Received: 6 February 2019 | Revised: 15 May 2019 | Accepted: 21 May 2019

DOI: 10.1111/mec.15136

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Evidence for the genetic basis and epistatic interactions underlying ocean- and river-maturing ecotypes of Pacific Lamprey (*Entosphenus tridentatus*) returning to the Klamath River, California

Keith A. Parker[1] | Jon E. Hess[2] | Shawn R. Narum[3] | Andrew P. Kinziger[1]

[1]Department of Fisheries Biology, Humboldt State University, Arcata, California

[2]Columbia River Inter-Tribal Fish Commission, Portland, Oregon

[3]Columbia River Inter-Tribal Fish Commission, Hagerman, Idaho

Correspondence
Andrew P. Kinziger, Department of Fisheries Biology, Humboldt State University, Arcata, CA 95521.
Email: Andrew.Kinziger@Humboldt.edu

Present Address
Keith A. Parker, Fisheries Department, Yurok Tribe, Klamath, California

Funding information
National Science Foundation-Graduate Research Fellowship; Robert & Patricia Switzer Foundation Fellowship

Abstract

Surveys of genomic variation have improved our understanding of the relationship between fitness-related phenotypes and their underlying genetic basis. In some cases, single large-effect genes have been found to underlie important traits; however, complex traits are expected to be under polygenic control and elucidation of multiple gene interactions may be required to fully understand the genetic basis of the trait. In this study, we investigated the genetic basis of the ocean- and river-maturing ecotypes in anadromous Pacific lamprey (*Entosphenus tridentatus*). In Pacific lamprey, the ocean-maturing ecotype is distinguished by advanced maturity of females (e.g., large egg mass) at the onset of freshwater migration relative to immature females of the river-maturing ecotype. We examined a total of 219 adult Pacific lamprey that were collected at-entry to the Klamath River over a 12-month period. Each individual was genotyped at 308 SNPs representing known neutral and adaptive loci and measured at morphological traits, including egg mass as an indicator of ocean- and river-maturing ecotype for females. The two ecotypes did not exhibit genetic structure at 148 neutral loci, indicating that ecotypic diversity exists within a single population. In contrast, we identified the genetic basis of maturation ecotypes in Pacific lamprey as polygenic, involving two unlinked gene regions that have a complex epistatic relationship. Importantly, these gene regions appear to show stronger effects when considered in gene interaction models than if just considered additive, illustrating the importance of considering epistatic effects and gene networks when researching the genetic basis of complex traits in Pacific lamprey and other species.

KEYWORDS

association testing, duplicate dominant epistasis, *Entosphenus tridentatus*, epistasis, gene–gene interactions, ocean-maturing, Pacific Lamprey, river-maturing

1 | INTRODUCTION

The ability to survey genomic variation in natural populations of organisms has paved the way for the identification of specific genes that are associated with fitness-related traits (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013). This genetic information can improve our understanding of phenotypic variation and offers the potential to focus conservation efforts on the genomic regions that are associated with traits that are likely influenced by selection (Allendorf, Hohenlohe, & Luikart, 2010). While ecological variation is generally considered in defining conservation units (e.g., Waples, 1991), how to incorporate the novel insights generated by genotype–phenotype studies into conservation planning is unclear and currently under debate (McMahon, Teeling, & Hogland, 2014; Shafer et al., 2015; Waples & Lindley, 2018).

The first step to effectively apply genomics data to conservation is to identify the single gene or multigene combinations that underpin the focal trait of interest. In some cases, single large-effect genes may underlie important traits (Barson et al., 2015; Hess, Zendt, Matala, & Narum, 2016; Narum, Genova, Micheletti, & Maass, 2018; Pearse, Miller, Abadía-Cardoso, & Garza, 2014; Prince et al., 2017; Thompson et al., 2019). However, the reigning paradigm in quantitative genetics posits that most complex traits are controlled by many genes of small effect (Bernatchez, 2016; Lynch & Walsh, 1998; Pavey et al., 2015). When many genes are involved, complex interactions between genes (or epistasis) may influence genotype–phenotype associations (Phillips, 2008). Epistatic effects are ubiquitous in natural systems but can be difficult to identify owing to the large number of interactions that must be statistically tested and the potential for a gene's effects to be obscured due to interactions with other loci (Phillips, 2008; Yang et al., 2009). Epistatic interactions can result in genetic markers that show little effect on a trait when considered individually but strong effect when considered in combination with other loci (Cordell, 2002). Emerging research suggests that epistatic interactions play important roles in the structure and function of gene networks and are critical for understanding complex traits (Boyle, Li, & Pritchard, 2017; Callaway, 2017).

We investigated the association of genetic variation with ecotypic differentiation in Pacific lamprey (*Entosphenus tridentatus*). The Pacific lamprey is an anadromous species inhabiting coastal rivers and nearshore marine waters from Hokkaido Island, Japan, to southern California, USA (Moyle, 2002). Pacific lamprey provide direct subsistence to indigenous people in the Pacific Northwest of North America, but due to declines in abundance, the species is considered to be at risk of extinction (Close et al., 1995; Close, Fitzpatrick, & Li, 2002; Murphey, 1959; Petersen-Lewis, 2009). In the Klamath River of northern California, two Pacific lamprey ecotypes have been documented, termed ocean- and river-maturing (Clemens, van de Wetering, Sower, & Schreck, 2013). The primary difference between the ecotypes is maturity at the onset of freshwater migration, with egg mass of the river-maturing ecotype less than half of the ocean-maturing ecotype (1.2%–2.8% vs. 5.5% ratio between egg mass and egg-free body mass) (Clemens et al., 2013). Owing to differences in maturity at entry, it is hypothesized that the ocean-maturing ecotype would likely spawn several weeks after entering freshwater, whereas the river-maturing ecotype would spend one year or more in freshwater prior to spawning (Beamish, 1980; Clemens et al., 2013; Figure 1). Thus, river-maturing fish holding over from the previous year may potentially interbreed with ocean-maturing fish from the current year migration. If interbreeding is occurring, then we would predict that the maturation ecotypes should exhibit homogenous allele frequencies at neutral loci.

The evolution of ecotypic diversity in Pacific lamprey appears to be a bet-hedging strategy similar to that exhibited by Pacific salmonids, a taxonomic group that shares a similar distribution and anadromous life history (Schindler et al., 2010). The Pacific lamprey ocean- and river-maturing ecotypes appear superficially most analogous to summer- and winter-run ecotypes described in anadromous steelhead (*Oncorhynchus mykiss*) (Clemens et al., 2013). Winter-run steelhead enters natal rivers from December to May in a more mature state, whereas summer-run steelhead enters natal rivers from May to October in an immature state (Myers et al., 2006). Genome-wide association testing has identified a single large-effect locus associated with steelhead ecotypic variation (Hess et al., 2016; Micheletti, Hess, Zendt, & Narum, 2018; Prince et al., 2017). For Pacific lamprey, it is unknown whether there is a genetic basis to ecotypic differentiation or whether the ecotypes initiate their freshwater migrations during different seasons. Pacific lamprey have been observed to initiate their anadromous

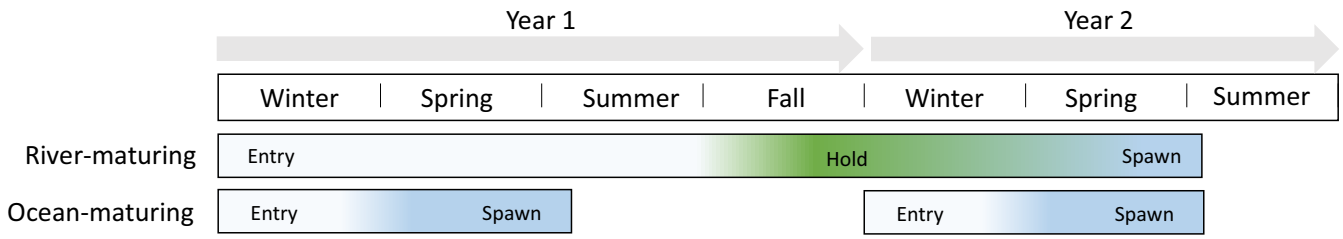

FIGURE 1 Hypothesized freshwater migration strategies for ocean- and river-maturing ecotypes of Pacific lamprey returning to the Klamath River, California. Ocean-maturing ecotype enters freshwater at an advanced state of maturity and likely spawn several weeks after entering fresh water, whereas the river-maturing enter in an immature state and would therefore spend one year in freshwater prior to spawning. Indicated are the estimated peak timing and duration of freshwater entry (white), holding (green) and spawning (blue) for ocean- and river-maturing ecotypes

FIGURE 2 (a) Aerial view of the collection site at the mouth of the Klamath River, Del Norte County, California, USA. Site was photographed in 2016. Inset provides location of study area in northwestern California. (b) Adult Pacific lamprey collected at-entry to the Klamath River at the initiation of their freshwater migration. (c) Yurok Native American eel hook used for collection of Pacific lamprey. (d) Variation in Pacific lamprey egg mass for individuals collected on the same day, 14 April 2017. Egg mass ranged from the smallest of the study at 1.6 g (third from left) to 22.7 g (second from right). (e) Female Pacific lamprey gut cavity prior to egg excision. The individual pictured represents the largest egg mass of the study (25.5 g)

migrations almost year around, but abundance peaks from early March to late June (Larson & Belchik, 1998; Moyle, 2002; Murphey, 1959; Petersen, 2006; Petersen-Lewis, 2009). In some basins, such as the Klamath River of northern California, USA, two distinct spawning runs of Pacific lamprey, a spring- and fall-run, have been suggested but not well documented (Anglin, 1994).

