

## ORIGINAL RESEARCH

# Discovery and characterization of single nucleotide polymorphisms in two anadromous alosine fishes of conservation concern

Diana S. Baetscher<sup>1,2</sup> | Daniel J. Hasselman<sup>3</sup> | Kerry Reid<sup>2,3</sup> | Eric P. Palkovacs<sup>3</sup> | John Carlos Garza<sup>1,2</sup> 

<sup>1</sup>Department of Ocean Sciences, University of California, Santa Cruz, CA, USA

<sup>2</sup>Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, CA, USA

<sup>3</sup>Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA

**Correspondence**

John Carlos Garza, Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, CA, USA.  
Email: carlos.garza@noaa.gov

**Present address**

Daniel J. Hasselman, Columbia River Inter-Tribal Fish Commission, Hagerman, ID, USA

**Funding information**

National Fish and Wildlife Federation, Grant/Award Number: 0104.14.041425; Atlantic States Marine Fisheries Commission, Grant/Award Number: 15-0105; The Nature Conservancy, Grant/Award Number: 0104.14.041425

**Abstract**

Freshwater habitat alteration and marine fisheries can affect anadromous fish species, and populations fluctuating in size elicit conservation concern and coordinated management. We describe the development and characterization of two sets of 96 single nucleotide polymorphism (SNP) assays for two species of anadromous alosine fishes, alewife and blueback herring (collectively known as river herring), that are native to the Atlantic coast of North America. We used data from high-throughput DNA sequencing to discover SNPs and then developed molecular genetic assays for genotyping sets of 96 individual loci in each species. The two sets of assays were validated with multiple populations that encompass both the geographic range and the known regional genetic stocks of both species. The SNP panels developed herein accurately resolved the genetic stock structure for alewife and blueback herring that was previously identified using microsatellites and assigned individuals to regional stock of origin with high accuracy. These genetic markers, which generate data that are easily shared and combined, will greatly facilitate ongoing conservation and management of river herring including genetic assignment of marine caught individuals to stock of origin.

**KEYWORDS**

SNPs, alewife, blueback herring, anadromous, conservation, fisheries management

## 1 | INTRODUCTION

Genetic data are routinely used to inform ecological investigation and formulate conservation and management plans for fish and wildlife. Elucidation of population structure and patterns of connectivity are often the first steps in the use of genetic data to understand a species' biology. In addition, common applications for such data include reconstructing pedigree relationships, inferring historical demography, individual identification for mark/recapture type analyses, and evaluating patterns of natural selection and the identification of individuals

to population of origin (Morin, Luikart, & Wayne, 2004; Narum et al., 2008).

Alewife (*Alosa pseudoharengus*) and blueback herring (*Alosa aestivalis*)—collectively known as “river herring”—are migratory sea-run (i.e., anadromous) fishes that reproduce in lakes and rivers along the east coast of North America, but typically migrate to the Atlantic Ocean as juveniles to grow and reach sexual maturity before returning to their natal freshwater spawning grounds to reproduce (Loesch, 1987). River herring once supported an important commercial fishery, but spawning adult abundances have declined by 93% since 1970, and many

spawning populations (hereafter “populations”) are now at historically low levels and are of increasing conservation concern (Hightower et al., 1996; Limburg & Waldman, 2009; Atlantic States Marine Fisheries Commission [ASMFC] 2012). Both species have broad distributions along the Atlantic coast (alewife: Labrador, Canada to North Carolina, USA; blueback herring: Gulf of St. Lawrence, Canada to Florida, USA), and resolving the range-wide spatial scale of population genetic structure is an important component of conservation efforts and fishery management plans.

Recent genetic studies of alewife and blueback herring used polymorphic microsatellite markers to resolve the spatial scale of population genetic structure (McBride, Willis, Bradford, & Bentzen, 2014; Palkovacs et al., 2014), examine range-wide patterns of hybridization (Hasselman et al., 2014), assess the influence of stocking activities on genetic structure (McBride, Hasselman, Willis, Palkovacs, & Bentzen, 2015), and determine the origin of river herring bycatch in commercial fisheries (Hasselman et al., 2016). Palkovacs et al. (2014) used data from microsatellites to reveal that US alewife populations ( $n = 21$ ) were nested within three regional genetic stocks (Northern New England, Southern New England, and Mid-Atlantic), whereas US blueback herring populations ( $n = 21$ ) were nested within four regional genetic stocks (Northern New England, Southern New England, Mid-Atlantic, and South Atlantic), with similar but not identical boundaries. Hasselman et al. (2016) also found that these same data had sufficient statistical power to confidently assign river herring bycatch in commercial fisheries to regional genetic stocks. Given their propensity for natal philopatry, the conservation and management of river herring requires a “population-level” approach, and there is a need for molecular tools that can resolve population genetic structure at spatial scales finer than regional genetic stock. Moreover, for anadromous fishes, such as river herring, that migrate substantial distances across jurisdictional boundaries and are subject to capture as bycatch in mixed-stock fisheries, a method that generates portable genetic data that can be easily shared and allows unambiguous assignment of individuals to population of origin is an important conservation and management tool (Clemento, Crandall, Garza, & Anderson, 2014; Morin et al., 2004; Starks, Clemento, & Garza, 2016).

