Title

Use of single nucleotide polymorphisms identifies backcrossing and species misidentifications among three San Francisco estuary osmerids

Journal

Conservation Genetics, 19(3)

ISSN

1566-0621

Authors

Benjamin, Alyssa
Sağlam, İsmail K
Mahardja, Brian
et al.

Publication Date

2018-06-01

DOI

10.1007/s10592-018-1048-9

Peer reviewed

Conservation Genetics
https://doi.org/10.1007/s10592-018-1048-9

RESEARCH ARTICLE

Use of single nucleotide polymorphisms identifies backcrossing and species misidentifications among three San Francisco estuary osmerids

Alyssa Benjamin[1] · İsmail K. Sağlam[1,2] · Brian Mahardja[3] · James Hobbs[4] · Tien-Chieh Hung[5] · Amanda J. Finger[1]

Received: 8 September 2017 / Accepted: 11 January 2018

Abstract

Two threatened osmerid species native to the San Francisco Estuary (SFE)—Delta Smelt (*Hypomesus transpacificus*) and Longfin Smelt (*Spirinchus thaleichthys*)—are subject to broad human influence, including significant habitat alteration and the presence of the introduced osmerid, Wakasagi (*Hypomesus nipponensis*). The identification of these closely related species and their hybrids is difficult in field collected specimens which are subject to damage through handling and may be difficult to identify morphologically, especially when young. In addition, it is known that these three species hybridize, but the extent and effect of hybridization is difficult to quantify and monitor. We developed assays for 24 species-specific single nucleotide polymorphisms (SNPs) that identify whether a sample is a pure species (Delta Smelt, Longfin Smelt, or Wakasagi), a first generation (F_1) hybrid, or a backcross. We used this SNP panel to genetically identify wild osmerids collected in Yolo Bypass from 2010 to 2016 and detected nine Delta Smelt × Wakasagi F_1 hybrids and two Wakasagi × (Delta Smelt × Wakasagi) backcross hybrids; all assayed hybrids had Wakasagi as the maternal parent. The backcrossing into Wakasagi suggests that hybridization may only occur in one direction and thus preclude introgression to Delta Smelt. We also found substantial morphological field misidentifications (32.7%) in the Yolo Bypass samples resulting in more Wakasagi and fewer Delta Smelt than previously recorded when based on morphology. The SNP panel described in this study constitutes a valuable resource for monitoring hybridization in the SFE and assigning species identifications with accuracy and efficiency.

Keywords Hybridization · SNP · RADseq · Delta Smelt · *Hypomesus transpacificus* · San Francisco Estuary

Introduction

Correct identification of species is critical to ecological monitoring. Biological surveys rely on field identifications to provide information about population abundance, density, and distribution—all measurements which are used to set conservation priorities. While many studies have addressed issues related to incomplete detection, monitoring programs often make data inferences with an assumption of 100% identification accuracy (Elphick 2008; Kellner and Swihart 2014; Kirsch et al. 2014). However, field identifications can be difficult due observer bias, non-distinctive morphology between species, or the presence of interspecific hybrids, resulting in inaccurate population assessments for a number of species (Beerkircher et al. 2009; Hull et al. 2010; Shea et al. 2011). Genetic tools can offer a powerful supplemental approach when combined with traditional field surveys. Diagnostic molecular markers can be used to identify

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10592-018-1048-9) contains supplementary material, which is available to authorized users.

✉ Alyssa Benjamin
abenjamin@ucdavis.edu

[1] Department of Animal Science, University of California-Davis, One Shields Avenue, Davis, CA 95616, USA

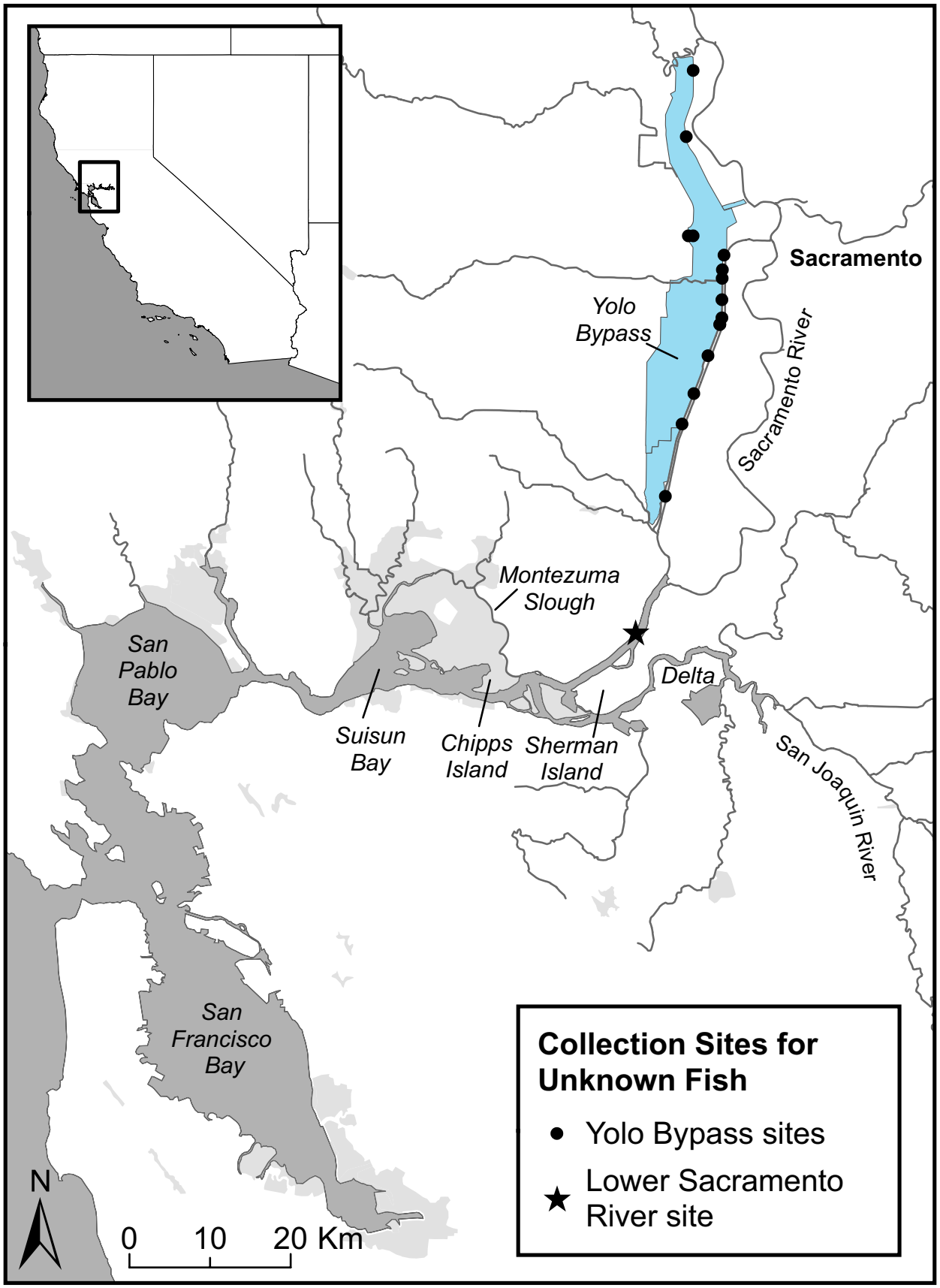
[2] Ecological Sciences Research Laboratories, Department of Biology, Hacattepe University, Beytepe, Ankara 06800, Turkey