Genome-wide scans of genetic variation in Pacific lamprey have identified 162 SNP loci that are likely associated with adaptive variation (Hess, Campbell, Close, Docker, & Narum, 2013). Linkage analysis shows that these 162 loci are distributed across four linkage groups, termed linkage groups A, B, C and D (Hess et al., 2013). Preliminary comparisons to the sea lamprey genome (Smith et al., 2018) indicate that linkage groups A, B, C and D align to four different chromosomes and that SNPs from each linkage group localize to the same gene region of each of those chromosomes (unpublished data). Loci on linkage groups A, B and C exhibit strong associations with adult body metrics (primarily length) and migration distance in the Columbia River Basin, but equally strong phenotypic associations of the adaptive loci on group D have yet to be elucidated (Hess et al., 2015, 2014). Studies of neutral genetic loci in Pacific lamprey suggest the species is geographically structured, including three genetic groups: northern British Columbia, U.S. West Coast (including the Columbia River) and "dwarf" adults (Goodman, Reid, Docker, Hass, & Kinziger, 2008; Hess et al., 2013; Spice, Goodman, Reid, & Docker, 2012). One study has resolved evidence of inter- and intra-annual temporal genetic structure in Pacific lamprey in the Willamette River, Oregon (Clemens et al., 2017).

For this study, we collected river- and ocean-maturing ecotypes of Pacific lamprey as they initiated their anadromous migration to spawn in freshwater and genotyped each individual with a known set of neutral and adaptive markers to investigate the following questions:

1. Is there evidence for a genetic basis of river- and ocean-maturing ecotypes and do gene–gene interactions have greater predictive power for this trait compared to a single-locus model?
2. Do neutral markers indicate panmixia between ocean- and river-maturing ecotypes or do these ecotypes exhibit some degree of reproductive isolation?
3. Is there evidence for a genetic basis of adult body size similar to the pattern found in the Columbia River basin?

2 | MATERIALS AND METHODS

Adult Pacific lamprey were collected at-entry to the Klamath River from the Pacific Ocean (Table S1; Figure 2) (coordinates 41.544, −124.079). Specimens were captured within 0.5 km of the mouth of the Klamath River and collection effort was distributed across a 12-month period (June 2016 to May 2017), representing a full year of collections. Specimens were obtained via creel survey of the Yurok Tribe subsistence fishery (Figure 2) or captured by the first author using a traditional Native American eel hook (60–120 cm pole with a 30- to 60-cm attached wire terminating in a hook) (Figure 2) or a long-handled dip net (152- to 183-cm-long handle with a 51-cm bow of fine mesh). All noncreel individuals were euthanized by severing the notochord just posterior, and dorsal, of the eyes, followed by pithing with a metal rod through the brain. A fin clip (2 cm^2) was collected from the dorsal fin of each lamprey and preserved in 95% ethanol until DNA extraction. The following were recorded for each adult Pacific lamprey: day of year, total length (1 mm), body mass (0.1 g), girth just behind the posterior-most gill pore (1 mm), interdorsal distance (1 mm) defined as the distance between the insertion of the posterior-most ray of the first dorsal fin to the insertion of the anterior-most ray on the second dorsal fin and sex. For females, we also recorded egg mass (0.1 g), consisting of the total weight of all eggs without the membrane and used this to calculate the ratio between egg mass and egg-free body mass following Clemens et al. (2013). Condition factor was calculated as the ratio between actual body mass (somatic mass + egg mass), and the predicted body mass derived from a length-weight model created from log-transformed lengths and body masses.

2.1 | Molecular methods

Single nucleotide polymorphism genotypes were generated using the Genotyping-in-Thousands by sequencing (GT-seq) custom amplicon method described in Campbell, Harmon, and Narum (2015). The discovery, selection and development of a sufficient number of SNP markers to characterize Pacific lamprey population variability were the result of Hess et al. (2013) and Hess et al. (2015). A SNP panel of 308 loci were selected to be representative of neutral and adaptive loci across the geographic range of Pacific lamprey, representing a subset of markers developed from the paired-end consensus reads from the Hess et al. (2013) RAD-seq data set. The selection of loci and steps in development are described in the supporting information. Final optimization resulted in 308 markers that worked best in GT-seq genotyping. There are 230 neutral SNPs (between the 10th and 90th percentiles of the F_{ST} probability distribution from LOSITAN), 41 adaptive SNPs ($p > 0.995$) and a set of 31 "intermediate" SNPs that did not fit definitions of either putative neutral and putatively adaptive (i.e., $p < 0.100$ or $0.900 > p < 0.995$; Hess et al., 2013). Finally, four loci are species diagnostic (Hess et al., 2015), and 2 loci were duplicated. Therefore, there were 302 unique markers available for these association analyses out of the total 308 that were genotyped. These markers include 38 SNPs that were mostly adaptive loci that were categorized into the following four groups of linked loci: A ($N = 10$), B ($N = 13$), C ($N = 7$) and D ($N = 8$, Hess et al., 2013).

2.2 | Single gene association tests

Genotype–phenotype associations were tested using a general linearized model (GLM) and a mixed linearized model (MLM) using the software TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) v. 5.2.50 (Bradbury et al., 2007). Analysis using a GLM reduces the risk of false positives due to population structure by including population membership estimates as covariates in the model. Population structure was estimated based upon principal components analysis (first three PC axes) of 76 neutral SNPs (as defined in Hess et al., 2015) using the software TASSEL. These 76 neutral SNPs are a subset of the 85 neutral SNPs that were characterized in detail by Hess et al. (2015) for rangewide collections and were successfully integrated into the GT-seq assay panel. Therefore, these 76 SNPs were expected to be informative to estimate neutral population structure in rangewide assessments of Pacific lamprey, including the Klamath River. Genotype–phenotype associations were also investigated using MLM, an approach that controls for false positives that may arise from both population and family structure and is therefore a more stringent association analysis than GLM. The MLM analysis used the population structure as estimated by the principal component analysis (see above) and a kinship matrix ("scaled IBS" method; Endelman & Jannink, 2012) which was calculated using TASSEL.

The data set used for association testing was restricted to include females because maturation state was not recorded from males. The data set consisted of 302 SNPs genotyped in 92 female adult Pacific lamprey collected at-entry to the Klamath River. Eight traits were used for genotype–phenotype association testing including egg mass, the ratio between egg mass and egg-free body mass, total length, body mass, girth, interdorsal distance, river entry date (day of year) and condition factor. Pearson's correlation coefficient and tests of significance were used to examine intercorrelations among the traits.

The large number of tests for phenotype–genotype associations (8 traits × 302 loci = 2,416 tests) increased the possibility of detecting a significant association by chance. To account for multiple tests, only those associations with p-values less than the critical value as determined using the false discovery rate procedure described by Benjamini and Hochberg (1995) were considered significant. The Benjamini and Hochberg (1995) false discovery rate approach has more power to detect significant differences than sequential Bonferroni correction (Narum, 2006; Rice, 1989). Critical values were calculated using the function p.adjust using the R package stats (R Core Team, 2018).

2.3 | Gene–gene interaction association tests

To identify the combined effects of multiple genes and gene–gene interactions on egg mass at-entry in Pacific lamprey, we employed

Generalized Multifactor Dimensionality Reduction (GMDR). GMDR is part of a family of machine learning approaches based upon Multifactor Dimensionality Reduction (MDR), which have become popular for identifying gene interactions underlying complex human diseases (Gola, John, van Steen, & Konig, 2016; Li, Guo, Wang, Liu, & Zou, 2015). An explanation of MDR approaches and its extension to quantitative traits, termed GMDR, have been described previously (Lou et al., 2007; Ritchie, Hahn, & Moore, 2003; Richie et al., 2001). Overall, MDR approaches are advantageous because they are nonparametric and do not make assumptions about specific genetic models. However, the GMDR method is computationally intensive and interpretation of gene interactions is not always straight-forward. We selected GMDR because it accommodates quantitative phenotypes such as egg mass at entry, our indicator of ocean- and river-maturing ecotypes.