Single nucleotide polymorphisms (SNPs) are bi-allelic markers, ubiquitous in the genome of most species (Morin et al., 2004), that are relatively simple to genotype and provide data that are easily portable between laboratories and instruments. Recent higher-throughput SNP genotyping technologies allow samples to be processed efficiently and in a cost-effective manner (Clemento, Abadía-Cardoso, Starks, & Garza, 2011; Larson, Seeb, Pascal, Templin, & Seeb, 2014; Seeb, Pascal, Ramakrishnan, & Seeb, 2009). SNP marker data have utility for a variety of ecological and evolutionary questions, and a suitable number of SNPs have been demonstrated to provide sufficient statistical power for resolving the spatial scale of population genetic structure in anadromous fishes (Clemento et al., 2014; Narum et al., 2008; Starks et al., 2016) and identifying pedigree relationships (Anderson & Garza, 2006). SNP data are also useful for assignment of individuals of unknown provenance to population of origin, often called genetic stock identification (GSI), and can be particularly informative when some of

those SNPs have been affected by divergent selection between populations (Ackerman et al., 2011; Nielsen et al., 2012).

Here, we describe the development of two sets of 96 SNP assays, one specific to alewife and the other to blueback herring. These SNP panels are suitable for resolving range-wide population genetic structure and have applications for GSI, investigating patterns of hybridization and introgression, and addressing issues of ecological and evolutionary relevance in a conservation and fisheries management framework. We used samples collected from across the ranges of both species for SNP discovery to minimize ascertainment bias (Albrechtsen, Nielsen, & Nielsen, 2010; Clark, Hubisz, Bustamante, Williamson, & Nielsen, 2005) and assess the power of the SNP data to accurately resolve previously described genetic stock structure for both species. The SNPs described herein will provide more power for population genetic investigations, enable higher throughput genotyping than with microsatellites, and allow for more effective data sharing across laboratories and management agencies.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection

Muscle plugs or fin tissue was obtained from alewife and blueback herring captured across the species' ranges (Table 1, Figure 1). All samples were obtained from adult fish and were collected in freshwater. Tissue was preserved in 95% ethanol until DNA extraction.

### 2.2 | SNP discovery and assay development

Tissue samples were removed from ethanol and air-dried before extracting genomic DNA using DNeasy 96 Blood and Tissue kits with a BioRobot 3000 (Qiagen, Inc.). To identify potential SNPs for alewife and blueback herring, we used double digest Restriction Associated DNA sequencing (ddRADseq), a genome-reduction technique that uses two restriction enzymes to create DNA fragments with identical fixed endpoints for annealing sequencing adapters (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). To ensure range-wide coverage and reduce the risk of ascertainment bias, samples chosen for SNP discovery included individuals from at least one population from each of the regional genetic stocks for alewife and blueback herring previously identified by Palkovacs et al. (2014).

We performed ddRADseq library construction, sequencing, and SNP identification separately for each species, following the same protocol. Undiluted genomic DNA from 48 alewife from 18 populations and 12 blueback herring from four populations (Table S1), representative of genetic lineages throughout the species' geographic ranges (McBride et al., 2014; Palkovacs et al., 2014) was digested using two restriction enzymes—Sph1 and EcoR1. Next, we performed size selection for 350-bp fragments using the Pippin Prep system (Sage Science, Inc.). Following the addition of adapters, sequencing was performed on a MiSeq instrument (Illumina Inc.). We used two 600-cycle sequencing reactions with paired-end reads for alewife

**TABLE 1** Summary statistics of SNP assays in validation populations of (a) alewife and (b) blueback herring. Sample size consists of samples included in analyses.  $H_E$  is unbiased, expected heterozygosity,  $H_O$  is observed heterozygosity, No. of alleles is the mean number of alleles per locus in that population. Mean minor allele freq. is the frequency of the minor allele in the Quinnipiac River for alewife and the Monument River for blueback herring