[3] California Department of Water Resources, 3500 Industrial Boulevard, West Sacramento, CA 95691, USA

[4] Department of Wildlife, Fish, and Conservation Biology, University of California-Davis, One Shields Avenue, Davis, CA 95616, USA

[5] Department of Biological and Agricultural Engineering, University of California-Davis, One Shields Avenue, Davis, CA 95616, USA

Published online: 17 January 2018


Sacramento
Yolo Bypass
Sacramento River
Montezuma Slough
San Pablo Bay
Suisun Bay
Chipps Island
Sherman Island
Delta
San Joaquin River
San Francisco Bay
N
0 10 20 Km
Collection Sites for Unknown Fish
Yolo Bypass sites
Lower Sacramento River site

◂**Fig. 1** Map of the San Francisco Estuary, including the location of the Yolo Bypass and sample collection sites for the unknown Yolo Bypass and lower Sacramento River osmerids

species and their hybrids without total reliance on highly variable morphological features (Schwartz et al. 2007).

Two intensively monitored fish species in California include the Delta Smelt (DSM; *Hypomesus transpacificus*) and the Longfin Smelt (LFS; *Spirinchus thaleichthys*), fish in the Osmeridae family native to the San Francisco Estuary (SFE). Both species have experienced steep declines since the 1980s within the SFE and are part of the larger pelagic organism decline (Sommer et al. 2007; Mac Nally et al. 2010). DSM were federally-listed as threatened in 1993 and state-listed as endangered in 2010 (USFWS 1993; Newman 2008; CDFW 2017). Meanwhile, LFS were state-listed as threatened in 2009 (CDFG 2009). Declines to these native osmerids are most notably attributed to habitat modifications (e.g. extensive hydrological engineering of the Bay Delta for water extraction), entrainment in the pumps used for water diversions, and competition or predation resulting from changes in the estuary's biota (Moyle et al. 2016). Indeed, the SFE is one of the most highly invaded ecosystems in the world (> 200 exotic species; Cohen and Carlton 1998) and is the nexus for state and federal water deliveries to agriculture and cities in California. Consequently, listed fish are subject to some of the most intensive monitoring efforts in the world so that water deliveries are balanced with species protection.

Despite the significant investment and interest in the status of osmerids in the SFE, current monitoring efforts for DSM and LFS rely on morphological identifications, which is problematic for multiple reasons. Fish caught in traps are often found dead and decomposition can make identification based on morphological features difficult. Additionally, the methods used to capture fish can damage specimens, removing distinctive features needed for species assignment. Another significant confounding factor is the co-occurrence of a closely related introduced fish: the Wakasagi (WKS; *Hypomesus nipponensis*). The California Department of Fish and Wildlife introduced WKS from Japan to California reservoirs in 1959 as forage for salmonids but by the mid-1990s WKS had escaped reservoirs and established reproductive populations in the SFE (Moyle 2002). Field differentiation of DSM and WKS is notably difficult for juvenile fish (Wang et al. 2005) but correct species identification is essential to accurate reporting of each species range and abundance. Furthermore, prior genetic studies have documented hybridization of DSM with both LFS and WKS (May 1996; Trenham et al. 1998; Fisch et al. 2014). Although the extent and impacts of SFE osmerid hybridization are undetermined, hybridization can waste reproductive efforts and, in some cases, compromise the genetic integrity of entire populations (Allendorf et al. 2001). Because the frequency of hybridization can change rapidly and is ultimately irreversible in a system, it is important that SFE monitoring efforts reliably identify not only the pure species but also the hybrid progeny of DSM, LFS, and WKS.

One location within the SFE of interest for monitoring efforts in recent years is the Yolo Bypass—a flood plain of the Lower Sacramento River where the ranges of DSM, LFS, and WKS overlap (Fig. 1). Though surveys have not detected large numbers of WKS throughout the SFE, they have been detected with increasing frequency in recent years in the Yolo Bypass (Frantzich et al. 2013a, b; Ikemiyagi et al. 2014, 2015; Mahardja et al. 2016). The consistent co-occurrence between DSM and WKS in the Yolo Bypass is of particular concern given the two species' similarity in appearance (Moyle 2002) and past detections of hybrids in the region (Fisch et al. 2014).

Prior studies have developed genetic markers for identifying DSM, WKS, and LFS and their hybrids, but these markers have been either inefficient or unreliable (May 1996; Trenham et al. 1998; Fisch et al. 2014). With high-throughput sequencing, it is possible to develop large numbers of diagnostic markers for multiple uses, including species identification and hybrid detection (Twyford and Ennos 2012). Restriction site-associated DNA sequencing (RADseq) is particularly effective for identifying markers in non-model organisms and has been used successfully for hybridization studies in other species (Hohenlohe et al. 2011; Amish et al. 2012; Pujolar et al. 2014). Here we use RADseq to design species-specific single nucleotide polymorphism (SNP) assays for DSM, LFS, and WKS. We use these assays to detect and quantify levels of hybridization and species misidentification in a field sample from the Yolo Bypass.

Materials and methods

SNP discovery

Genomic DNA from all samples included in this experiment was extracted from fin clips using the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, California) in accordance with the manufacturer's protocols. We prepared RADseq libraries using the Sbf1 restriction enzyme, including DNA from DSM, WKS, and LFS (Table 1) according to the RAD protocol described in Ali et al. (2015). Genomic data was generated by paired-end, 100 base pair (bp) sequencing on the Illumina HiSeq platform. Sequences were sorted into individuals using unique 8 bp barcodes and matching read pairs (forward and reverse) of each individual were written into separate files. A de novo set of reference RAD-contigs was built using sequences from LFS individuals following the custom procedure given in Saglam et al. (2016). This

Table 1 Species, collection location and year, and number of individual fish (N) used for RADseq libraries or assay validation

Species	Collection location	Collection year	N (RADseq/validation)
DSM	San Francisco Estuary, CA	1996–1999 2013–2016	110/56
DSM × LFS F_1 hybrid	UC Davis FCCL, CA	2017	0/32
LFS	Lake Washington, WA	2008	20/0
LFS	Columbia River, OR/WA	2011–2016	21/0
LFS	Yakutat Bay, AK	2013	20/32
LFS	Pitt Lake, Canada	2006	20/0
LFS	Harrison Lake, Canada	2000	13/0
LFS	Skeena River, Canada	2011	10/0
LFS	Humboldt Bay, Canada	2012–2016	8/0
LFS	San Francisco Estuary, CA	2007–2015	76/26
WKS	San Francisco Estuary, CA	2010–2016	11/52

procedure produced 23,525 unique RAD-contigs ranging from 250 to 800 bp which served as a reference set of RAD-contigs for alignment and subsequent downstream analysis (for more details see Online Resource 1).

Individuals from all populations (DSM, WKS, and LFS) were aligned to this set of reference RAD contigs using the BWA-MEM algorithm (Li and Durbin 2010; Li 2013), outputted as bam files and indexed after sorting for proper pairs and removing PCR duplicates. Mean individual read depth was around 4× and mean alignment success was around 68–70%. Raw read counts, number of raw aligned reads, and number of alignments after filtering for proper pairs and removing PCR duplicates are summarized in Online Resource 2. Potential paralogous RAD-loci were tagged using ngsParalogs (https://github.com/tplinderoth/ngsParalog) and removed from further analysis.