The GMDR approach uses scores to identify genotype combinations associated with the trait of interest. Scores were calculated using a linear regression model with egg mass as a response variable. The score was used to classify status (e.g., ocean-mature/large egg mass or river-maturing/small egg mass) of each genotype combination. We conducted an exhaustive search for all possible one- to four-locus models. The best model was defined as the model with the maximal cross-validation consistency and highest testing balanced accuracy. Cross-validation consistency is defined as how many times the same model is identified in all 10 training data sets. Training data sets were constructed by randomly partitioning the data into 10 equal or nearly equal subdivisions, one subdivision is used as the testing set and the remaining serve as the independent training set. Testing balanced accuracy is calculated as ((TP/(TP + FN)) + (TN/(TN + FP))/2, where TP = true positive, FP = false positive, TN = true negative and FN = false negative. Testing balanced accuracy is preferable when the data do not have the same number of instances in each class. Statistical significance of the best model was evaluated by comparing the testing balanced accuracy from the observed data to the distribution under the null hypothesis of no associations from 1,000 permutations. We used the software GMDR version 0.9 (Chen et al., 2011; Lou et al., 2007) to first identify the best models. We then determined testing balanced accuracy and conducted permutation analysis for the best models using a Perl script. For GMDR analysis, only those loci with no missing genotypes were included. In addition, only one locus from each linkage group was included: A = Etr_2287, B = Etr_2791, C = Etr_965 and D = Etr_2878. The final data set included 92 females and 152 loci. We conducted a parallel analysis using total length as the response variable because previous work has indicated multiple linkage groups (linkage groups A, B and C) were associated with total length in Pacific lamprey (Hess et al., 2015, 2014).

2.4 | Neutral genetic structure

To evaluate the extent of neutral genetic differentiation between river- and ocean-maturing ecotypes in Klamath River Pacific lamprey, the panel of 308 SNPs was filtered as follows. First, the four species identification loci, the two duplicated loci and all loci missing ≥5% of their genotypes were removed. Next, those loci identified as non-neutral using the software LOSITAN in Hess et al. (2013) were removed. Lastly, linked loci were identified using the software TASSEL (Bradbury et al., 2007), and one locus from each linked pair identified at a significance level of <0.01 was eliminated from the data set. Filtering resulted in a data set consisting of 148 SNP loci that were used to evaluate neutral patterns of genetic structure. Tests for conformance to Hardy–Weinberg proportions and estimates of observed and expected heterozygosity were calculated using the software GENODIVE 2.0b27 (Meirmans & Van Tienderen, 2004).

To examine the data for evidence of neutral genetic structure, two approaches were used: (a) Bayesian cluster analysis (Pritchard, Stephens, & Donnelly, 2000) and (b) K-means clustering (Meirmans, 2012). Both approaches do not require population hierarchy to be defined a priori and allow assessment as to whether any significant genetic structure is present. However, the two approaches employ different statistical frameworks, thereby allowing assessment of consistency across analytical approaches.

The Bayesian clustering algorithm implemented in the software STRUCTURE v 2.3.4 (Pritchard et al., 2000) was used to estimate the number of discrete genetic clusters (K) of individuals in the data and the probability of assignment of each individual to each cluster. Analyses were run for 200,000 steps (with 100,000 discarded as burn-in), and 20 independent runs were conducted for each value of *K*. Analyses were run assuming the data consisted of $K = 1 \ldots 5$ clusters. Summaries of replicate runs were calculated using the software STRUCTURE HARVESTER (Earl & vonHoldt, 2012), and the *K* with the highest probability was used as an indicator of the number of genetically distinct groups in the data.

The K-means cluster method implemented in the software GENODIVE was used to sort individuals into an arrangement that maximized variance among groups but minimized within-group diversity. The sums of squares from an AMOVA were used to calculate a distance matrix between individuals (Meirmans, 2012), and simulated annealing from 50,000 steps (repeated 20 times) was used to perform K-means clustering. Analyses were run assuming the data consisted of $K = 1 \ldots 5$ clusters, and selection of optimal K was based on Bayesian Information Criterion (BIC).

3 | RESULTS

3.1 | Single gene association tests

Egg mass at-entry, the trait used to define ocean- and river-maturing ecotypes, exhibited significant associations with 15 loci in the GLM analysis ($p \leq 0.00097896$; Table 1; Figure 3). A total of 14 of 15 loci found to exhibit significant associations with egg mass were previously suspected to be adaptive (Hess et al., 2013), and these loci were distributed across two linkage groups, including six loci on linkage group B and eight loci on linkage group D (Table 1). One locus, Etr_3383, was of unknown linkage relationship and not previously identified as adaptive by Hess et al. (2013). Mean egg mass at

TABLE 1 *p*-Values for traits and loci exhibiting significant genotype–phenotype associations in the GLM analysis. Only those associations identified as significant as determined using the false discovery rate procedure of Benjamini and Hochberg (1995) are reported ($p \leq 0.00097896$). The linkage group designations follow Hess et al. (2013)

Locus	Egg mass	Ratio between egg mass and egg-free body mass	Total length	Body mass	Girth	Interdorsal distance
Linkage Group A						
Etr_2287			3.77E-10	1.39E-06	3.34E-05	
Etr_3069			2.35E-09	1.22E-07	7.64E-06	
Etr_6363			2.35E-09	1.22E-07	7.64E-06	
Etr_5317			1.66E-09	5.07E-08	5.24E-06	
Etr_3638			6.42E-09	1.29E-06	6.73E-05	
Etr_2603			5.66E-07	3.26E-06	1.04E-04	3.57E-04
Linkage Group B						
Etr_2791	3.56E-05	3.29E-04				
Etr_4455	3.56E-05	3.29E-04				
Etr_1613	5.95E-05	3.85E-04				
Etr_2730	5.95E-05	3.85E-04				
Etr_2151	1.60E-04	9.79E-04				
Etr_1509	2.03E-04					
Linkage Group D						
Etr_464	2.65E-05	5.63E-05				
Etr_211B	1.20E-04	3.10E-04				
Etr_2878	1.20E-04	3.10E-04				
Etr_4156	1.20E-04	3.10E-04				
Etr_1944	2.34E-04	5.42E-04				
Etr_8649	2.34E-04	5.42E-04				
Etr_1378	2.55E-04	6.59E-04				
Etr_2097	7.02E-04					
Linkage Group NA						
Etr_4889			1.47E-09	6.07E-08	3.57E-06	
Etr_3885			7.18E-08	3.96E-07	1.87E-06	
Etr_3383	1.53E-04	4.85E-05				
Etr_8681						1.26E-04
Etr_2451						4.59E-04
Etr_3837				6.69E-04		

a linkage group B locus, Etr_2791, was largest at the AA genotype (12.5 g), smallest at the GG genotype the (5.9 g) and intermediate at the heterozygous genotype (6.9 g) (Figure 4). Mean egg mass at a linkage group D locus, Etr_2878, was largest at the TT genotype (12.4 g), smallest at the CC genotype the (5.8 g) and intermediate at the heterozygous genotype (8.4 g) (Figure 4). The ratio between egg mass and egg-free body mass exhibited a strong correlation with egg mass (0.95, Table 2), and therefore, these two traits exhibited similar associations (Table 1).

To visualize the strength of the association between the loci in the B and D linkage groups and eggs mass, a heatmap of each individuals' multilocus genotype was constructed (Figure 5). For standardization, genotypes of each individual were coded as homozygous for small egg mass/river-maturing, homozygous for large egg mass/ocean-maturing or heterozygous. The allele with the highest frequency among the large egg mass/ocean-maturing individuals was used to designate the ocean-maturing ecotype allele. The heatmap indicates a change in allele frequency at approximately 12.5 g egg mass. The large egg mass (≥12.5 g) or ocean-maturing individuals were generally homozygous for the large egg mass allele at both linkage groups B and D. In contrast, the small egg mass (<12.5 g) or river-maturing individuals exhibited higher genetic diversity including both river and ocean alleles at both linkage group B and D loci.

Total length exhibited significant associations with eight loci in the GLM analysis ($p \leq 0.00097896$; Table 1; Figure 3). A total of six of eight loci found to exhibit significant associations with total length were previously suspected to be adaptive (Hess et al., 2013), and these loci were located on linkage group A (Table 1). Representative

