Received: 1 April 2020 | Revised: 26 July 2020 | Accepted: 14 August 2020

DOI: 10.1111/mec.15605

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Genomic islands of divergence infer a phenotypic landscape in Pacific lamprey

Jon E. Hess[1] | Jeramiah J. Smith[2] | Nataliya Timoshevskaya[2] | Cyndi Baker[3] | Christopher C. Caudill[4] | David Graves[1] | Matthew L. Keefer[4] | Andrew P. Kinziger[5] | Mary L. Moser[6] | Laurie L. Porter[1] | Greg Silver[1] | Steven L. Whitlock[7] | Shawn R. Narum[8]

[1]Columbia River Inter-Tribal Fish Commission, Portland, OR, USA

[2]Department of Biology, University of Kentucky, Lexington, KY, USA

[3]Oregon Department of Fish and Wildlife, Clackamas, OR, USA

[4]University of Idaho, Moscow, ID, USA

[5]Humboldt State University, Arcata, CA, USA

[6]NOAA NWFSC, Seattle, WA, USA

[7]University of Washington, Seattle, WA, USA

[8]Columbia River Inter-Tribal Fish Commission, Hagerman, ID, USA

Correspondence
Jon E. Hess, Columbia River Inter-Tribal Fish Commission, 700 NE Multnomah Street Suite 1200, Portland, OR 97232, USA.
Email: hesj@critfc.org

Funding information
Bonneville Power Administration, Grant/Award Number: 2008-524-00

Abstract

High rates of dispersal can breakdown coadapted gene complexes. However, concentrated genomic architecture (i.e., genomic islands of divergence) can suppress recombination to allow evolution of local adaptations despite high gene flow. Pacific lamprey (*Entosphenus tridentatus*) is a highly dispersive anadromous fish. Observed trait diversity and evidence for genetic basis of traits suggests it may be locally adapted. We addressed whether concentrated genomic architecture could influence local adaptation for Pacific lamprey. Using two new whole genome assemblies and genotypes from 7,716 single nucleotide polymorphism (SNP) loci in 518 individuals from across the species range, we identified four genomic islands of divergence (on chromosomes 01, 02, 04, and 22). We determined robust phenotype-by-genotype relationships by testing multiple traits across geographic sites. These trait associations probably explain genomic divergence across the species' range. We genotyped a subset of 302 broadly distributed SNPs in 2,145 individuals for association testing for adult body size, sexual maturity, migration distance and timing, adult swimming ability, and larval growth. Body size traits were strongly associated with SNPs on chromosomes 02 and 04. Moderate associations also implicated SNPs on chromosome 01 as being associated with variation in female maturity. Finally, we used candidate SNPs to extrapolate a heterogeneous spatiotemporal distribution of these predicted phenotypes based on independent data sets of larval and adult collections. These maturity and body size results guide future elucidation of factors driving regional optimization of these traits for fitness. Pacific lamprey is culturally important and imperiled. This research addresses biological uncertainties that challenge restoration efforts.

KEYWORDS

genomic islands of divergence, highly dispersive species, local adaptations

1 | INTRODUCTION

Highly dispersive species such as Pacific lamprey (*Entosphenus tridentatus*) present an evolutionary conundrum for adaptation. Adaptation is facilitated when particular combinations of gene variants (i.e., coadapted gene complexes) that confer optimal fitness in an environment can be passed on to the next generation. The action of recombination to rearrange genes operates regardless of the rate of gene flow. However, convention suggests that with high rates of gene flow recombination breaks down coadapted gene complexes which in turn disrupts local adaptation. Yet there is evidence from Pacific lamprey and other dispersive species that local adaptation may occur despite these high rates of gene flow. For example, Pacific lamprey body size is correlated with upstream migration distance in the Columbia River (Hess et al., 2014; Keefer, Moser, Boggs, Daigle, & Peery, 2009) and traits in other dispersive species appear to be optimized for specific environments within their broader range (Asaduzzaman et al., 2019; Miller et al., 2019; Phair, Toonen, Knapp, & von der Heyden, 2019).