To determine the overall quality of our data set and to confirm that samples were accurately diagnosed before SNP assays we conducted a principal component analysis (PCA) of genetic structure using the NGSCOVAR module of NGSTOOLS (Fumagalli et al. 2013, 2014). Prior to PCA, we sub-sampled all individual BAM files down to 80,000 mapped reads to remove any bias that might arise from variation in read numbers, which have the potential to dominate PC1. Subsampled BAM files were only used in PCA and all other analyses were conducted using the full number of mapped reads for each individual. PCA results showed clear separation of all three species along PC1 and no misidentified individuals (Online Resource 3). We conclude that our genomic data are of high quality and can reliably distinguish between species hence, suitable for SNP assay discovery.

To locate diagnostic sites for SNP assays, we genotyped each species separately. Estimate of per site minor allele frequencies (MAF), genotype probabilities and SNP discovery were conducted in ANGSD (Korneliussen et al. 2014) and sites were designated polymorphic only if MAFs were over 0.05 and the probability of the site not being polymorphic was less than 10–12. We filtered out any site that was not present in at least half of the individuals and did not meet a minimum phred quality score of 20 and mapping quality of 10. Major and minor alleles were inferred from genotype likelihoods using the method described in Nielsen et al. (2012). Genotypes were called using a posterior cutoff of 80% and outputted as a set of base pairs (AA, AC, AG, etc.) along with the major and minor allele at each site. Using the generated genotypes we identified sites where all three species were homozygous but where two of the three species were fixed for the same allele while the third were differentially fixed for an alternative allele (i.e. a species specific allele that identified one species of three; Online Resource 4). All such sites which had < 10% missing genotype calls were designated as candidate diagnostic sites for SNP assays. The set of diagnostic sites between each species were further filtered to contain only SNPs that were located at least 60 bp down and upstream of the ends of the contigs to enable optimum assay design. The same procedure was used to obtain candidate diagnostic sites for differentiating between DSM and WKS, DSM and LFS, and LFS and WKS.

Assay validation

Sequences containing SNPs were submitted to Fluidigm (Fluidigm Corporation, South San Francisco, CA) for SNP Type assay design and assays with high design rank were ordered for in vitro validation. For validation, we tested the assays on adult fish (56 DSM, 52 WKS, and 58 LFS) not included in the RAD library (Table 1). To test assays on hybrids, we used 32 F_1 DSM × LFS hybrids created by strip spawning a LFS mother with a DSM father at the UC Davis Fish Conservation and Culture Lab (FCCL). To account for the lack of availability of DSM × WKS and LFS × WKS F_1 hybrids, we replicated the Stephens et al. (2009) strategy of testing the assays on 'composite' hybrids (created by mixing equal ratios of genomic DNA from pairs of each parental type) and included two of each pair (DSM × WKS, DSM × LFS, and LFS × WKS) for analysis. We followed the recommended specific target amplification (STA) protocol from Fluidigm to pre-amplify regions containing SNP loci and diluted the STA product 1:100 in 2 mM Tris buffer. Assays and diluted STA products were loaded on 192.24 or 48.48 integrated fluidic circuits (IFCs) and run on the Fluidigm IFC Controller RX/MX and Biomark platforms according to the manufacturer's instructions. We determined genotypes using the Fluidigm SNP Genotyping Analysis

Software version 4.3.2 (Fluidigm Corporation, South San Francisco, CA). For our in vitro filtration of assays, we discarded assays that (1) failed to amplify both alleles in all three species, (2) had ambiguous plots, or (3) were non-species specific.

To validate our ability to detect backcrossed hybrids with our SNP panel we computationally simulated backcross hybrid genotypes using the three pure parental genotypes (DSM, LFS, WKS) and the three F_1 hybrid genotypes (DSM × WKS, DSM × LFS, LFS × WKS). To simulate backcross genotypes, we randomly selected one of the two alleles for each locus from the F_1 genotype and crossed them with the pure parental type allele. We repeated this selection for all SNP loci and hybrid types in order to generate 500 genotypes for each of the six first generation backcross types.

We assigned genetic identifications to all groups (pure references, F_1 hybrids, 'composite' hybrids, and simulated backcrosses) by first comparing each individual genotype to those expected for pure DSM, WKS, and LFS. Each fish whose genotype that did not match a pure genotype was considered a hybrid. Next we determined the genetic species composition of putative hybrids using STRUCTURE version 2.3.3 (Pritchard et al. 2000) which applies a Bayesian approach to assign individuals to clusters. We selected the model assuming admixture and ran 10 iterations at $K = 3$ with 10^6 MCMC reps and a 10^5 burn-in period. For each individual (i), STRUCTURE outputs q_i values, which range from 0 to 1 and represent the estimated membership coefficient for a particular individual in a given cluster. We compiled the STRUCTURE outputs for the 10 iterations and tested for multimodality using the Greedy K algorithm in the software CLUMPAK (Kopelman et al. 2015). Hybrid class identifications were assigned for each unknown individual by comparing q_i values to those of known reference samples. Individuals were identified as backcrosses when their q_i values were in the range between the q_i values observed for pure and F_1 references. We prepared graphical representations of STUCTURE bar plots using DISTRUCT version 1.1 (Rosenberg 2004).

Genetic identification of Yolo Bypass smelts

Following validation, we applied the SNP panel to 384 wild smelt collected in the Yolo Bypass (Fig. 1) from 2010 to 2016 by the Yolo Bypass Fish Monitoring Program (YBFMP) conducted by the California Department of Water Resources (DWR; Frantzich et al. 2013a, b; Ikemiyagi et al. 2014, 2015; Mahardja et al. 2016). DWR personnel assigned species identifications to the wild smelts—a combination of juveniles and adults ranging in size from 26 to 99 mm fork length (FL)—using morphological criteria described in Moyle (2002) and Wang et al. (2005). In addition, we used the SNP panel on two osmerids collected in the lower Sacramento River near Sherman Island (Fig. 1) suspected to be hybrids based on morphological features and length-at-date (95 and 92 mm FL). Negative and positive controls were included on each IFC.

For the hybrids that were detected, we wanted to determine if hybridization was sex-specific. To determine the maternal species of the genetically identified hybrid fish, we amplified and sequenced the mitochondrial gene CO1 using universal fish primers (VF2_t1, FR1d_t1; Ivanova et al. 2007). Consensus sequences were obtained using Sequencher version 4.8 (Gene Codes Corporation), and species identifications were assigned using the NCBI nucleotide basic local alignment search tool (BLAST).

Results

SNP panel design and validation

We discovered a total of 21 differentially fixed (i.e. diagnostic) sites to distinguish LFS from WKS and DSM and an additional 50, 30 and 6 differentially fixed sites between DSM and WKS, DSM and LFS, and LFS and WKS, respectively (Online Resource 4). A subset of 58 assays that passed in silico filtration were designed into assays. Of these, 24 assays passed in vitro filtration and were included in the final assay panel for use in all further genotyping and genetic assignment (Table 2).