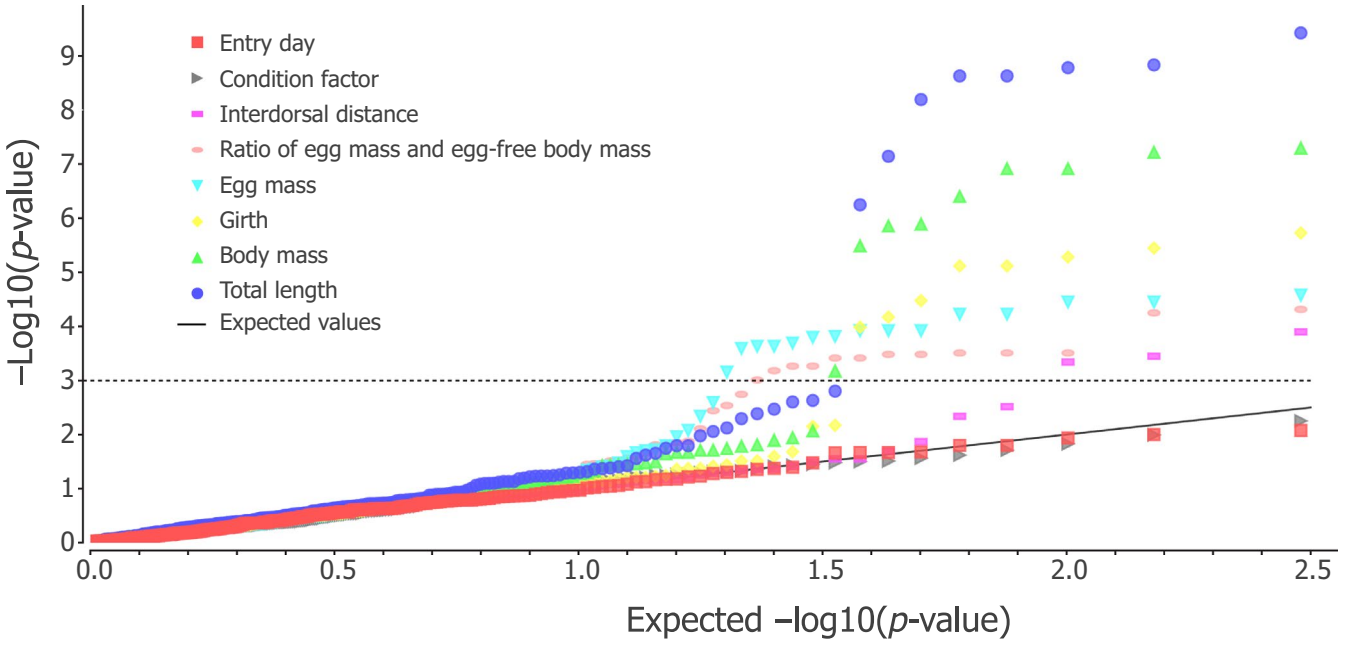

FIGURE 3 Quantile-Quantile plot from the GLM for eight traits and 302 SNP loci in 92 adult Pacific lamprey collected at entry to the Klamath River, California. Points above the dotted horizontal line identify significant associations as determined using the false discovery rate procedure of Benjamini and Hochberg (1995) ($p \leq 0.00097896$)

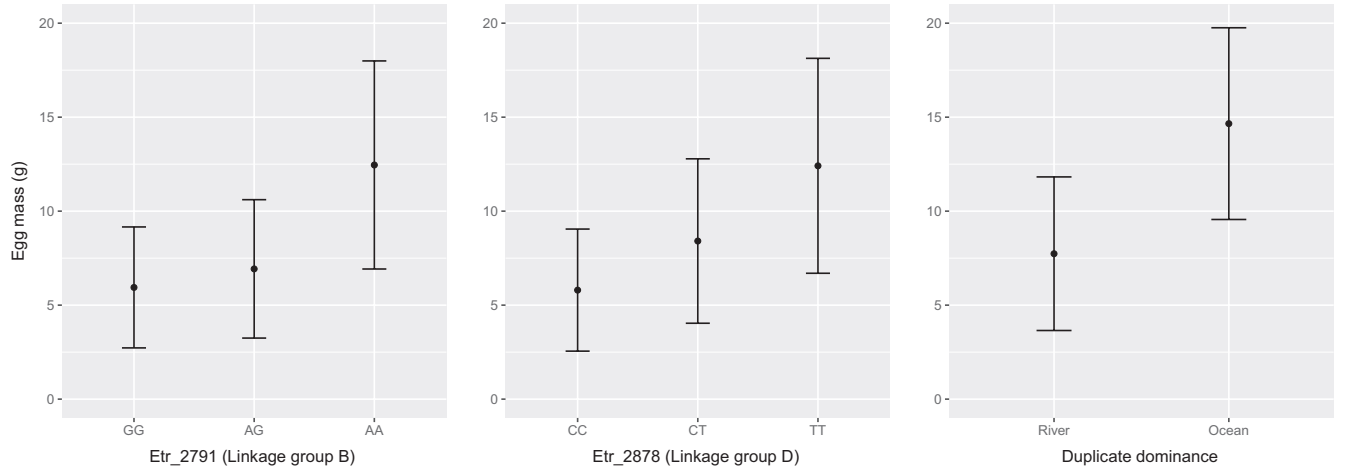

FIGURE 4 Mean egg mass (and standard deviation) of Pacific lamprey collected at entry to the Klamath River, California, at linkage group B (Etr_2791), linkage group D (Etr_2878) and under a model of duplicate dominance epistasis. Under this model, if Etr_2791 = AA and Etr_2878 = TT, then an individual was classified as ocean-maturing, whereas all other genotype combinations were assigned as river-maturing

loci from linkage group A have also been shown to be highly associated with body length in the Columbia River (Hess et al., 2014). Two loci, Etr_3885 and Etr_4889, are of unknown linkage relationship and not previously identified as adaptive by Hess et al. (2013). Total length exhibited a moderate correlation with body mass (0.82) and girth (0.66) (Table 2), and therefore, these intercorrelated traits exhibited significant associations with the same loci as total length (Table 1). The only difference was that body mass exhibited significant associations with one additional locus (9 total). However, associations of linkage group A were stronger for total length (mean *p*-value 8.15E-08) than for body mass (mean *p*-value 7.513E-05) or girth (mean *p*-value 2.88E-05; Table 1; Figure 3).

To visualize the strength of the association between the loci in the A linkage group and total length, a heatmap of each individuals' multilocus genotype was constructed (Figure 6). Visual inspection indicated a change in allele frequency at a total length of 625 mm. Based upon this, genotypes of each individual were coded as homozygous for large size, homozygous for small size or heterozygous. The allele with the highest frequency among the large individuals (≥625 mm total length) was used to designate the "large" allele.

TABLE 2 Pearson correlation coefficients (below diagonal) and significance tests (above diagonal) for the eight traits used for genotype–phenotype association testing in the 92 female adult Pacific lamprey collected at-entry to the Klamath River

	Day	Total length	Body mass	Girth	Interdorsal distance	Egg mass	Ratio between egg mass and egg-free body mass	Condition factor
Day	–	0.8512	0.2549	0.6461	0.2551	0	0	0.0397
Total length	0.02	–	0	0	0.0015	0.1823	0.341	0.9073
Body mass	−0.12	0.82	–	0	0.0002	0.1434	0.2239	0
Girth	0.05	0.66	0.87	–	0.0187	0.8869	0.0297	0
Interdorsal distance	−0.12	0.33	0.38	0.24	–	0.9915	0.3583	0.0789
Egg mass	−0.49	0.14	0.15	0.02	0	–	0	0.3066
Ratio between egg mass and egg-free body mass	−0.47	−0.1	−0.13	−0.23	−0.1	0.95	–	0.716
Condition factor	−0.21	−0.01	0.55	0.57	0.18	0.11	−0.04	–

The GLM analysis also detected significant associations between interdorsal distance and three loci (Table 1). One locus found to exhibit significant association with interdorsal distance was previously suspected to be adaptive (Hess et al., 2013), and this locus was located on linkage group A (Table 1). Two loci, Etr_2451 and Etr_8681, were of unknown linkage relationship and not previously identified as adaptive by Hess et al. (2013).

The MLM analysis produced similar results to the GLM, except that the more stringent MLM approach detected a subset of the significant associations produced by GLM, including: egg mass (one D group locus), the ratio between egg mass and egg-free body mass (one D group locus), total length (six A group loci and two loci of unknown linkage), body mass (six A group loci and two loci of unknown linkage), girth (five A group loci and two loci of unknown linkage) and interdorsal distance (one locus of unknown linkage; Benjamini & Hochberg, 1995; $p \leq 0.00050881$; Table S2).

3.2 | Gene–gene interaction association tests

For egg mass, the GMDR model with the maximal cross-validation accuracy (10/10) and highest testing balance accuracy (84%) was a two-locus interaction model including Etr_2791 (linkage group B) and Etr_2878 (linkage group D) (Figure 7; Table 3). A 7% increase in testing balanced accuracy was realized when using the two-locus interaction model (84%) in comparison with the best single-locus model (77%). Under the best two-locus model, if Etr_2791 = AA and Etr_2878 = TT, then an individual was classified as ocean-maturing/large egg mass, whereas all other genotype combinations were classified as river-maturing/small egg mass (Figure 7). Using this model, mean egg mass for the ocean-maturing ecotype (14.7 g) was almost double that of the river-maturing ecotype (7.8 g) (Figure 4).

For total length, two GMDR models emerged with no clear indication which one was the best fit: (a) a single-locus model including Etr_2287 (linkage group A) with cross-validation accuracy (10/10) and testing balance accuracy (83%), and (b) a two-locus model including Etr_2287 (linkage group A) and Etr_2791 (linkage group B) with cross-validation accuracy (8/10) and testing balance accuracy (87%) (Table 4; Figure 7). Under the best two-locus model, if Etr_2791 = AA or AG and Etr_2287 = TT, then an individual was classified as having larger total length (Figure 7). Models involving three or more loci had considerably lower cross-validation accuracy than the one- and two-locus models.