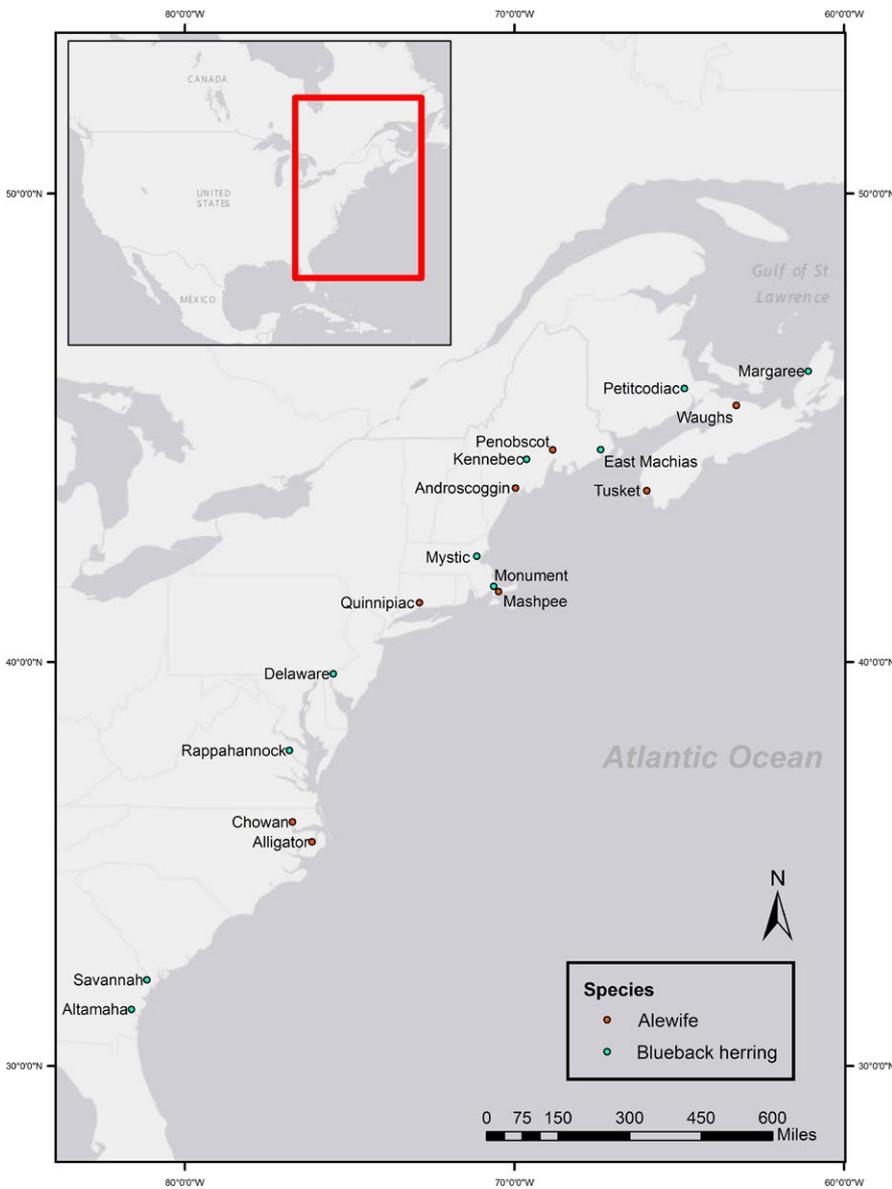
Population	Sample size	Loci typed	$H_E$	$H_O$	No. of alleles	Percent polymorphic loci	Mean minor allele freq.
(a)							
Waugh's	27	92	0.259	0.259	1.92	92.4	0.199
Tusket	45	92	0.265	0.265	1.95	94.6	0.206
Penobscot	44	93	0.267	0.264	1.96	95.7	0.213
Androscoggin	47	93	0.240	0.233	1.92	92.5	0.201
Mashpee	40	93	0.262	0.265	1.95	94.6	0.202
Quinnipiac	36	93	0.265	0.257	1.94	93.5	0.190
Chowan	41	93	0.216	0.223	1.88	88.2	0.180
Alligator	43	92	0.225	0.223	1.83	82.6	0.186
(b)							
Margaree	40	93	0.294	0.275	1.97	97.8	0.230
Petitcodiac	27	93	0.266	0.296	1.86	86.0	0.211
East Machias	45	94	0.298	0.295	1.98	97.9	0.238
Kennebec	47	96	0.301	0.305	1.99	97.9	0.240
Mystic	44	96	0.285	0.290	1.96	95.8	0.215
Monument	47	96	0.282	0.288	1.91	90.6	0.210
Delaware	47	95	0.290	0.290	1.97	96.8	0.224
Rappahannock	42	96	0.302	0.313	1.98	97.9	0.234
Savannah	47	96	0.275	0.267	1.95	94.8	0.251
Altamaha	47	96	0.277	0.281	1.95	94.8	0.247

and a single such sequencing reaction for blueback herring. Sequence data from each species were processed, with homologous reads identified and SNPs called, using Stacks (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Loci were selected for assay design by identifying sequences with a single SNP that also met the following three criteria: (1) all three genotypes (both homozygotes and the heterozygote) were observed, (2) a minimum of 20 sequence reads per allele for alewife and 15 reads for blueback herring were detected, and (3) the sequence did not share high similarity (>80%) with any other sequence selected from Stacks when global alignment was evaluated using BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990). These criteria were used to choose unique and sufficiently polymorphic target loci for the development of SNPtype genotyping assays (Fluidigm Corporation).

A total of 166 SNPs in alewife and 141 SNPs in blueback herring were chosen for SNPtype assay (Fluidigm) design. Assays were evaluated for consistency and polymorphism by genotyping 382 alewife samples from eight populations and 474 blueback herring from 10 populations throughout the species ranges (Figure 1; Table 1). SNP genotyping was performed with 96.96 Dynamic Genotyping Arrays on an EP1 Genotyping system (Fluidigm), which combines 96 DNA samples with 96 assays for a total of 9,216 reactions on each nanofluidic array. SNP genotypes were called with the Fluidigm SNP Genotyping software package.

Assays were first evaluated for their ability to produce clearly and consistently distinct clusters of genotypes. Loci were excluded that produced ambiguous genotypes or for which all validation samples appeared to have either homozygote or heterozygote genotypes, indicating null alleles or a lack of Mendelian inheritance. Sets of 96 well-performing assays were then retained for the final alewife and blueback herring SNP panels. Details of these SNP genotyping assays, including target polymorphism, primer/probe sequences, and database accession numbers are in Table S2 (alewife) and Table S3 (blueback herring). Additionally, as there were more than 96 high-quality blueback herring loci remaining at this stage, minor allele frequencies in the validation populations were estimated and used as a proxy for the expected power of the markers in pedigree reconstruction applications (Anderson & Garza, 2006) and used to choose the final panel of 96 assays.

Following genotyping, samples with missing data for 10 or more loci were excluded from further analyses (Table 1). Of the 382 alewife samples genotyped, 59 were excluded from validation analyses due to missing data. For the 474 blueback herring samples genotyped, the number excluded was 41. Allele frequencies, expected ( $H_E$ ) and observed heterozygosity ( $H_O$ ) for each locus were estimated using the Microsatellite Toolkit (v.3.1; Park, 2001). Concordance with expectations of Hardy-Weinberg and linkage (gametic phase) equilibria was determined with GenePop (v.4.2; Rousset, 2008) using an alpha value  $p = .05$ .