Three clusters were identified from the STRUCTURE analysis, which corresponded with reference specimens for each species. All pure smelt validation samples were assigned to the expected clusters with a threshold of $q_i \geq 0.980$ and all genotypes matched those expected for each respective pure species, except for one WKS specimen (hybrid #13; Fig. 2b). For this individual, we observed DSM alleles present at five of 17 loci diagnostic for DSM and WKS. We re-inspected the genotype plots for these loci, confirming that amplification was clear and DSM allele calls were not attributable to erroneous scoring. Based on STRUCTURE analysis, this individual genetically assigned to both the WKS and DSM clusters (DSM $q_i = 0.150$, WKS $q_i = 0.838$, LFS $q_i = 0.012$; Fig. 2b) and had mitochondrial DNA (mtDNA) consistent with WKS (Table 3). Together, these results suggest that this fish is likely a WKS × (DSM × WKS) backcross.

The DSM × WKS hybrids ('composite') and DSM × LFS hybrids ('composite' and F_1) had the expected genotypes with STRUCTURE assigning each hybrid to the correct two parental clusters (Fig. 2a). The LFS × WKS 'composite' hybrids had the expected genotypes at all but two loci (LFS_03 and LFS_07). At these loci, heterozygous genotypes were expected but homozygous LFS genotypes were

Table 2 Twenty-four diagnostic SNP assays for DSM, LFS, and WKS species identification and hybrid detection

Assay Name	Locus	Fluidigm ID	DSM	WKS	LFS	SNP Type Primer Sequences (5′-3′)
DSM_01	R003880	GTA0176508	CC	TT	TT	ASP1: ACAGTTTTAGATTCTGGCACTCAGAAA
						ASP2: CAGTTTTAGATTCTGGCACTCAGAAG
						LSP: CCTTGGCAGAGCCAATTATAGCA
						STA: GAACCGACAAAAAAGAAGGAAAACA
DSM_02	R004792	GTA0176510	TT	CC	CC	ASP1: GCCTCAGGAAACATCATTAAGTGC
						ASP2: AGCCTCAGGAAACATCATTAAGTGT
						LSP: TGTTACTGAAAACCAAAAGAAACAAAACCGA
						STA: AAACGAAAACACTGGAAATAGCCT
DSM_03	R014607	GTA0176504	TT	CC	CC	ASP1: GTTCTCCGCTTGTTATTGGCTC
						ASP2: GTTCTCCGCTTGTTATTGGCTT
						LSP: AGACTGGCACCAGGAGAACG
						STA: GAATGCTCCACGGACACTT
DSM_04	R001236	GTA0176532	TT	CC	CC	ASP1: CCATGGAGAAGACGACCTGATTT
						ASP2: CCATGGAGAAGACGACCTGATTC
						LSP: GGCTGACACTGCACAACCA
						STA: CAACAGTTCACATCTAATCACCTCC
DSM_05	R003880	GTA0176563	CC	TT	TT	ASP1: CAGTTTTAGATTCTGGCACTCAGAAG
						ASP2: ACAGTTTTAGATTCTGGCACTCAGAAA
						LSP: CCTTGGCAGAGCCAATTATAGCA
						STA: GAACCGACAAAAAAGAAGGAAAACA
DSM_06	R007633	GTA0176548	AA	GG	GG	ASP1: GGGCTTTCATCAAAGGACCAACTA
						ASP2: GGCTTTCATCAAAGGACCAACTG
						LSP: GTATTTCACCTCAGCCAGGGC
						STA: CAGGGATGTTGTGAAACAGCT
DSM_07	R013671	GTA0176527	GG	TT	TT	ASP1: CTCGTGAAATGAAGAACAACACAGAC
						ASP2: CTCGTGAAATGAAGAACAACACAGAA
						LSP: ATGCAGTTGAAGTATCATTCTATAGTGGCT
						STA: CCATTTTAGCATTAGTACGAGGGAT
DSM_08	R015867	GTA0176538	GG	AA	AA	ASP1: TTATTCACACACACTGTGCATGC
						ASP2: GTTATTCACACACACTGTGCATGT
						LSP: CCTCTCGATCCTGGAGTAGCC
						STA: TTTCTGTTCACACTGCAGTGTT
DSM_09	R021884	GTA0176526	CC	AA	AA	ASP1: CCTAGCCCCTGCAAGAACG
						ASP2: CCTAGCCCCTGCAAGAACT
						LSP: CCCGCCCAGTTTGCTGT
						STA: CACGGCCCTCCACAAG
DSM_10	R021965	GTA0176552	AA	GG	GG	ASP1: CTGAAGATGATGCTCCTGTCTCTTT
						ASP2: TGAAGATGATGCTCCTGTCTCTTC
						LSP: AGGAACCTTTTTAGTCAGGCGGA
						STA: GCCACCATCGTCACCC
WKS_01	R001177	GTA0176539	AA	GG	AA	ASP1: CCATCCCATTCACTCAGACACA
						ASP2: CCATCCCATTCACTCAGACACG
						LSP: TGCCCTGGAAGAGGCACA
						STA: CCTATCCACCCTCCCTCC
WKS_02	R002608	GTA0176567	TT	CC	TT	ASP1: CCCTCCCAGATGCCAAACTA
						ASP2: CCCTCCCAGATGCCAAACTG
						LSP: CAGCAAGCAGAGGGTCCAG
						STA: ACAGCCTCACGTTTCAATTTCTC
WKS_03	R006717	GTA0176542	TT	GG	TT	ASP1: GATGTGGTGCCAGAAGCAA
						ASP2: GATGTGGTGCCAGAAGCAC

Table 2 (continued)