3.3 | Neutral genetic structure

Among the 148 SNPs identified as neutral and therefore suitable for the assessment of neutral genetic structure in river- and ocean-maturing ecotypes, 11 had an expected heterozygosity <0.15, 32 between 0.15 and 0.30, and 105 > 0.35. Tests for conformance to Hardy–Weinberg proportions (Table S3) revealed significant departures at 39 loci at a $p < 0.05$, and nine departures using a Bonferroni corrected p-value for multiple tests of 0.0003 (0.05/148 tests). Subsequent analyses of neutral structure compared results between the full 148 SNP loci vs. removal of the 39 loci with Hardy–Weinberg departures.

The at-entry collections of the Klamath River Pacific lamprey included both river- and ocean-maturing ecotypes, samples collected from 2016 and 2017, and samples collected across multiple months in each year, allowing the assessment of genetic differentiation among ecotypes and inter- and intra-annual temporal genetic structure. In the Bayesian cluster analysis using the software STRUCTURE, the highest log probability of the data was at $K = 1$ and visual inspection of $K > 1$ revealed that assignments were generally symmetric to all populations, indicative of the absence of ecotypic differentiation and inter- or intra-annual population genetic structure (Table S4). Similarly, K-means clustering using the software GENODIVE indicated the best clustering occurred at $K = 1$ according to BIC, suggesting there was no significant genetic structure in the data (Table S4).

Removal of the 39 loci that significantly departed from Hardy–Weinberg expectations, and Bayesian cluster analysis of a data set consisting of 109 SNP loci resolved patterns that were identical to the 148 SNP data set indicative of the absence of genetic differentiation between ocean- and river-maturing ecotypes and the absence of temporal genetic structure.

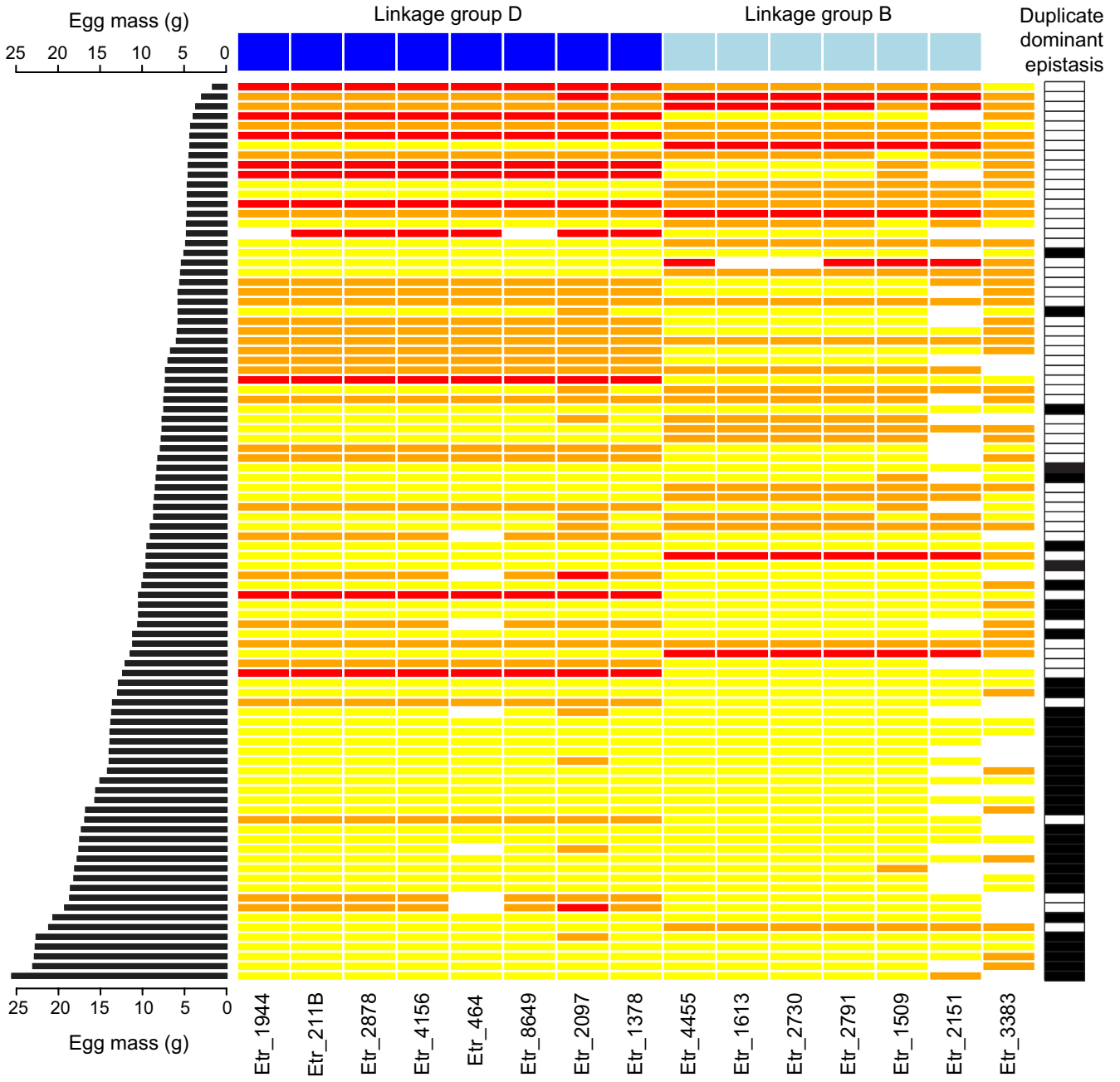

FIGURE 5 Genotype heatmap for the 15 SNP loci exhibiting significant association with egg mass in the GLM analysis of 92 adult Pacific lamprey collected at entry to the Klamath River, California. Each row indicates an individual multilocus genotype, coded homozygous for small egg/river-maturing (red), homozygous for large egg/ocean-maturing (yellow) and heterozygous (orange). Missing data are coded white. Individuals are ordered top to bottom from small to large egg mass, as indicated by the plot at the left. Loci are grouped by linkage group as indicated by the bar at the top. At right, categorization of individuals into river- or ocean-maturing according to duplicate dominant epistasis

4 | DISCUSSION

Our research resolved two primary results regarding ocean- and river-maturing ecotypes of Pacific lamprey returning to the Klamath River: (a) we identified the genetic basis of maturation ecotypes as polygenic, involving two unlinked gene regions (linkage groups B and D), and further found that the effects of the two gene regions did not appear to be additive but instead had complex interactive or epistatic relationship. (b) We found that Klamath River Pacific lamprey are panmictic at neutral loci indicating that maturation ecotypic diversity exists within a single population, presumably caused by interbreeding between hold-over river-maturing and current year and ocean-maturing ecotypes. Also, we observed large variation in egg mass among individuals collected on the same day (Figure 2), suggesting that river- and ocean-maturing ecotypes initiate freshwater migration simultaneously with each other (Figure 2). Our field observations did not resolve evidence for large-scale temporal separation in entry time between the river- and ocean-maturing ecotypes.