**FIGURE 1** Map of sampling locations for alewife and blueback herring. Sampling locations are indicated by diamonds with associated river names

To evaluate the utility and performance of the two sets of SNP assays for GSI, self-assignment analyses were conducted with GeneClass 2.0 (Piry et al., 2004), and the proportion of accurately assigned individuals per population and regional reporting group estimated. To evaluate utility for resolving range-wide population structure,  $F_{ST}$  values were estimated for all pairs of populations within each species using Genetix (v.4.05; Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme, 1996–2004). Model-based clustering was performed with *structure* (Pritchard, Stephens, & Donnelly, 2000) without prior location information and using the admixture and correlated allele frequencies model. Clustering was evaluated for hypothesized numbers of genetic groups,  $K = 2-6$ , with 10 iterations at each value of  $K$ . Plots were generated using CLUMPP (Jakobsson and Rosenberg, 2007) and DISTRICT (Rosenberg, 2004). Additionally, discriminate analysis of principal components (DAPC) (Jombart, 2008; Jombart, Devillard, & Balloux, 2010), implemented in the package *adegenet* (version 2.0.1) in R 3.3.1 (R Core Development Team 2016), was used to identify population structure.

### 3 | RESULTS

Sequencing yielded approximately 54 million reads in alewife and 33 million reads in blueback herring. After filtering with Stacks software (Catchen et al., 2013), 33,868 and 19,809 unique loci were found for alewife and blueback herring, respectively. Of these, 4,934 loci containing a single SNP were identified in alewife and 2,810 loci in blueback herring. From the loci selected for SNPTyping assay design, two unique sets (one for each species) containing 96 SNPs each were used for validation (Tables S2 and S3). Three loci (*Aps\_2060*, *Aps\_4755* and *Aps\_8787*) did not show consistent clustering in the alewife validation populations and were removed from further analyses.

Genetic variation of the validated SNPs was generally lower in alewife (Table S4) than in blueback herring (Table S5), with mean observed heterozygosity ( $H_O$ ) of 0.290 in blueback herring versus  $H_O$  of 0.249 in alewife. Similarly, the mean minor allele frequency across blueback herring populations was 0.230, whereas it was 0.197 in alewife. While

still adequate for parentage analyses (Anderson & Garza, 2006), diversity values are generally lower than in other anadromous species, such as Chinook salmon (e.g., Clemento et al., 2011).

There were 15 loci that deviated from HWE expectations in one alewife population, one locus that deviated in two populations (Aps\_14730), and one locus in four populations (Aps\_5844) (Table S4). For blueback herring, 21 loci were not in HWE in one population, four loci in two populations (Aae\_2120, Aae\_4985, Aae\_5780, Aae\_7796), one locus in three populations (Aae\_8427), and one locus in four populations (Aae\_5563). Significant linkage disequilibrium was identified in 58 (of 4,278) pairs of loci in alewife and 88 (of 4,562) pairs of loci in blueback herring. These are less than the numbers expected by chance alone, and no geographic pattern in the significant values was apparent for either species.

The majority of both blueback herring and alewife individuals were accurately assigned to their population and reporting group of origin in the self-assignment analyses. For alewife, 67% of all fish were correctly assigned to population of origin and 93% to reporting group of origin (Table 2a) when no probability criterion was used, and 83% to population and 97% to reporting group, when a 90% probability criterion was applied (Table 2b). For blueback herring, assignment accuracy was similar, with 67% accurately assigned to population of origin and 96% assigned to reporting group of origin with no probability criterion (Table 2c) and 79% to population and 98% to reporting group, when a 90% probability criterion was applied (Table 2d). Self-assignment for alewife was less accurate in northern populations than in the south, corresponding to increased geographic distance between rivers in the south, whereas there was no discernible pattern with blueback herring.

Comparisons across alewife populations found generally low differentiation across the species range, with significant  $F_{ST}$  values ranging from 0.006 to 0.140. The highest  $F_{ST}$  values were between the southernmost populations (Alligator and Chowan) and populations in the Northern New England reporting group (Penobscot and Androscoggin), with somewhat less differentiation seen between these populations and the northernmost ones in Canada (Waugh and Tuskett) (Table 3a). Pairwise  $F_{ST}$  was nonsignificant only between the Alligator and Chowan Rivers. Pairwise  $F_{ST}$  values in blueback herring showed a similar pattern of modest genetic differentiation across the range, with significant values ranging from 0.004 to 0.150. The two southernmost populations were again the most genetically distinct when compared to all other validation populations, with almost all values of  $F_{ST} > 0.1$ , whereas none of the other values exceeded 0.1 (Table 3). Pairwise  $F_{ST}$  was significantly different from zero between all pairs of populations, except those from the Delaware and Rappahannock Rivers.

Model-based clustering analysis with *structure* for the eight populations of alewife found that, at  $K = 2$ , the two southernmost populations, Chowan and Alligator Rivers, cluster together (Fig. S1), consistent with their higher  $F_{ST}$  values with northern populations and the nonsignificance between them. At  $K = 3$ , the Penobscot and Androscoggin River populations cluster together. At  $K = 4$ , the northernmost Waugh and Tuskett populations formed a distinct

cluster (Figure 2). DAPC identified the same four clusters (Figure 3a). Clustering analyses with the 10 blueback herring populations at  $K = 2$  again found that the two southernmost populations, the Savannah and Altamaha Rivers, formed a distinct cluster (Fig. S1). At  $K = 3$ , two proximate populations in the middle of the range, the Mystic and Monument Rivers, separated as a distinct cluster. At  $K = 4$ , the four northernmost populations separated (Figure 2), while at higher values of  $K$ , most individuals in multiple populations began to separate fractionally into different clusters, indicating that the analysis had exceeded the most likely value of  $K$ . Clustering with DAPC again identified four distinct groups (Figure 3b), which mostly corresponded to the clusters found by *structure*.