Assay Name	Locus	Fluidigm ID	DSM	WKS	LFS	SNP Type Primer Sequences (5′-3′)
						LSP: CGGTAGATCGCCAGGGCT
						STA: GGCATGTCTCCTTGGAAACG
WKS_04	R009220	GTA0176544	TT	AA	TT	ASP1: ATTACCTGGCCAACGTCACA
						ASP2: ATTACCTGGCCAACGTCACT
						LSP: CCCACGGACGACAGGTTG
						STA: CTAAGATCGAGAAACCCCTGGA
WKS_05	R010828	GTA0176553	AA	GG	AA	ASP1: CACACACCCCGACATAAGACA
						ASP2: ACACACCCCGACATAAGACG
						LSP: CCCACTCACGCAGTCAGTCT
						STA: CTGAGAGCACACACACACA
WKS_06	R020283	GTA0176530	GG	AA	GG	ASP1: GATGCTGTAACAGCAGCAAGTC
						ASP2: GATGCTGTAACAGCAGCAAGTT
						LSP: CACGCCAAACCGCCCC
						STA: GGCAGTAAACATCTCCCCATG
WKS_07	R020770	GTA0176554	AA	TT	AA	ASP1: GAGCTTGTTCCCTCTGTTCTGAT
						ASP2: AGCTTGTTCCCTCTGTTCTGAA
						LSP: CATGCCCATCCAACCATCCAA
						STA: AACAGGTACAGTAGGAAACAGGA
LFS_01	R000068	GTA0176495	CC	CC	AA	ASP1: GCACATCTATGACCAAACACACAA
						ASP2: GCACATCTATGACCAAACACACAC
						LSP: TGGTGGCAGGCTGGATAAACA
						STA: TCAGAAGGACCTGCACATCTAT
LFS_02	R004739	GTA0176513	GG	GG	CC	ASP1: ATGGAGGACAGTTTCACCCTC
						ASP2: ATGGAGGACAGTTTCACCCTG
						LSP: CCATCACCTTCCTCTGGAGACT
						STA: CACGATGACACTGAGAACATGG
LFS_03	R008663	GTA0185505	AA	AA	TT	ASP1: GCTTCACCAACATCCTGCTT
						ASP2: CGCTTCACCAACATCCTGCTA
						LSP: CAACACCGTTTTGCAGATGGC
						STA: TGGACAAACTTCTGCAGCG
LFS_04	R010547	GTA0185502	GG	GG	CC	ASP1: GCTCGTAGATGGTCATACGGC
						ASP2: GCTCGTAGATGGTCATACGGG
						LSP: CTCCTGCTCTTTTGTGCCCA
						STA: CGACCCATCATGTTCTCCTG
LFS_05	R012703	GTA0185508	CC	CC	TT	ASP1: TCCCACAGATCCTTCTACAGGA
						ASP2: CCCACAGATCCTTCTACAGGG
						LSP: AGGTGCCAATAGGACTGAGACC
						STA: CGCTAGATAAGCACCGTCACA
LFS_06	R015722	GTA0185501	CC	CC	TT	ASP1: GCCTGTTAAGTACCTGGTAATCACT
						ASP2: GCCTGTTAAGTACCTGGTAATCACC
						LSP: CCAGGAGCCCCAGCTGT
						STA: AGCCTGTTAAGTACCTGGTAATCA
LFS_07	R028969	GTA0185498	CC	CC	TT	ASP1: GTTGACAGGGGTCATACTTACAGTA
						ASP2: TTGACAGGGGTCATACTTACAGTG
						LSP: CACCTCACCAGTCACCGGA
						STA: ACTTGGTTGCCTTTCCTCATG

Assay names, RAD loci, Fluidigm assay IDs, expected genotypes, and Fluidigm SNP Type allele specific (ASP1, ASP2), locus specific (LSP), and specific target amplification (STA) primer sequences are included

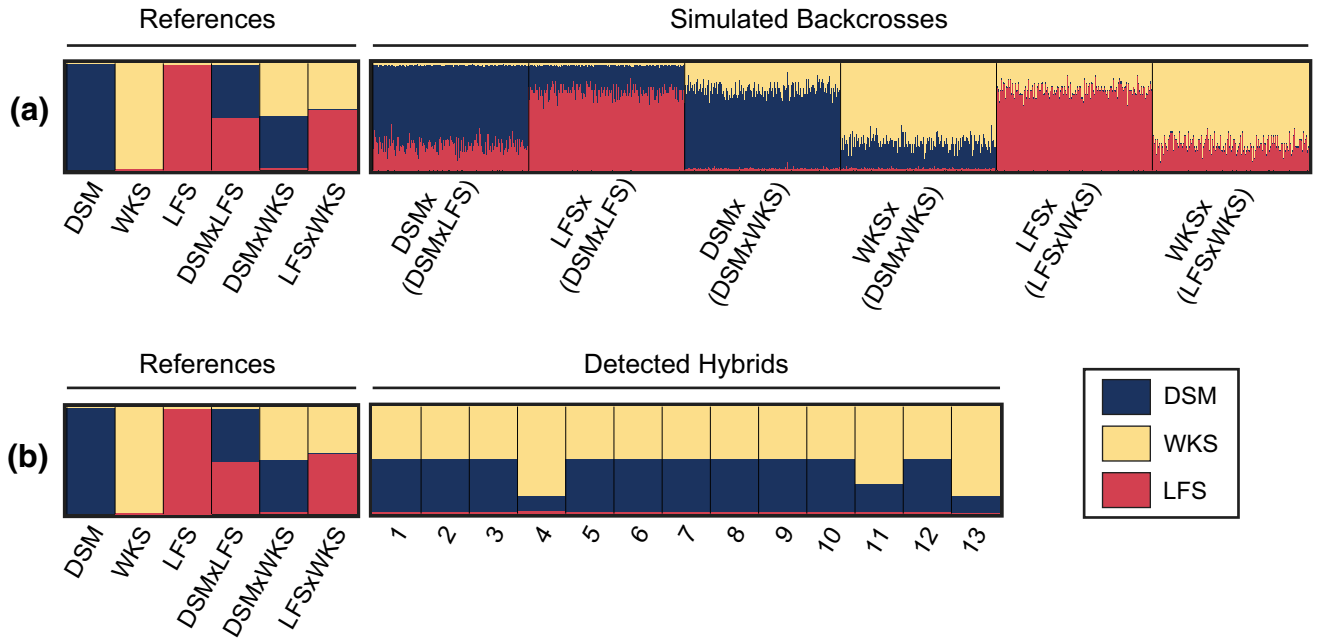


Fig. 2 Bar plots depicting compiled STRUCTURE q_i outputs (K = 3) for **a** simulated backcrosses and **b** hybrids detected using the assays. Each bar represents an individual and the height of each color represents the proportion of that individual's genotype assigned to a particular cluster. Representative pure and F_1 reference samples are included with each bar plot for comparison and are wider relative to simulated backcross results for visual clarity. We only included one representative individual for each type of pure and F_1 hybrid because STRUCTURE outputs were identical between individuals of the same type. Hybrids 1–11 are from the Yolo Bypass, hybrid 12 is the 92 mm fish from the lower Sacramento River and hybrid 13 is the WKS from the assay validation sample set that was identified as a backcross. (Color figure online)

Table 3 Year of capture and identification for hybrids based on SNP panel, mitochondrial DNA, and morphology

	Year	SNP ID	mtDNA	Morph. ID
1	2010	F_1-DSMxWKS	WKS	DSM
2	2013	F_1-DSMxWKS	WKS	DSM
3	2013	F_1-DSMxWKS	WKS	DSM
4	2013	BC-WKSx(DSMxWKS)	WKS	WKS
5	2014	F_1-DSMxWKS	WKS	DSM
6	2015	F_1-DSMxWKS	WKS	DSM
7	2015	F_1-DSMxWKS	WKS	DSM
8	2015	F_1-DSMxWKS	WKS	DSM
9	2015	F_1-DSMxWKS	WKS	DSM
10	2015	F_1-DSMxWKS	WKS	WKS
11	2016	BC-WKSx(DSMxWKS)	WKS	WKS
12	2016	F_1-DSMxWKS	WKS	DSM
13	2016	BC-WKSx(DSMxWKS)	WKS	WKS

Hybrids 1–11 are from the Yolo Bypass, hybrid 12 is the 92 mm fish from the lower Sacramento River, and hybrid 13 is the WKS from the assay validation sample set that was identified as a backcross

observed. For these two assays, we observed some minor separation of DSM from WKS in the genotype plots—WKS had slightly reduced endpoint fluorescence—despite having the same genotypes, indicating a difference in amplification efficiency between the two species. Weaker amplification of WKS and the fact that these are not true hybrids may explain why LFS × WKS 'composite' hybrids did not have the expected genotypes at these two loci.