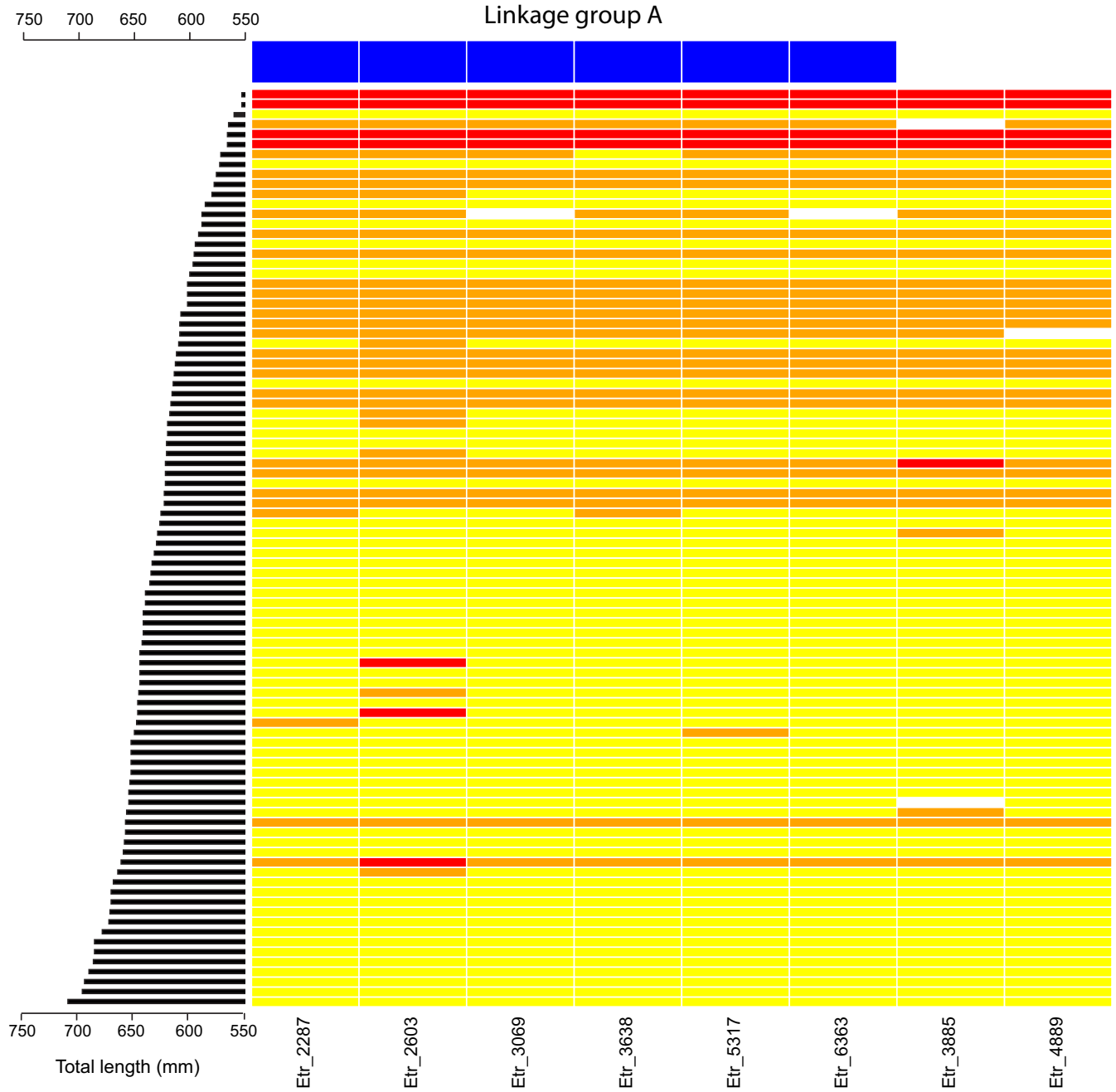

FIGURE 6 Genotype heatmap for the 8 SNP loci exhibiting significant association with total length in the GLM analysis of 92 adult Pacific lamprey collected at entry to the Klamath River, California. Each row indicates an individual multilocus genotype, coded homozygous for shorter length (red), homozygous for larger length (yellow) and heterozygous (orange). Missing data are coded white. Individuals are ordered top to bottom from small to large total length, as indicated by the plot of total length at the left. Loci are grouped by linkage group as indicated by the bar at the top

Our analysis indicated a role for gene interactions underlying river- and ocean-maturing ecotypes of Pacific lamprey, with a testing balance accuracy of the best gene interaction model at 84%. The best fit model from GMDR was consistent with a well-known form of epistasis, termed duplicate dominant epistasis (Miko, 2008). Under this form of epistasis, a dominant allele at either of two loci can mask the expression of recessive alleles at the two loci (Miko, 2008). For Pacific lamprey, the hypothesis is that the ocean-maturing ecotype would be expressed only when genes at both linkage groups B and D were homozygous recessive (if Etr_2791 = aa and Etr_2878 = tt), whereas only one dominant allele in either linkage group B (Etr_2791 = aG or GG) or D linkage group (Etr_2878 = Ct or CC) is necessary to express the river-maturing ecotype. This model suggests that the river-maturing ecotype carries more standing genetic variation than the ocean-maturing ecotype (Figure 5).

While our findings lend support for duplicate dominance epistasis for Pacific lamprey maturation ecotypes, assignment accuracy under this model was not 100% as would be expected if there was strict

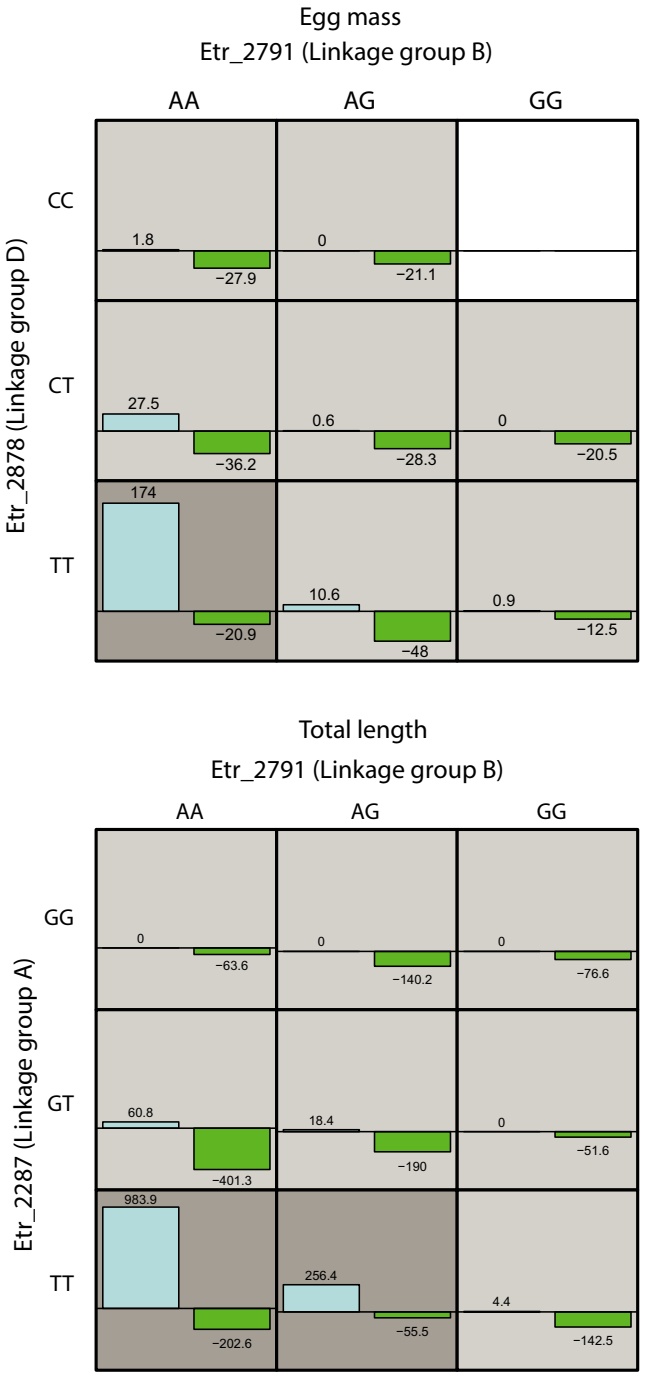

FIGURE 7 (Top) Plot of the best model from GMDR for egg mass. Dark grey boxes are genotype combinations associated with ocean-maturing/large egg mass, light grey boxes are genotype combinations associated with river-maturing/small egg mass, and white boxes are genotype combinations with no observations. The blue and green bars within each box correspond to the calculated scores for ocean-maturing/large egg mass and river maturing/small egg mass, respectively. (Bottom) Plot of the two-locus interaction model from GMDR for total length. Dark grey boxes are genotype combinations associated with large total length, and light grey are genotype combinations associated with small body size. The blue and green bars within each box correspond to the calculated scores for large total length and small total length, respectively. The number above each bar is the sum of scores, and the height of the bars is proportional to the sum of scores

adherence to this model (Miko, 2008). We hypothesize that the difference between model expectations and our observations is related to our inability to make perfect categorizations to maturity ecotype using egg mass as a proxy for quantifying maturation. Egg mass exhibited continuous variation (range 1.6–25.5 g), but we used it to make a binary diagnosis into ocean- and river-maturing ecotype. For example, a total of 11 individuals categorized as ocean-maturing according to the duplicate dominance model had intermediate egg masses ranging from 8.5 to 12.5 g (Figure 5). It is possible that our field collections may have intercepted individuals too early in gonadal development to allow for correct diagnosis of the maturation ecotypes. These intermediate individuals may potentially express their ocean-maturing ecotype by continuing egg maturation during migration. A better method for ecotype classification may serve to clarify adherence to the duplicate dominance model. Alternatively, Pacific lamprey maturation ecotypes may not be under exclusive genetic control, and environmental factors may likely play a role as is common with quantitative traits.

We found no evidence for neutral genetic differentiation among river- and ocean-maturing ecotypes of Pacific lamprey in the Klamath River. These findings are consistent with other population genetic assessments that have found that Pacific lamprey in a large area encompassing the U.S. West Coast (and Columbia River Basin) are genetically homogenous (Goodman et al., 2008; Hess et al., 2013; Spice et al., 2012). Our results support the hypothesis that river-maturing individuals holding over from the previous year likely interbreed with ocean-maturing fish from the current year migration (Figure 1). Thus, river- and ocean-maturing ecotypic diversity exists within a single random mating population, a pattern that also holds for summer- and winter-run steelhead (Hess et al., 2016).

The selective drivers maintaining maturation ecotypes within a single panmictic population are unknown; however, it has been hypothesized that rivers with low flow and warm water (>20°C) would select for ocean-maturing ecotypes and the opposite conditions (<20°C and high flow) would select for river-maturing ecotypes (Clemens, Schreck, Sower, & van de Wetering, 2016). The diversity of habitats available in many river basins suggests a role for balancing selection for maintaining both ecotypes. Understanding the selective forces acting upon river-maturing and ocean-maturing ecotypes will be important for ensuring the long-term maintenance of these two maturation ecotypes. For example, population genetic theory would predict that strong directional selection against the ocean-maturing ecotype would serve to maintain both ecotypes because the ocean-maturing allele is common within the river-maturing individuals (Figure 5). In contrast, strong directional selection against the river-maturing ecotype would drive the river-maturing allele towards elimination and could potentially cause loss of the river-maturing ecotype.