## 4 | DISCUSSION

Molecular genetic data and analysis have become a critical component of biological investigation and conservation for migratory species, particularly anadromous fishes that are harvested and often subject to multijurisdictional management (Clemento et al., 2014; Hasselman et al., 2016; Palkovacs et al., 2014). We describe here validated SNP assays for alewife and blueback herring that provide power for multiple applications, including GSI across the species' ranges, as well as pedigree reconstruction and phylogeography.

Self-assignment analyses with both alewife and blueback herring populations demonstrated clear delineation between regional genetic stocks previously identified using microsatellite data (Palkovacs et al., 2014). Alewife validation populations displayed differentiation with these SNP assays that mirror the regional population genetic structure previously identified by Palkovacs et al. (2014) and expand the utility of genetic identification into the northern portion of the species range. Pairwise  $F_{ST}$  values revealed significant differentiation between all sets of populations except the Chowan and Alligator Rivers (Table 3), which are geographically proximate and tributaries of the same coastal estuary (Albemarle Sound). Two-thirds of alewife samples assigned to their correct population/river basin of origin, but the proportion of accurate assignments by population ranged between 49% and 84% (Table 2) when no probability criterion was applied. When such a criterion was applied, the overall proportion of accurate assignments increased substantially (82%), and the proportion per population ranged from 74% to 97%, but nearly half of the samples remained unassigned, emphasizing the lack of fine-scale differentiation and population structure in alewife. When self-assignment was evaluated at the scale of the previously reported regional genetic stocks, accuracy was much higher, with 93% of samples assigned accurately to regional stock of origin and the accuracy per population of assignment to reporting unit ranged from 82% to 100% without a probability criterion and 91% to 100% with a probability criterion.

Model-based clustering results from *structure* were consistent with the self-assignment and DAPC results for alewife and found that populations within the same regional genetic stock generally clustered together (Figure 2a and 3a, Table 2a,b). The

**TABLE 2** Accuracy of leave-one-out self-assignment analyses to population and regional stock for alewife (a) without applying a probability criterion (i.e., all individuals assigned) and (b) with a 90% criterion. Blueback herring are also assigned (c) without applying a probability criterion (d) with a 90% criterion

True population	Waughts	Tusket	Androscoggin	Penobscot	Mashpee	Quinnipiac	Chowan	Alligator	Prop. assigned population	Prop. assigned regional stock		
<b>(a) Alewife, no probability criterion</b>												
Waughts	18	4			2	2		1	.667	.815		
Tusket	3	38	1	1	1	1			.844	.911		
Androscoggin			34	13					.723	1.000		
Penobscot		1	16	26	1				.591	.955		
Mashpee	4			1	25	9	1		.625	.850		
Quinnipiac	2				6	27		1	.750	.917		
Chowan						1	27	13	.659	.976		
Alligator							22	21	.488	1.000		
<b>(b) Alewife, 90% probability criterion</b>												
Waughts	12	2			1				.800	.933		
Tusket	1	32							.970	1.000		
Androscoggin			11	3					.786	1.000		
Penobscot			4	14	1				.737	.947		
Mashpee	2				18	2			.818	.909		
Quinnipiac				4		19			.826	1.000		
Chowan							13	4	.765	1.000		
Alligator							2	10	.833	1.000		
<b>(c) Blueback, no probability criterion</b>												
True population	Margaree	Petitcodiac	East Machias	Kennebec	Mystic	Monument	Delaware	Rappahannock	Savannah	Altamaha	Prop. assigned population	Prop. assigned regional stock
Margaree	23	1	7	6	1		2				.575	.925
Petitcodiac		27									1.000	1.000
East Machias	3		25	15	1		1				.556	.956

(Continues)

**TABLE 2** (Continued)

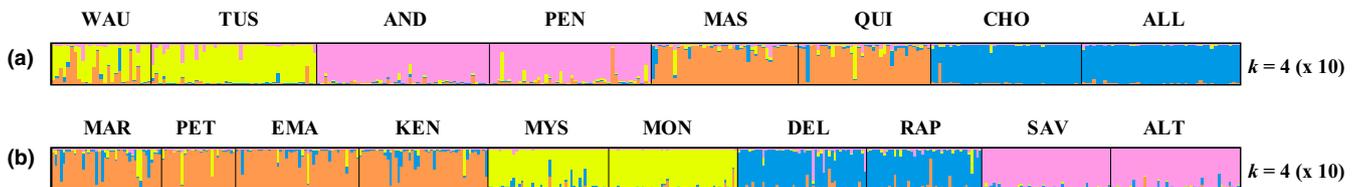
True population	Margaree	Petitcodiac	East Machias	Kennebec	Mystic	Monument	Delaware	Rappahannock	Savannah	Altamaha	Prop. assigned population	Prop. assigned to regional stock
Kennebec	4		9	31	1			2			.660	.936
Mystic				<b>38</b>	5	1					.864	.977
Monument				5	<b>42</b>						.894	1.000
Delaware			1			<b>23</b>	20		1	2	.489	.915
Rappahannock	1		3	1		23	<b>14</b>				.333	.881
Savannah								<b>33</b>	14	<b>33</b>	.702	1.000
Altamaha								14		<b>33</b>	.702	1.000
(d) Blueback, 90% probability criterion												
Margaree	<b>11</b>		1		1						.846	.923
Petitcodiac		<b>23</b>									1.000	1.000
East Machias		5	5	5	1						.455	.909
Kennebec		2	<b>14</b>					1			.824	.941
Mystic					<b>30</b>	3					.909	1.000
Monument					1	<b>32</b>					.970	1.000
Delaware							<b>6</b>	8			.429	1.000
Rappahannock				1			8	<b>3</b>			.250	.91
Savannah									<b>17</b>	6	.739	1.000
Altamaha									3	<b>13</b>	.813	1.000

Bold = assignment to correct population of origin.