Genomic architecture appears to be one factor that can influence local adaptation in highly dispersive species. In general, the closer two genes occur in the genome the smaller the chance for recombination events that may separate an optimal combination of variants (Yeaman & Whitlock, 2011). Inversions suppress recombination between inverted haplotypes and can effectively lock an optimal combination of variants together over longer distances within the inverted segment (sometimes referred to as a supergene). The fitness conferred by these inversions can help maintain them as a polymorphism in a population through both forces of balancing and divergent selection (Faria, Johannesson, Butlin, & Westram, 2019; Wellenreuther & Bernatchez, 2018). In Pacific lamprey, if there are particular phenotypes that have a polygenic basis and confer differential fitness across environments, we might expect these genes to be maintained together within inverted segments which would appear as long polymorphic intervals of DNA sequence.

Several traits in Pacific lamprey have been found to have a genetic basis. These include body size, reproductive migration-timing (Hess et al., 2014, 2015), and advanced maturity of females at onset of freshwater migration (i.e., ocean-maturing versus river-maturing ecotypes, Parker, Hess, Narum, & Kinziger, 2019). There also appears to be evidence for statistical linkage of multiple loci that show high divergence in the species' range (Hess, Campbell, Close, Docker, & Narum, 2013). One thing that is unclear is whether range-wide divergence that has been observed can be explained by phenotype-by-genotype associations reported thus far. Phenotypic traits are often interrelated, which can obscure the true target of selection (Powell & MacGregor, 2011). Testing a large variety of phenotypic trait associations with genotypes at different sites in the species' range can help to disentangle these correlations and help elucidate the true target of selection (Chanock et al., 2007). Once phenotype-by-genotype associations are confirmed across geographic sites, these relationships can be exploited to extrapolate a phenotype across large geographic areas in which only genotypes have been measured. This genetic tool then becomes a powerful predictor and can generate hypothesis testing frameworks to guide future studies aimed to validate these predicted phenotypic distributions across the range and elucidate factors driving regional optimization of these traits.

Pacific lamprey is a culturally important and imperiled anadromous fish with a parasitic ocean phase distribution in the North Pacific Ocean and a freshwater distribution throughout the Pacific Northwest and Alaska (Close, Fitzpatrick, & Li, 2002; Orlov, Beamish, Vinnikov, & Pelenev, 2009). There are still a number of biological uncertainties that remain before complete characterization of the life history diversity and complex life cycle of Pacific lamprey is achieved, especially for the marine phase of its life cycle (Clemens et al., 2019). Basic biological information and life-history research is needed to explain observed trait variation and inform management actions that are being implemented to help restore abundance of the species within the Columbia River basin (an area that covers Washington, Oregon, and Idaho of the Pacific Northwest). For example, proactive conservation measures such as translocations in which adults are transferred from a source of abundance to a target site of low abundance are currently being implemented in the interior Columbia River (Ward et al., 2012). Information on which life history traits are regionally optimized could help inform these types of actions (Hess et al., 2014).

In this study we addressed four major objectives: (a) Divergence mapping: Test whether previously observed genomic divergence across the species' range is either concentrated or diffusely organized in the genome; (b) Association testing: Test phenotypic trait associations with genotypes across geographic sites to identify robust phenotype-by-genotype relationships; (c) Association mapping: Test whether phenotypic-by-genotype associations mapped to the genome can explain genomic divergence across the species' range; and (d) Extrapolation of spatiotemporal phenotypic distributions: Use candidate SNP genotypic distributions across time and space to characterize the ecological conditions that may favor particular life history traits. Our findings supported a high concentration of genomic divergence to regions within four chromosomes, referred to as genomic islands. Two of these four genomic islands showed robust correlation with maturity and body-size traits and could be used to predict their spatiotemporal distributions across the species' range.

2 | MATERIALS AND METHODS

2.1 | Divergence mapping

Two new Pacific lamprey genome assemblies (83 linkage groups/chromosomes) were constructed using the whole genome sequence from the milt and blood from a male (representing the gametic and somatic genomes; Genbank Accession #: JAAVTP010000000) and the blood of a female (Genbank Accession #: JAAXLI010000000),