Simulated backcross fish were correctly identified with 99.7% accuracy (Fig. 2a). Of the 3000 analyzed simulated backcross genotypes, only nine were not assigned the correct identification. Four were simulated DSM × (DSM × WKS) backcrosses and four were DSM × (DSM × LFS) backcrosses, for which STRUCTURE identified DSM as the primary genetic component but failed to distinguish whether the secondary genetic contribution was from WKS or LFS. The other simulated genotype was a LFS × (DSM × LFS) backcross that was misidentified as a LFS × (LFS × WKS) backcross.

Identification of smelts from Yolo Bypass

Of the 384 Yolo Bypass fish we genetically analyzed, two fish amplified poorly and could not be genetically identified. Of the 382 remaining Yolo Bypass fish, 125 individuals (32.7%) were misidentified, i.e., their genetic identification did not match the field identification (Table 4). All but five of these misidentified fish were thought to be DSM, but were genetically identified as WKS (82 individuals),

Table 4 The total number of DSM, LFS, and WKS detections in the Yolo Bypass based on SNP panel followed by the original morphological identifications

Species Genetic ID	N Morph. assigned		
	DSM	WKS	LFS
DSM (187)	**185**	2	–
WKS (146)	82	**64**	–
LFS (38)	30	–	**8**
Hybrid (11)	8	3	–

Bold indicates morphological and genetic identifications match. e.g. 82 fish morphologically identified as DSM were WKS

LFS (30 individuals), or hybrids (eight individuals). The other five misidentified fish were morphologically identified as WKS but were genetically identified as DSM (two individuals) or hybrids (three individuals). Of the 11 hybrids, nine were DSM × WKS F_1 hybrids and two were WKS × (DSM × WKS) backcrosses (hybrids #1–11; Fig. 2b; Table 3). We report the SNP panel-based numbers of Yolo Bypass DSM, WKS, LFS, and hybrids separated by collection year and year-class in Table 5.

For the two large fish collected in the lower Sacramento River, one was genetically identified as a pure DSM (95 mm) and the other a DSM × WKS F_1 hybrid (92 mm; hybrid #12; Fig. 2b; Table 3). MtDNA from all hybrids indicated that WKS was the maternal parent.

Discussion

Here we described the discovery and development of a SNP panel from high throughput sequencing data for application on non-model organisms of conservation concern. The SNP panel we developed for identifying DSM, LFS and WKS provided 100% correct assignment of pure and F_1 hybrid fish and 99.7% assignment for simulated backcrosses, suggesting that this assay panel can reliably detect hybrids and confirm species for three osmerids in the SFE. However, we experienced two unexpected results during assay validation. First, we identified one individual in the WKS pure adult reference samples that is likely a backcrossed WKS by DSM [WKS × (DSM × WKS)]. Though adult DSM and WKS are morphologically distinguishable, backcrossed WKS would probably be difficult to morphologically distinguish from pure WKS. The fact that DSM alleles were found at multiple diagnostic loci suggests that deviation from the expected genotype is likely due to backcrossing rather than genotyping error or low-level polymorphisms at the loci.

Second, both LFS × WKS 'composite' hybrids were scored as homozygous at two loci (LFS_03 and LFS_07) rather than heterozygous, even though the "parents" were assumed pure and genotyped as expected. We surmise that this occurred due to a combination of two factors—reduced amplification efficiency of WKS for these assays and the imperfection of using mixed DNA samples rather than true LFS × WKS F_1 hybrids. However, we kept these two assays in the panel because we found so few assays distinguishing LFS from the other two osmerids and these two assays are still useful for distinguishing pure species and DSM × LFS F_1 hybrids. Further, no evidence to date indicates that LFS × WKS hybrids exist, suggesting that the unexpected genotype results for these two loci are likely to be inconsequential for most data sets. Still, amplification of LFS × WKS F_1 hybrids using these two assays should be validated with true hybrids.

Table 5 Numbers of DSM, WKS, LFS, and hybrids detected in the Yolo Bypass based on SNP panel

Collection Year	DSM		WKS		LFS		F_1-DSM × WKS		BC-WKS × (DSM × WKS)		% Hybridization	
	YC	A	YC	A	YC	A	YC	A	YC	A	YC	A
2010	2	4	1	0	0	0	1	0	0	0	25	0
2011	1	3	11	0	0	0	0	0	0	0	0	0
2012	6	29	24	1	4	0	0	0	0	0	0	0
2013	15	39	29	2	32	0	2	0	1	0	6	0
2014	5	26	11	2	2	0	1	0	0	0	6	0
2015	36	10	19	0	0	0	5	0	0	0	8	0
2016	6	5	46	0	0	0	0	0	1	0	2	0
Total	71	116	141	5	38	0	9	0	2	0	5	0

Counts are broken down by collection year and further separated into the numbers of year-class fish produced in that collection calendar year (YC) and adult fish born in prior years (A). Hybridization percentages were calculated based on the number of hybrid fish divided by the number of DSM, WKS, and hybrids

Species identification

Correct identification of osmerids in the SFE is critical for both conservation efforts and water operations. Government agencies conduct extensive annual surveys in the SFE to determine fish abundance indices (Finstad and Baxter 2016); survey data informs policy decisions that affect not only the ecosystem's organisms but also farmers, industry, urban development, water pumping projects, commercial fisheries, and recreation. The Yolo Bypass region that the YBFMP operates appeared to have increased incidence of osmerid hybridization relative to other regions in the SFE (Fisch et al. 2014) and a substantial amount of the juvenile osmerids collected by the program may have been decaying up to 24 h at the time of sampling due to high (~ 50%) mortality associated with sampling via rotary screw trap. For these reasons, the relatively high frequency of osmerid species misidentification may be specific to the YBFMP, which is not one of the long-term monitoring surveys used for U.S. Endangered Species Act considerations. Nonetheless, our results suggest that the accuracy of osmerid species identification in other SFE monitoring programs may need to be evaluated using genetic species identification.

Overall, we found that nearly one-third of the osmerids collected from the Yolo Bypass for this study were misidentified and that juvenile WKS were the most commonly misidentified individuals. We observed a bimodal distribution in fork length of Yolo Bypass fish (corresponding with juvenile and adult fish) and over 95% of the fish that were misidentified had fork lengths less than or equal to 50 mm (Fig. 3). It is not surprising that juvenile fish are more difficult to identify in the field—DSM, WKS, and LFS are all morphologically similar as juveniles. Notably, we found that a majority of the misidentifications (66%) were WKS mistakenly identified as DSM. In the field, identifications between DSM and WKS were primarily based on isthmus pigmentation (typically only zero or one chromatophores in DSM, many chromatophores in WKS) and the pattern of pigmentation at the upper base of the caudal fin ("V" shaped pigmentation in DSM, random scattered pigmentation in WKS; Wang et al. 2005). However, Wang and Hess (2000) report ecophenotypic variation between lake/reservoir WKS and estuary WKS; the estuary WKS have much lighter pigmentation, with some individuals lacking isthmus pigmentation (like DSM). Additionally, WKS in the estuary only start to develop isthmus pigmentation when they reach approximately 25 to 30 mm (Wang and Hess 2000) and appear to develop more chromatophores as they grow larger (Sweetnam 1995; Aasen et al. 1998). Thus, juvenile WKS in the estuary may have fewer chromatophores than expected, explaining their misidentifications as DSM. Variation in pigmentation is likely one cause for morphological misidentifications so we caution against using pigmentation as the primary feature for species identification.