**TABLE 3** Pairwise  $F_{ST}$  values. Significance assessed with 200 permutations

Population	Tusket	Androscoggin	Penobscot	Mashpee	Quinnipiac	Chowan	Alligator		
Alewife									
Waugh's	0.029	0.075	0.060	0.030	0.034	0.083	0.075		
Tusket		0.067	0.055	0.061	0.059	0.088	0.080		
Androscoggin			0.006	0.064	0.067	0.140	0.124		
Penobscot				0.050	0.051	0.131	0.115		
Mashpee					0.015	0.077	0.066		
Quinnipiac						0.075	0.064		
Chowan							<b>0.000</b>		
Population	Petitcodiac	East Machias	Kennebec	Mystic	Monument	Delaware	Rappahannock	Savannah	Altamaha
Blueback herring									
Margaree	0.049	0.010	0.017	0.063	0.056	0.039	0.035	0.122	0.109
Petitcodiac		0.048	0.054	0.092	0.075	0.091	0.086	0.150	0.138
East Machias			0.004	0.069	0.063	0.046	0.043	0.134	0.119
Kennebec				0.065	0.062	0.042	0.040	0.131	0.113
Mystic					0.025	0.057	0.057	0.123	0.112
Monument						0.078	0.075	0.123	0.116
Delaware							<b>-0.002</b>	0.087	0.075
Rappahannock								0.087	0.076
Savannah									0.006

Bold italics = nonsignificant value.

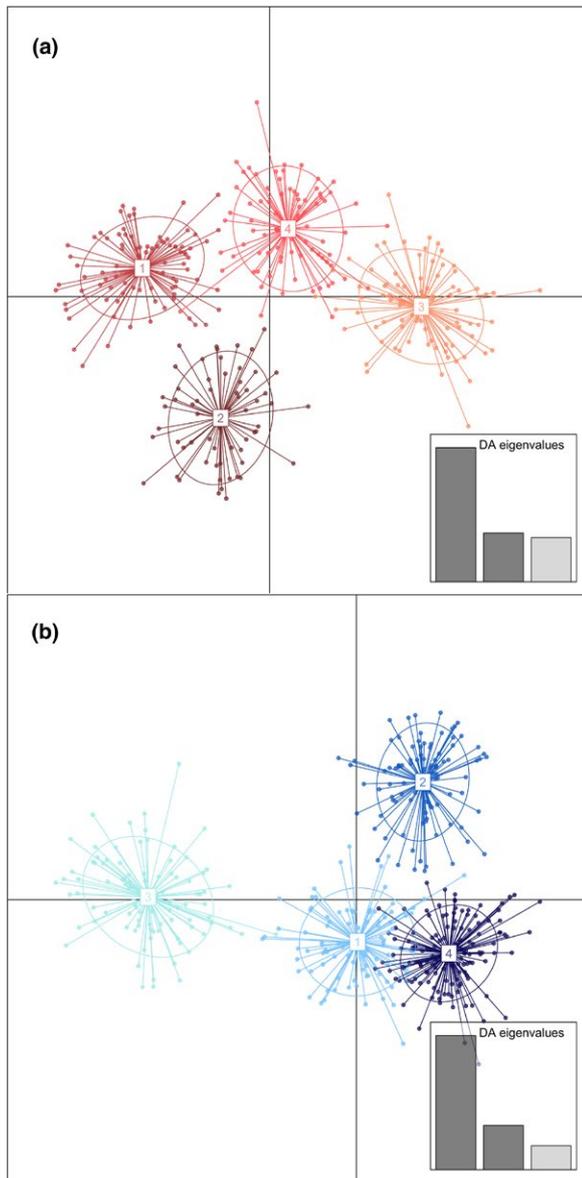


**FIGURE 2** Bayesian clustering analyses for (a) alewife and (b) blueback herring. The vertical lines represent fractional ancestry of individual fish partitioned into  $K = 4$  clusters, as indicated by colors. Alewife: WAU, Waugh's River; TUS, Tusket River; AND, Androscoggin River; PEN, Penobscot River; MAS, Mashpee River; QUI, Quinnipiac River; CHO, Chowan River; ALL, Alligator River. Blueback herring: MAR, Margaree River; PET, Petitcodiac River; EMA, East Machias River; KEN, Kennebec River; MYS, Mystic River; MON, Monument River; DEL, Delaware River; RAP, Rappahannock River; SAV, Savannah River; ALT, Altamaha River

two southernmost populations separated first, followed by the two Southern New England populations (Quinnipiac and Mashpee Rivers; Palkovacs et al., 2014), at  $K = 3$ . The two Northern New England populations and the two northernmost populations, in Canada, separated at higher  $K$  in some iterations. The two northernmost populations are both in Nova Scotia but geographically distinct, as one (Waugh's) is in the Gulf of the Saint Lawrence Seaway and the other (Tusket) borders the Bay of Fundy, where strong population structure has been identified in other anadromous species: America shad (*A. sapidissima*, Hasselman et al., 2010, 2013), Atlantic salmon (*Salmo salar*, McConnell et al., 1997; Verspoor et al., 2002; Spidle et al., 2003), striped bass (*Morone saxatilis*, Bradford et al., 2012), and rainbow smelt (*Osmerus mordax*, Coulson et al., 2014). Previous work on alewife (McBride

et al., 2014) found elevated levels of differentiation within the Bay of Fundy, as well as different life history traits and homing patterns between inner Bay of Fundy rivers and those outside the Gulf of Maine in North America.

