association of Fish and Wildlife Agencies*, **40**, 194–205.

Onisto M, Slongo L, Graziotto R *et al.* (2001) Evidence for FSH-dependent upregulation of SPATA2 (spermatogenesis-associated protein 2). *Biochemical and Biophysical Research Communications*, **283**, 86–92.

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.

Philipp DP, Claussen JE (1995) Fitness and performance differences between two stocks of largemouth bass from different river drainages within Illinois. *American Fisheries Society Symposium*, **15**, 236–243.

Philipp DP, Childers WF, Whitt GS (1983) A biochemical genetic evaluation of the northern and Florida subspecies of largemouth bass. *Transactions of the American Fisheries Society*, **112**, 1–20.

Pritchard V, Abadía-Cardoso A, Garza J (2012) Discovery and characterization of a large number of diagnostic markers to discriminate *Oncorhynchus mykiss* and *O. clarkii*. *Molecular Ecology Resources*, **12**, 918–931.

Pritchard VL, Campbell NR, Narum SR, Peacock MM, Garza JC (2013) Discovery and characterization of novel genetic markers for use in the management of Lahontan cutthroat trout (*Oncorhynchus clarkii henshawi*). *Molecular Ecology Resources*, **13**, 276–288.

Redenbach Z, Taylor E (2003) Evidence for bimodal hybrid zones between two species of char (Pisces: Salvelinus) in northwestern North America. *Journal of Evolutionary Biology*, **16**, 1135–1148.

Rice AM, Rudh A, Ellegren H, Qvarnström A (2011) A guide to the genomics of ecological speciation in natural animal populations. *Ecology Letters*, **14**, 9–18.

Richter CA, Martyniuk CJ, Annis ML *et al.* (2014) Methylmercury-induced changes in gene transcription associated with neuroendocrine disruption in largemouth bass (*Micropterus salmoides*). *General and Comparative Endocrinology*, in Press.

Rideout EJ, Marshall L, Grewal SS (2012) Drosophila RNA polymerase III repressor Maf1 controls body size and developmental timing by modulating tRNAiMet synthesis and systemic insulin signaling. *Proceedings of the National Academy of Sciences*, **109**, 1139–1144.

Rodrigues SM, Farb CR, Bauer EP, LeDoux JE, Schafe GE (2004) Pavlovian fear conditioning regulates Thr286 autophosphorylation of Ca2+/calmodulin-dependent protein kinase II at lateral amygdala synapses. *The Journal of Neuroscience*, **24**, 3281–3288.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Seyoum S, Barthel BL, Tringali MD *et al.* (2013) Isolation and characterization of eighteen microsatellite loci for the largemouth bass, *Micropterus salmoides*, and cross amplification in congeneric species. *Conservation Genetics Resources*, **5**, 697–701.

Shen Y, Catchen J, Garcia T *et al.* (2012) Identification of transcriptome SNPs between *Xiphophorus* lines and species for assessing allele specific gene expression within F1 interspecies hybrids. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **155**, 102–108.

Slate J, Gratten J, Beraldi D *et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica*, **136**, 97–107.

Stephens MR, Clipperton NW, May B (2009) Subspecies-informative SNP assays for evaluating introgression between native golden trout and introduced rainbow trout. *Molecular Ecology Resources*, **9**, 339–343.

Thurston M, Field D (2005) MSATFINDER: detection and characterisation of microsatellites. Distributed by the authors at http://www.bioinformatics.org/groups/?group_id=469. CEH Oxford, Mansfield Road, Oxford OX1 3SR.

Tucci S, Primassin S, Spiekerkoetter U (2010) Fasting-induced oxidative stress in very long chain acyl-CoA dehydrogenase-deficient mice. *FEBS Journal*, **277**, 4699–4708.

Wang W-A, Groenendyk J, Michalak M (2012) Calreticulin signaling in health and disease. *The International Journal of Biochemistry & Cell Biology*, **44**, 842–846.

Williams DJ, Kazianis S, Walter RB (1998) Use of random amplified polymorphic DNA (RAPD) for identification of largemouth bass subspecies and their intergrades. *Transactions of the American Fisheries Society*, **127**, 825–832.

Yorikawa C, Shibata H, Waguri S *et al.* (2005) Human CHMP6, a myristoylated ESCRT-III protein, interacts directly with an ESCRT-II component EAP20 and regulates endosomal cargo sorting. *Biochemical Journal*, **387**, 17–26.

Zhao L, Li Y, He M *et al.* (2014) The Fanconi anemia pathway sensitizes to DNA alkylating agents by inducing JNK-p53-dependent mitochondrial apoptosis in breast cancer cells. *International Journal of Oncology*, **45**, 129–138.

E.P. planned the project and provided funding. E.P., C.L. and S.G. wrote the manuscript. C.L. and E.P. carried out data analysis. B.B., A.A., S.G., H.K. and L.K. assisted in sample collection and preparation and provided critical editing of the manuscript. S.G., A.A., W.T., H.K. and L.K. carried out the SNP validation.

Data accessibility

Transcriptome assemblies have been deposited to NCBI's Transcriptome Shotgun Assembly (TSA) under Accessions GBFM00000000 (*M. floridanus*), GBGA0 0000000 (*M. salmoides*) and GBFO00000000 (F1). Raw RNA-seq reads are archived at NCBI Sequence Read Archive (SRA) under Accession SRP042097. Composite assembly contigs are available upon request. Filtered SNP primer information for the Sequenom multiplex is provided in Table S6.

Supporting Information

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

Table S1 Sequencing statistics from *Micropterus* sp. RNA-seq samples.

Table S2 Summary of Trinity *de novo* assembly results of Illumina RNA-seq data from *M. salmoides*, *M. floridanus*, F1 and the composite assembly.

Table S3 Statistics of simple sequence repeats (SSRs) identified from *M. floridanus* and *M. salmoides* transcriptomes.

Table S4 Details of the simple sequence repeats (SSRs) identified from *M. floridanus* and *M. salmoides* transcriptomes.

Table S5 Details of the 3674 putative fixed-allelic SNPs, with contig ID, genotypes, coverage and annotation information.

Table S6 Sequenom-generated SNP calls for the 25-SNP multiplex run and multiplex primer sequence information.

Table S7. Summary statistics of six largemouth populations based on 25 SNP loci.