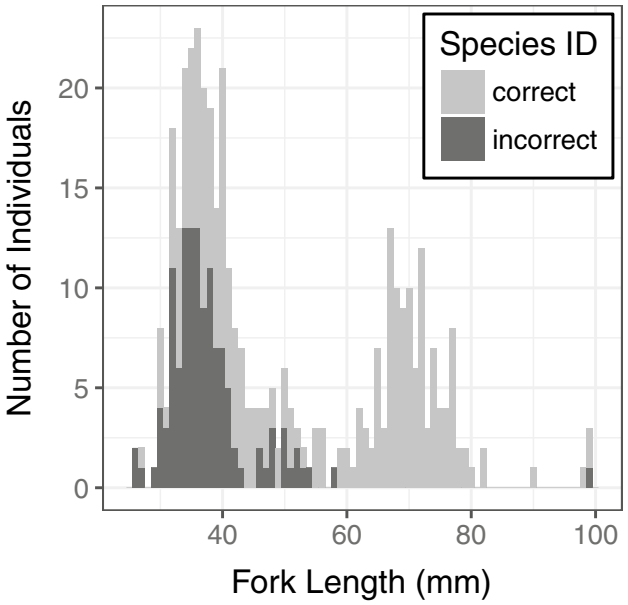


Fig. 3 Histogram showing fork length in millimeters and the number of individuals from the Yolo Bypass that were either correctly or incorrectly morphologically identified when compared to genetic species identifications

The considerable misidentification frequency of WKS as DSM is important beyond merely exposing morphological identification errors—it reveals higher proportions of WKS to DSM in the Yolo Bypass than previously recorded. This rise is troubling given WKS' likely competition with DSM for food and spawning sites, possible predation on DSM larvae, and their capacity for hybridization (Moyle 2002).

Hybridization

Hybridization presents a challenge to the management of legally protected endangered species, particularly when it stems from anthropogenic influences (Wayne and Shaffer 2016). Before selecting a management approach, it is necessary to understand the extent and direction of admixture in a system (Allendorf et al. 2001). We found evidence of backcrossing of DSM and WKS in contrast to prior analyses using allozyme and SNP markers (May 1996; Trenham et al. 1998; Fisch et al. 2014), which suggested that F_1 hybrids are infertile. Though the microsatellite approach in Fisch et al. (2014) identified 18 backcrossed individuals, the prevalence of backcrossing may have been overestimated due to the non-diagnostic nature of microsatellites. We are more confident that we detected true backcrosses in this study because we addressed the problems afflicting the prior genetic approaches by (1) using bi-allelic diagnostic markers and (2) using more markers in our SNP panel. In total, we identified 10 DSM × WKS

F_1 hybrids and three WKS × (DSM × WKS) backcrossed individuals, suggesting the occurrence of unidirectional introgressive gene flow toward WKS. Furthermore, mtDNA sequences for all 10 of the DSM × WKS F_1 hybrids identified WKS as the maternal parent, indicating that interspecific breeding resulting in F_1 progeny may only occur between DSM males and WKS females. An experiment from Wang et al. (2007) provides supporting evidence for sex-specific hybridization: when researchers crossed a female DSM with a male WKS by artificial fertilization, all larvae died before reaching first feeding stage.

Unlike prior studies (May 1996; Fisch et al. 2014), we did not find any DSM × LFS hybrids. However, we focused on fish from the Yolo Bypass whereas Fisch et al. (2014) and May (1996) detected DSM × LFS F_1 hybrids in Montezuma Slough and near Chipps Island—locations in the SFE where LFS are more prevalent than they are in the Yolo Bypass (Wang 2007; Tempel 2016). Overall, we detected very few hybrids of any type and the apparent unidirectional gene flow suggests that introgression does not currently pose a risk to the genetic integrity of DSM in the Yolo Bypass. Nevertheless, the persistence of low level hybridization, even without gene flow, can be detrimental to native populations through competition, disease transmission, or wasted reproductive effort that can reduce genetic diversity and impair the viability of native fish (Laikre et al. 2010). Moreover, hybridization rates can vary over time and space so additional genetic monitoring efforts are needed to quantify hybridization rates in other areas of the SFE.

The SNP markers described in this study will provide a valuable resource for managers working on conservation of native osmerids in the SFE. They are more accurate for species identifications than using morphological features across all life stages, can be used to study hybridization, and offer a high throughput approach that is efficient and cost-effective. Using these assays as a complement to traditional survey methods in the SFE will help improve the monitoring efforts that guide conservation.

Acknowledgements We are grateful to Rene Reyes (U.S. Bureau of Reclamation), Luke Ellison (FCCL), Naoaki Ikemiyagi (DWR), Jared Frantzich (DWR), Brian Schreier (DWR) as well as past and present staff of the Yolo Bypass Fish Monitoring Program for collection of field specimens. We thank Luke Ellison and the FCCL crew for providing Delta Smelt by Longfin Smelt hybrids. We also thank Bernie May and three anonymous reviewers for helpful comments that improved the manuscript. This research was funded by the California Department of Water Resources (Contract #4600011196).

References

Aasen GA, Sweetnam DA, Lynch LM (1998) Establishment of the Wakasagi, *Hypomesus nipponensis*, in the Sacramento-San Joaquin Estuary. Calif Fish Game 84(1):31–35

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR (2015) RAD capture (Rapture): flexible and efficient sequence-based genotyping. Genetics 202:389–400. https://doi.org/10.1534/genetics.115.183665

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16(11):613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Amish SJ, Hohenlohe PA, Painter SA, Leary RF, Muhlfeld C, Allendorf FW, Luikart G (2012) RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays. Mol Ecol Resour 12:653–660. https://doi.org/10.1111/j.1755-0998.2012.03157.x

Beerkircher L, Arocha F, Barse A, Prince E, Restrepo V, Serafy J, Shivji M (2009) Effects of species misidentification on population assessment of overfished white marlin *Tetrapturus albidus* and roundscale spearfish *T. georgii*. Endanger Species Res 9:81–90. https://doi.org/10.3354/esr00234

CDFG (California Department of Fish and Game) (2009) A status review of the longfin smelt (Spirinchus thaleichthys) in California. Report to the fish and game commission

CDFW (California Department of Fish and Wildlife) (2017) State and federally listed endangered and threatened animals of California. California Department of Fish and Wildlife, The Natural Resources Agency, North Highlands

Cohen AN, Carlton JT (1998) Accelerating invasion rate in a highly invaded esturary. Science 279:555–557. https://doi.org/10.1126/science.279.5350.555

Elphick CS (2008) How you count counts: the importance of methods research in applied ecology. J Appl Ecol 45:1313–1320. https://doi.org/10.1111/j.1365-2664.2008.01545.x

Finstad S, Baxter R (2016) 2015 Status and trends report for pelagic fishes of the upper San Francisco Estuary. Interag Ecol Program Newsl 29(2):12–27

Fisch KM, Mahardja B, Burton RS, May B (2014) Hybridization between DSM and two other species within the family Osmeridae in the San Francisco SFE. Conserv Genet 15:489–494. https://doi.org/10.1007/s10592-013-0555-y

Frantzich J, Ikemiyagi N, Conrad JL (2013a) 2010–2011 Yolo Bypass fisheries monitoring status and trends report. Interag Ecol Program Newsl 26(1):45–52