For blueback herring, pairwise  $F_{ST}$  values revealed significant differentiation between all sets of populations except the Delaware and Rappahannock Rivers (Table 3). This is consistent with movement between the Delaware River and Chesapeake Bay (likely via the Chesapeake and Delaware Canal), as has been shown by otolith microchemistry (Turner, Limburg, & Palkovacs, 2015). The proportion of blueback herring correctly self-assigned to population of origin was similarly variable, ranging from 33% to 100% without a probability criterion (Table 2). Using a probability criterion increased overall accuracy of assignments (72%), with a range of 43%–100%



**FIGURE 3** Scatter plots of Discriminant Analysis of Principal Components (DAPC, Jombart, 2008) for (a) alewife and (b) blueback herring. In both species, the lowest BIC values corresponded with clusters of  $K = 4$ , and partitions populations geographically. The eigenvalues for the first three principal components are indicated in the insets for each species. The clusters for alewife represent (1) AND and PEN, (2) WAU and TUS, (3) CHO and ALL, and (4) MASH and QUI. For blueback herring, the clusters represent (1) DEL and RAP, (2) MYS and MON, (3) SAV and ALT, and (4) MAR, PET, EMA, and KEN. Population codes are as in Fig. 2

per population. Geographic proximity and population connectivity appear to explain many of the misassignments. For example, a high frequency of misassignments involved the Delaware and Rappahannock Rivers, which are connected via the Chesapeake and Delaware Canal, and the Savannah and Altamaha, which are geographically proximate (Figure 1). Similar to alewife, assignment to previously reported regional stocks was much more accurate, with overall assignment of 96% without a probability criterion and 97% with a probability criterion (Table 2).

The *structure* clustering results with blueback herring populations again mirrored the self-assignment and DAPC results (Figure 2b and 3b, Table 2c,d). Although the validation samples encompass most of the species range, the spatial distribution of populations is uneven, and the proximate populations consistently grouped. At  $K = 2$ , the southernmost populations, in the Altamaha and Savannah Rivers, clustered separately (Fig S1), whereas at  $K = 3$ , the proximate Southern New England populations separated. At  $K = 4$ , the four regional genetic stocks within the US range, previously identified by Palkovacs et al. (2014), were recovered (Figure 2b); however, the northernmost populations (Margaree and Petitcodiac) grouped with the Northern New England populations (Kennebec and East Machias). At  $K = 5$ , the Petitcodiac separated, whereas the Margaree continued to group with Northern New England (Fig. S1).

The 96 locus SNP set for blueback herring provided clustering concordant with the genetic stocks previously identified with microsatellite markers, and the four northernmost populations, which had not previously been evaluated together, formed a distinct cluster. The alewife SNP panel extends the geographic range for GSI, and also recovers clusters consistent with the regional genetic stocks found with microsatellites, yet the ability for the assays to discriminate population structure at small spatial scales, especially among rivers exchanging frequent migrants, may prove difficult.

Data from SNP assays are unambiguous and easily portable between laboratories. Generating these markers using high-throughput sequencing of genomic DNA enhances our ability to confidently distinguish populations of alewife and blueback herring that are genetically distinct across both species' ranges. This ability to identify stock of origin for fish caught at sea is critical for management of populations experiencing diminished spawning returns and of increasing conservation concern (Hasselman et al., 2016). Using these new tools, samples from mixed-stock assemblages can be quickly and efficiently genotyped, allowing new insights into marine movement patterns and impacts of marine fisheries.

## ACKNOWLEDGMENTS

We thank C. Columbus, E. Correa, and E. Campbell for assistance in the laboratory and P. Bentzen, M. McBride, and the Massachusetts Department of Marine Fisheries for providing samples. This work was supported by the National Fish and Wildlife Federation (NFWF #0104.14.041425), Atlantic States Marine Fisheries Commission (ASMFC #15-0105, #15-0102), and The Nature Conservancy. EPP was partially supported by the NOAA Cooperative Institute for Maine Ecosystems and Climate.

## AUTHOR CONTRIBUTIONS

D. S. Baetscher conducted laboratory work and performed analyses. D. J. Hasselman collected samples and performed analyses. K. Reid performed analyses. E. P. Palkovacs collected samples and obtained funding. J. C. Garza performed analyses and obtained funding. All authors participated in study design and manuscript composition.

## CONFLICT OF INTEREST

None declared.

## DATA ACCESSIBILITY

Genotype data for all fish with all assays, FASTA files with consensus sequences from all ddRAD loci with at least one SNP, as well as haplotype calls from all combinations of locus and individual are in the DRYAD digital repository (doi:10.5061/dryad.v4q83).