Our analysis also identified a strong genetic basis for body size (linkage group A) in Klamath River Pacific lamprey, with the strongest associations identified when using total length as a metric of body size (Figure 3). These findings are consistent with the association

TABLE 3 Generalized Multifactor Dimensionality Reduction for egg mass, including the number of loci, SNPs in the model, cross-validation consistency, testing balanced accuracy and *p*-values from permutation testing

Number of loci	SNPs in the model (linkage group)	Cross-validation consistency[a]	Testing balanced accuracy[b]	*p*-Value*
1	Etr_2791(B)	8/10	0.7709	0.004
2	Etr_2791(B), Etr_2878(D)	10/10	0.8391	0.001
3	Etr_2791(B), Etr_4750(NA), Etr_7292(NA)	3/10	0.8147	0.010
4	Etr_2791(B), Etr_1551(NA), Etr_1684(NA), Etr_3107(NA)	2/10	0.9091	0.002

[a]Cross-validation consistency is defined as the number of times the same model is identified in all 10 training data sets.
[b]Testing balanced accuracy is ((TP/(TP + FN)) + (TN/(TN + FP))/2, where TP = true positive, FP = false positive, TN = true negative and FN = false negative.
*Statistical significance was evaluated by comparing the testing balanced accuracy from the observed data to the distribution under the null hypothesis of no associations from 1,000 permutations.

TABLE 4 Generalized Multifactor Dimensionality Reduction for total length, including the number of loci, SNPs in the model, cross-validation consistency, testing balanced accuracy and *p*-values from permutation testing

Number of loci	SNPs in the model (linkage group)	Cross-validation consistency[a]	Testing balanced accuracy[b]	*p*-Value*
1	Etr_2287 (A)	10/10	0.8290	0.003
2	Etr_2287(A), Etr_2791(B)	8/10	0.8731	0.002
3	Etr_2287(A), Etr_2791(B), Etr_4750(NA)	3/10	0.8607	0.002
4	Etr_2287(A), Etr_2791(B), Etr_7382 (NA), Etr_899 (NA)	2/10	0.9391	0.004

[a]Cross-validation consistency is defined as the number of times the same model is identified in all 10 training data sets.
[b]Testing balanced accuracy is ((TP/(TP + FN)) + (TN/(TN + FP))/2, where TP = true positive, FP = false positive, TN = true negative and FN = false negative.
*Statistical significance was evaluated by comparing the testing balanced accuracy from the observed data to the distribution under the null hypothesis of no associations from 1,000 permutations.

results involving loci on the same linkage group and body metrics of adult Pacific lamprey in the Columbia River basin (Hess et al., 2015, 2014). This strong association of linkage group A loci with body size is remarkable because of the consistent strength of its associations observed across different geographic regions (Columbia River and Klamath River). It can be difficult to disentangle the intercorrelations among measured phenotypes and sets of genetic markers, and so these studies confirming consistent, and very strong, associations are critical and illustrate that selection on body size may occur throughout the geographic range of Pacific lamprey.

Improved understanding of the relationship between egg mass and body size, and the genetic underpinnings of these traits, is important for elucidating the precise nature of ecotypic variation in Pacific lamprey. Egg mass generally exhibits a positive relationship with body size in anadromous fishes (Hearsey & Kinziger, 2015; McGurk, 2000), and due to this intercorrelation, it would be expected that egg mass and body size would be associated with the same genetic markers. However, our results suggest a more complex scenario. We surprisingly found that body mass and egg mass were not positively related to one another in Pacific lamprey (correlation coefficient 0.14; $p > 0.05$; Table 2), but there is evidence for a common genetic basis for both traits on linkage group B (Egg mass: this study; body size: Hess et al., 2014; Hess et al., 2015; gene interaction analysis this study) as well as evidence that these traits are associated with a separate genetic regions (body size: linkage group A; egg mass: linkage group D).

Across the geographic range of Pacific lamprey ecotypic variation has been shown to be related to body size, egg mass and coloration, defining ecotypes termed normal and dwarf (Hess et al., 2013; Kostow, 2002), day eels and night eels (Close, Jackson, Conner, & Li, 2004), normal and praecox (i.e., dwarf parasitic) (Docker, 2009) and river- and ocean-maturing (Clemens et al., 2013; this study). Many of these ecotypes appear to be restricted in geographic distribution. Given that many of these ecotypes are defined by the same traits (e.g., body size and egg mass), we hypothesize that the same gene regions are responsible for producing the different ecotypes but that regional environmental factors play an important role in mediating

the ecotype observed. Importantly, it seems that considering epistatic effects and gene networks will be critical when researching the genetic basis of complex traits in Pacific lamprey and other species.

At a much broader scale, our findings show that evolutionary forces have worked in parallel to drive a portfolio of ecotypic diversity in both Pacific lamprey and steelhead, two species representing highly divergent taxonomic groups but that share an anadromous life history and a similar geographic distribution. However, the ecotypic adaptations and their genetic underpinnings within each species appear to be distinctive. For example, the Pacific lamprey maturation ecotypes initiate freshwater migration simultaneously and the genetic basis is polygenic, whereas in steelhead the summer- and winter-run initiate freshwater migration in different seasons and the genetic underpinnings involve a single locus of major effect.

In recognition of the cultural importance of Pacific lamprey to fishing tribes of northwestern North America, we recommend distinguishing the river-maturing and ocean-maturing ecotypes by adopting the names ke'ween (lamprey "eel") and tewol (ocean), respectively, using terms from the Yurok language.

ACKNOWLEDGEMENTS

Andre Buchheister, Darren Ward and Damon Goodman provided comments on earlier drafts of this manuscript. Travis Jacobson supported SNP genotyping. Leslie Rodelander and Lonyx Landry provided logistical support. Josh Murphy and William Ihne contributed photographs used in Figure 2. Tingting Hou and Xiang-Yang Lou provided assistance with GMDR. The methods used in this study were reviewed and approved by the Humboldt State University Institutional Animal Care and Use Committee (No. 15/16.F.105-A, 11 May 2016). This project was funded by a National Science Foundation-Graduate Research Fellowship and a Robert & Patricia Switzer Foundation Fellowship to Keith Parker.

AUTHOR CONTRIBUTIONS

All authors jointly conceived and designed this study. K.A.P. performed field collections, genotyping and data analysis. S.R.N. coordinated genotyping. A.P.K. and J.E.H. assisted with data analysis. K.A.P. wrote the first version of the manuscript, with subsequent contribution by the coauthors.

DATA ACCESSIBILITY

Trait data and SNP genotypes from GT-seq are provided as a supplementary file: Genotype and Trait Data for Pacific Lamprey maturation ecotypes.csv.

ORCID

Jon E. Hess https://orcid.org/0000-0002-3643-202X
Shawn R. Narum https://orcid.org/0000-0002-7470-7485
Andrew P. Kinziger https://orcid.org/0000-0002-8776-6230

REFERENCES

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Anglin, D. R. (1994). *Lower Klamath River instream flow study: Scoping evaluation for the Yurok Indian Reservation*. Vancouver, WA: USFWS Lower Columbia River Fishery Resource Office.

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., … Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*, 405–408. https://doi.org/10.1038/nature16062

Beamish, R. J. (1980). Adult biology of river lamprey and the Pacific lamprey from the Pacific coast of Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*, 1906–1923.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes. *Journal of Fish Biology*, *89*, 2519–2556. https://doi.org/10.1111/jfb.13145

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Callaway, E. (2017). Genome studies attract criticism. *Nature*, *546*, 463.

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, *15*, 855–867. https://doi.org/10.1111/1755-0998.12357

Chen, G. B., Xu, Y., Xu, H. M., Li, M. D., Zhu, J., & Lou, X. Y. (2011). Practical and theoretical considerations in study design for detecting gene-gene interactions using MDR and GMDR approaches. *PLoS ONE*, *6*, e16981. https://doi.org/10.1371/journal.pone.0016981

Clemens, B. J., Schreck, C., Sower, S. A., & van de Wetering, S. (2016). The potential roles of river environments in selecting for stream- and ocean-maturing Pacific lamprey. In A. Orlov & R. Beamish (Eds.), *Jawless fishes of the world* (pp. 299–322). Cambridge, UK: Scholars Publishing.

Clemens, B. J., van de Wetering, S., Sower, S. A., & Schreck, C. B. (2013). Maturation characteristics and life-history strategies of the Pacific lamprey, *Entosphenus tridentatus*. *Canadian Journal of Zoology*, *91*, 775–788.

Clemens, B. J., Wyss, L., McCoun, R., Courter, I., Schwabe, L., Peery, C., … Docker, M. F. (2017). Temporal genetic population structure and interannual variation in migration behavior of Pacific lamprey *Entosphenus tridentatus*. *Hydrobiologia*, *794*, 223–240. https://doi.org/10.1007/s10750-017-3096-4

Close, D. A., Fitzpatrick, M., Li, H., Parker, B., Hatch, D., & James, G. (1995). *Status report of the Pacific lamprey* (Lampetra tridentata*) in the Columbia River Basin*. Portland, OR: Bonneville Power Administration.