Frantzich J, Rojas L, Ikemiyagi N, Conrad JL (2013b) 2011–2012 Yolo Bypass fisheries monitoring status and trends report. Interag Ecol Program Newsl 26(3):16–24

Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sanchez E, Albrechtsen A, Nielsen R (2013) Quantifying population genetic differentiation from next-generation sequencing data. Genetics 195(3):979–992. https://doi.org/10.1534/genetics.113.154740

Fumagalli M, Vieira FG, Linderoth T, Nielsen R (2014) ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics 30(10):1486–1487. https://doi.org/10.1093/bioinformatics/btu041

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. Mol Ecol Resour 11:117–122. https://doi.org/10.1111/j.1755-0998.2010.02967.x

Hull JM, Fish AM, Keane JT, Mori SR, Sacks BN, Hull AC (2010) Estimation of species identification error: implications for raptor migration counts and trend estimation. J Wildl Manag 74(6):1326–1334. https://doi.org/10.2193/2009-255

Ikemiyagi N, Carlson H, Frantzich J, Schreier B (2014) 2012–2013 Yolo Bypass fisheries monitoring status and trends report. Interag Ecol Program Newsl 27(1):29–38

Ikemiyagi N, Tung A, Frantzich J, Mahardja B, Schreier B (2015) 2013–2014 Yolo Bypass fisheries monitoring status and trends report. Interag Ecol Program Newsl 28(2):16–24

Ivanova NV, Zemlak TS, Hanner R, Hebert PDN (2007) Universal primer cockails for fish DNA barcoding. Mol Ecol Notes 7:544–548. https://doi.org/10.1111/j.1471-8286.2007.01748.x

Kellner K, Swihart (2014) Accounting for imperfect detection in ecology: a quantitative review. PLoS One 9(10):1–8. https://doi.org/10.1371/journal.pone.0111436

Kirsch J, Marshall M, Smith L (2014) Fish identification accuracy and implications to monitoring within the San Francisco Estuary, CA. Interag Ecol Program Newsl 27(2):37–43

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resourc 15(5):1179–1191. https://doi.org/10.1111/1755-0998.12387

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. BMC Bioinform 15:356. https://doi.org/10.1186/s12859-014-0356-4

Laikre L, Schwartz MK, Waples RS, Ryman N (2010) Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends Ecol Evol 25:520–529. https://doi.org/10.1016/j.tree.2010.06.013

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26(5):589–595. https://doi.org/10.1093/bioinformatics/btp698

Mac Nally R, Thomson JR, Kimmerer WJ, Feyrer F, Newman KB, Sih A, Bennett WA, Brown L, Fleishman E, Culberson SD, Castillo G (2010) Analysis of pelagic species decline in the upper San Francisco Estuary using multivariate autoregressive modeling (MAR). Ecol Appl 20(5):1417–1430. https://doi.org/10.1890/09-1724.1

Mahardja B, Ikemiyagi N, Farruggia MJ, Agundes J, Frantzich J, Schreier B (2016) 2014–2015 Yolo Bypass fisheries monitoring status and trends report. Interag Ecol Program Newsl 29(2):32–40

May B (1996) Identification of smelt species and their interspecific hybrids in the Sacramento-San Joaquin Estuary by allozyme analysis. Interag Ecol Program Newsl 9(3):9–10

Moyle PB (2002) Inland fishes of California. University of California Press, Berkeley

Moyle PB, Brown LR, Durand JR, Hobbs JA (2016) Delta Smelt: life history and decline of a once-abundant species in the San Francisco Estuary. San Franc Estuary Watershed Sci 14(2):1–30. https://doi.org/10.15447/sfews.2016v14iss2art6

Newman KB (2008) Sample design-based methodology for estimating Delta Smelt abundance. San Franc Estuary Watershed Sci 6(3):1–18

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS ONE 7(7):e37558. https://doi.org/10.1371/journal.pone.0037558

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Magnussen E, Jonsson B, Jiang X, Cheng L, Bekkevold D, Maes GE, Bernatchez L, Hansen MM (2014) Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. Heredity 112:627–637. https://doi.org/10.1038/hdy.2013.145

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Sağlam İK, Baumsteiger J, Smith MJ, Linares-Casenave J, Nicholas NL, O'Rourke SM, Miller MR (2016) Phylogenetics supports an ancient common origin of two scientific icons: Devils Hole and Devils Hole pupfish. Mol Ecol 25:3962–3973. https://doi.org/10.1111/mec.13732

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22(1):25–33. https://doi.org/10.1016/j.tree.2006.08.009

Shea CP, Peterson JT, Wisniewski JM, Johnson NA (2011) Misidentification of freshwater mussel species (Bivalvia:Unionidae): contributing factors, management implications, and potential solutions. Am Benthol Soc 30(2):446–458. https://doi.org/10.1899/10-073.1

Sommer T, Armor C, Baxter R, Breuer R, Brown L, Chotkowski M, Culberson S, Feyrer F, Gingras M, Herbold B, Kimmerer W, Mueller-Solger A, Nobriga M, Souza K (2007) The collapse of pelagic fishes in the upper San Francisco Estuary. Fisheries 32(6):270–277

Stephens MR, Clipperton NW, May B (2009) Subspecies-informative SNP assays for evaluating introgression between native golden trout and introduced rainbow trout. Mol Ecol Resour 9:339–343. https://doi.org/10.1111/j.1755-0998.2008.02407.x

Sweetnam D (1995) Field identification of Delta Smelt and Wakasagi. Interag Ecol Program Newsl 8(2):1–3

Tempel T (2016) Smelt larva survey summary. Interag Ecol Program Newsl 29(2):43–45

Trenham PC, Shaffer HB, Moyle PB (1998) Biochemical identification and assessment of population subdivision in morphologically similar native and invading smelt species(Hypomesus) in the Sacramento San Joaquin estuary, California. Trans Am Fish Soc 127:417–424

Twyford AD, Ennos RA (2012) Next-generation hybridization and introgression. Heredity 108:179–189. https://doi.org/10.1038/hdy.2011.68

USFWS (1993) Endangered and threatened wildlife and plants: determination of threatened status for the DSM. US Department of the Interior, Fish and Wildlife Service. Federal Register

Wang JCS (2007) Spawning, early life stages, and early life histories of the osmerids found in the Sacramento-San Joaquin Delta of California. Tracy fish facility studies California, vol 38. U.S. Bureau of Reclamation, Denver

Wang J, Hess L (2000) Similarities between hatchery reared Delta Smelt and wild Wakasagi from the Sacramento-San Joaquin Delta. Interag Ecol Program Newsl 13(1):49–51

Wang JCS, Lynch L, Bridges B, Grimaldo L (2005) Using morphometric characteristics to identify the early life stages of two sympatric osmerids (Delta Smelt and Wakasagi, Hypomesus transpacificus and H. *nipponensis*) in the Sacramento-San Joaquin Delta, California. Tracy Fish Facility Studies. Vol 30, U. S. Bureau of Reclamation, Mid-Pacific Region and Denver Technical Service Center

Wayne RK, Shaffer HB (2016) Hybridization and endangered species protection in the molecular era. Mol Ecol 25:2680–2689. https://doi.org/10.1111/mec.13642