## REFERENCES

- Ackerman, M. W., Habicht, C., & Seeb, L. W. (2011). Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. *Transactions of the American Fisheries Society*, 140, 865–881.
- Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*, 24, 1–20.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410.
- Anderson, E. C., & Garza, J. C. (2006). Single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172, 2567–2582.
- ASMFC (Atlantic States Marine Fisheries Commission) (2012). *River herring benchmark stock assessment*, vol. 1. Stock Assessment Report No. 12-02 of the Atlantic States Marine Fisheries Commission. Washington, DC.
- Belhkir, K., Borsari, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (1996–2004). *GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations*. Laboratoire Genome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.
- Bradford, R. G., LeBlanc, P., & Bentzen, P. (2012). *Update Status Report on Bay of Fundy Striped Bass (Morone saxatilis)*, DFO Canadian Science Advisory Secretariat 2012/02.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.
- Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H., & Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, 15, 1496–1502.
- Clemente, A. J., Abadía-Cardoso, A., Starks, H. A., & Garza, J. C. (2011). Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, *Oncorhynchus tshawytscha*. *Molecular Ecology Resources*, 11(Suppl. 1), 50–66.
- Clemente, A. J., Crandall, E. D., Garza, J. C., & Anderson, E. C. (2014). Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook Salmon (*Oncorhynchus tshawytscha*) in the California Current large marine ecosystem. *Fishery Bulletin*, 112, 112–130.
- Coulson, M. (2014). *Historical and contemporary processes shaping population genetic structure in an anadromous fish (Osmerus mordax)*. Ph.D. dissertation, Dalhousie University, Canada.
- Hasselman, D. J. (2010). *Spatial distribution of neutral genetic variation in a wide ranging anadromous clupeid, American shad (Alosa sapidissima)*. Ph.D. dissertation, Dalhousie University, Canada.
- Hasselman, D. J., Anderson, E. C., Argo, E. E., Bethoney, N. D., Gephard, S. R., Post, D. M., ... Palkovacs, E. P. (2016). Genetic stock composition of marine bycatch reveals disproportional impacts on depleted river herring genetic stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, 73, 951–963.
- Hasselman, D. J., Argo, E. E., McBride, M. C., Bentzen, P., Schultz, T. F., Perez-Umphrey, A. A., & Palkovacs, E. P. (2014). Human disturbances causes the formation of hybrid swarm between two naturally sympatric fish species. *Molecular Ecology*, 23, 1137–1152.
- Hasselman, D. J., Ricard, D., & Bentzen, P. (2013). Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular Ecology*, 22, 1558–1573.
- Hightower, J. E., Wicker, A. M., & Endres, K. M. (1996). Historical Trends in Abundance of American Shad and River Herring in Albermarle Sound, North Carolina. *North American Journal of Fisheries Management*, 16, 257–271.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94.
- Larson, W. A., Seeb, J. E., Pascal, C. E., Templin, W. D., & Seeb, L. W. (2014). Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 71, 1–11.
- Limburg, K. E., & Waldman, J. R. (2009). Dramatic declines in North Atlantic diadromous fishes. *BioScience*, 59, 955–965.
- Loesch, J. G. (1987). Overview of life history aspects of anadromous alewife and blueback herring in freshwater habitats. *American Fisheries Society Symposium*, 1, 89–103.
- McBride, M. C., Hasselman, D. J., Willis, T. V., Palkovacs, E. P., & Bentzen, P. (2015). Influence of stocking history on the genetic structure of anadromous alewife (*Alosa pseudoharengus*) in Maine rivers. *Conservation Genetics*, 16, 1209–1223.
- McBride, M. C., Willis, T. V., Bradford, R. G., & Bentzen, P. (2014). Genetic diversity and structure of two hybridizing anadromous fishes (*Alosa pseudoharengus*, *Alosa aestivalis*) across the northern portion of their ranges. *Conservation Genetics*, 15, 1281–1298.
- McConnell, S. K. J., Ruzzante, D. E., O'Reilly, P. T., & Hamilton, L. (1997). Microsatellite loci reveal highly significant genetic differentiation among Atlantic salmon (*Salmo salar* L.) stocks from the east coast of Canada. *Molecular Ecology*, 6, 1075–1089.
- Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19, 208–216.
- Narum, S. R., Banks, M., Beacham, T. D., Bellinger, M. R., Campbell, M. R., Dekoning, J., ... Garza, J. C. (2008). Differentiating salmon populations at broad and fine geographic scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, 17, 3464–3477.
- 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).
- Palkovacs, E. P., Hasselman, D. J., Argo, E. E., Gephard, S. R., Limburg, K. E., Post, D. M., ... Willis, T. V. (2014). Combining genetic and demographic information to prioritize conservation efforts for anadromous alewife and blueback herring. *Evolutionary Applications*, 7, 212–226.
- Park, S. D. E. (2001). *Trypanotolerance in West African cattle and the population genetic effects of selection*. Ph.D. dissertation, University of Dublin, Ireland.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7(5), e37135.

- Piry, S., Alapetite, A., Cornuet, J. M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GeneClass2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, *95*, 536–539.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <https://www.R-project.org/>
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, *4*, 137–138.
- Rousset, F. (2008). Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106.
- Seeb, J. E., Pascal, C. E., Ramakrishnan, R., & Seeb, L. W. (2009). SNP genotyping by the 5'-nuclease reaction: Advances in high throughput genotyping with non-model organisms. In A. Komar (Ed.), *Methods in Molecular Biology, Single Nucleotide Polymorphisms*, 2nd ed. (pp. 277–292). New York, NY: Humana Press.
- Spidle, A. P., Kalinowski, S. T., Lubinski, B. A., Perkins, D. L., Beland, K. F., Kocik, J. F., & King, T. L. (2003). Population structure of Atlantic salmon in Maine with reference to populations from Atlantic Canada. *Transactions of the American Fisheries Society*, *132*, 196–209.
- Starks, H. A., Clemento, A. J., & Garza, J. C. (2016). Discover and characterization of single nucleotide polymorphisms in coho salmon, *Oncorhynchus kisutch*. *Molecular Ecology Resources*, *16*, 277–287.
- Turner, S. M., Limburg, K. E., & Palkovacs, E. P. (2015). Can different combinations of natural tags identify river herring natal origins at different levels of stock structure? *Canadian Journal of Fisheries and Aquatic Sciences*, *72*, 845–854.
- Verspoor, E., O'Sullivan, M., Arnold, A. L., Knox, D., & Amiro, P. G. (2002). Restricted matrilineal gene flow and regional differentiation among Atlantic salmon (*Salmo salar* L.) populations within the Bay of Fundy, eastern Canada. *Heredity*, *89*, 465–472.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Baetscher DS, Hasselman DJ, Reid K, Palkovacs EP, Garza JC. Discovery and characterization of single nucleotide polymorphisms in two anadromous alosine fishes of conservation concern. *Ecol Evol*. 2017;00:1–11. <https://doi.org/10.1002/ece3.3215>