Close, D. A., Fitzpatrick, M. S., & Li., H. W., (2002). The ecological and cultural importance of a species at risk of extinction, Pacific lamprey. *Fisheries*, *27*, 19–25. https://doi.org/10.1577/1548-8446(2002)027<0019:TEACIO>2.0.CO;2

Close, D. A., Jackson, A. D., Conner, B. P., & Li, H. W. (2004). Traditional ecological knowledge of Pacific lamprey (*Entosphenus tridentatus*) in northeastern Oregon and southeastern Washington from indigenous peoples of the Confederated Tribes of the Umatilla Indian Reservation. *Journal of Northwest Anthropology*, *38*, 141–162.

Cordell, H. J. (2002). Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*, *11*, 2463–2468. https://doi.org/10.1093/hmg/11.20.2463

Docker, M. F. (2009). A review of the evolution of nonparasitism in lampreys and an update of the paired species concept. In L. Brown, S. Chase, M. Mesa, R. Beamish, & P. Moyle (Eds.), *Biology, management, and conservation of lampreys in North America* (pp. 71–114). Bethesda, Maryland:American Fisheries Society.

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Endelman, J. B., & Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genetics*, *2*, 1405–1413. https://doi.org/10.1534/g3.112.004259

Gola, D., John, J. M., van Steen, K., & Konig, I. R. (2016). A roadmap to multifactor dimensionality reduction methods. *Briefings in Bioinformatics*, *17*, 293–308. https://doi.org/10.1093/bib/bbv038

Goodman, D. H., Reid, S. B., Docker, M. F., Hass, G. R., & Kinziger, A. P. (2008). Mitochondrial DNA evidence for high levels of gene flow among populations of a widely distributed anadromous lamprey *Entosphenus tridentatus* (Petromyzontidae). *Journal of Fish Biology*, *72*, 400–417. https://doi.org/10.1111/j.1095-8649.2007.01731.x

Hearsey, J., & Kinziger, A. P. (2015). Diversity in sympatric Chinook salmon runs: Timing, relative fat content and maturation. *Environmental Biology of Fishes*, *98*, 413–423. https://doi.org/10.1007/s10641-014-0272-5

Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., & Narum, S. R. (2013). Population genomics of Pacific Lamprey: Adaptive variation in a highly dispersive species. *Molecular Ecology*, *22*, 2898–2916. https://doi.org/10.1111/mec.12150

Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., … Narum, S. R. (2015). Use of genotyping-by-sequencing data to develop a high-throughput and multi-functional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, *15*, 187–202. https://doi.org/10.1111/1755-0998.12283

Hess, J. E., Caudill, C. C., Keefer, M. L., McIlraith, B. J., Moser, M. L., & Narum, S. R. (2014). Genes predict long distance migration and large body size in a migratory fish, Pacific lamprey. *Evolutionary Applications*, *7*, 1192–1208. https://doi.org/10.1111/eva.12203

Hess, J. E., Zendt, J. S., Matala, A. R., & Narum, S. R. (2016). Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20153064.

Kostow, K. (2002). *Oregon lampreys: Natural history, status, and analysis of management issues*. Portland, OR: Oregon Department of Fish and Wildlife.

Larson, Z. S., & Belchik, M. R. (1998). *A preliminary status review of eulachon and Pacific lamprey in the Klamath River basin*. Klamath, CA: Yurok Tribal Fisheries Program.

Li, P., Guo, M., Wang, C., Liu, X., & Zou, Q. (2015). An overview of SNP interactions in genome-wide association studies. *Briefings in Functional Genomics*, *14*, 143–155. https://doi.org/10.1093/bfgp/elu036

Lou, X.-Y., Chen, G. B., Yan, L., Ma, J. Z., Zhu, J., Elston, R. C., & Li, M. D. (2007). A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *The American Journal of Human Genetics*, *80*, 1125–1137. https://doi.org/10.1086/518312

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates Inc.

McGurk, M. D. (2000). Comparison of fecundity-length-latitude relationships between nonanadromous (kokanee) and anadromous sockeye salmon (*Oncorhynchus nerka*). *Canadian Journal of Zoology*, *78*, 1791–1805.

McMahon, B. J., Teeling, E. C., & Hoglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, *7*, 999–1007. https://doi.org/10.1111/eva.12193

Meirmans, P. G. (2012). AMOVA-based clustering of population genetic data. *Journal of Heredity*, *103*, 744–750. https://doi.org/10.1093/jhered/ess047

Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, *4*, 792–794. https://doi.org/10.1111/j.1471-8286.2004.00770.x

Micheletti, S., Hess, J. E., Zendt, J. S., & Narum, S. R. (2018). Selection at a genomic region of major effect is responsible for complex life histories in anadromous steelhead. *BMC Evolutionary Biology*, *18*, 140.

Miko, I. (2008). Epistasis: Gene interaction and phenotype effects. *Nature Education*, *1*, 197.

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

Murphey, E. V. A. (1959). *Indian uses of native plants (3rd printing)*. Fort Bragg, CA: Mendocino County Historical Society.

Myers, J. M., Busack, C., Rawding, D., Marshall, A. R., Teel, D. J., Van Doornik, D. M., & Maher, M. T. (2006). *Historical population structure of Pacific salmonids in the Willamette River and lower Columbia River basins. NOAA technical memorandum NMFS-NWFSC-73*. Seattle, WA: NOAA NWFSC.

Narum, S. R. (2006). Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics*, *7*, 783–787. https://doi.org/10.1007/s10592-005-9056-y

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*, 2841–2847. https://doi.org/10.1111/mec.12350

Narum, S. R., Genova, A. D., Micheletti, S. J., & Maass, A. (2018). Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20180935. https://doi.org/10.1098/rspb.2018.0935

Pavey, S. A., Gaudin, J., Normandeau, E., Dionne, M., Castonguay, M., Audet, C., & Bernatchez, L. (2015). RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Current Biology*, *25*, 1666–1671. https://doi.org/10.1016/j.cub.2015.04.062

Pearse, D. E., Miller, M. R., Abadía-Cardoso, A., & Garza, J. C. (2014). Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20140012.

Petersen, R. S. (2006). *The role of traditional ecological knowledge in understanding a species and river system at risk: Pacific lamprey in the lower Klamath basin*. Unpublished Master's thesis, Oregon State University.

Petersen-Lewis, R. S. (2009). Yurok and Karuk traditional ecological knowledge: Insights into pacific lamprey populations of the lower Klamath basin. In L. Brown, S. Chase, M. Mesa, R. Beamish, & P. Moyle (Eds.), *Biology, management, and conservation of lampreys in North America* (pp. 1–39). Bethesda, Maryland: American Fisheries Society.

Phillips, P. C. (2008). Epistasis – The essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*, 855–867. https://doi.org/10.1038/nrg2452

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., … Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, *3*, e1603198. https://doi.org/10.1126/sciadv.1603198

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

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, *43*, 223–225. https://doi.org/10.1111/j.1558-5646.1989.tb04220.x

Ritchie, M. D., Hahn, L. W., & Moore, J. H. (2003). Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genetic Epidemiology*, *24*, 150–157. https://doi.org/10.1002/gepi.10218

Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F., & Moore, J. H. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *American Journal of Human Genetics*, *69*, 138–147. https://doi.org/10.1086/321276

Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, *465*, 609–612. https://doi.org/10.1038/nature09060

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*, 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Smith, J. J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M. C., Parker, H. J., … Amemiya, C. T. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nature Genetics*, *50*, 270–277. https://doi.org/10.1038/s41588-017-0036-1

Spice, E. K., Goodman, D. H., Reid, S. B., & Docker, M. F. (2012). Neither philopatric nor panmictic: Microsatellite and mtDNA evidence suggests lack of natal homing but limits to dispersal in Pacific lamprey. *Molecular Ecology*, *21*, 2916–2930. https://doi.org/10.1111/j.1365-294X.2012.05585.x

Thompson, T. Q., Bellinger, M. R., O'Rourke, S. M., Prince, D. J., Stevenson, A. E., Rodrigues, A. T., … Miller, M. R. (2019). Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 177–186. https://doi.org/10.1073/pnas.1811559115

Waples, R. S. (1991). Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. *Marine Fisheries Review*, *53*, 11–22.

Waples, R. S., & Lindley, S. T. (2018). Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evolutionary Applications*, *11*, 1518–1526. https://doi.org/10.1111/eva.12687

Yang, C., Zengyou, H., Wan, X., Yang, Q., Xue, H., & Yu, W. (2009). SNPHarvester: A filtering-based approach for detecting epistatic interactions in genome-wide association studies. *Bioinformatics*, *25*, 504–511. https://doi.org/10.1093/bioinformatics/btn652

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Parker KA, Hess JE, Narum SR, Kinziger AP. Evidence for the genetic basis and epistatic interactions underlying ocean- and river-maturing ecotypes of Pacific Lamprey (*Entosphenus tridentatus*) returning to the Klamath River, California. *Mol Ecol*. 2019;00:1–15. https://doi.org/10.1111/mec.15136