TABLE 1 Summary of data sets analysed in this study

Data set	Description	Life stage	Collection year (s)	Latitude	Longitude	*N*	Loci	Traits	Objective
Range-wide F_{ST}	16 N > 20 collections (Hess et al., 2013)	Adult/Larvae	1995–2011	16 sites		482	7,716	None	Identify range-wide F_{ST} outliers
JDD	John Day Dam	Adult	2014–2015	45.714800	−120.693700	656	302	1,2,3	Association testing
S_BON	Bonneville Dam Flume Experiment	Adult	2014	45.644300	−121.940600	295	302	1,2,3,6	Association testing
T_BON	Bonneville Dam Adult Fish Facility	Adult	2014	45.644300	−121.940600	883	302	1,2,3	Association testing
WFA♀	Willamette Falls Harvest	Adult	2016	45.351100	−122.619300	133	302	1,2,5	Association testing
WFA♂	Willamette Falls Harvest	Adult	2016	45.351100	−122.619300	136	302	1,2,5	Association testing
GAR	Common garden experiment	Larvae	2015	45.669007	−118.620700	337	302	4	Association testing
WFA2014	Willamette Falls fish ladder	Adult	2014	45.351100	−122.619300	868	302	1,2	Temporal distribution of genes
WFA2015	Willamette Falls fish ladder	Adult	2015	45.351100	−122.619300	581	302	1,2	Temporal distribution of genes
Larval/juvenile range-wide	Primarily larvae and juveniles from unique families (Table S1)	Larvae	2011–2015	57 sites		2,581	85	none	Spatial distribution of genes

Note: The measured traits were categorized into the following groups: (a) Body size - length, weight, girth, and interdorsal distance; (b) migration timing (day of arrival); (c) migration distance (distance in river kilometer travelled upstream from collection site); (d) growth following hatch; (e) maturity (measured by gonad weight); (f) swimming performance. All data sets were newly genotyped in this study with the exception of the range-wide F_{ST} data set (from Hess et al., 2013) and a portion of the Laval/juvenile range-wide data set (portions genotyped by Hess et al., 2013 and Hess et al., 2015, Table S1).

and using a high density linkage map (Smith et al., 2018) to validate and extend higher order scaffolding of chromosomes (Appendix S1).

For characterization of SNP densities and F_{ST} statistics, we used a set of 7,716 unique SNP loci from previously published RAD-seq data sets (Hess et al., 2013; Smith et al., 2018), which passed a set of population genetic quality control filters (Appendix S1). This set of 7,716 unique SNPs was a combination of overlapping groups of SNPs from a previous data set (Hess et al., 2013; SNPs N = 8,772) and a de novo linkage mapping data set (Smith et al., 2018; SNPs N = 7,977). BOWTIE2 (Langmead & Salzberg, 2012) was used to align these two data sets to the male reference assembly to define homologous loci. For the 7,716 total SNPs passing the QC filters, 4,046 loci were unique to Hess et al. (2013), 1,418 loci were unique to Smith et al. (2018), and 2,252 SNPs were shared across data sets. Marker positions based on BOWTIE2 alignments were compared between Pacific lamprey male and female genomes and the Pacific lamprey male and sea lamprey (*Petromyzon marinus*) male gametic genome (GenBank assembly accession: GCA_002833325.1) to characterize synteny.

Using these 7,716 SNPs genotyped for the same individuals from Hess et al. (2013; i.e., 16 collections with >20 individuals which totalled 482 individuals; Table 1), LOSITAN (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008) was run using parameter settings of 50,000 simulations, confidence interval of 0.99, false discovery rate set to 0.1, subsample size of 20, simulated F_{ST} of 0.019 and an attempted F_{ST} of 0.021. We considered loci candidates for positive selection (adaptive loci) above a probability level of 0.995, and neutral loci were defined as falling between the 10th and 90th quantiles of the F_{ST} distribution. Any remaining SNPs were conservatively considered undetermined (neither adaptive nor neutral).

Genes located within adaptive regions were identified using published sea lamprey gene annotations that were found in the homologous regions corresponding to the following Pacific lamprey male genome positions (a) chromosome 01 positions: 8,939,466…14,772,759 (sea lamprey scaf_00003: 6,777,250…13,554,086); (b) chromosome 02 positions: 3,351,206…18,794,404 (sea lamprey scaf_00006: 1,198,871–13,859,281); (c) chromosome 04 positions: 6,408,032…19,202,839 (sea lamprey scaf_00005: 2,591,251…16,864,119); and (d) chromosome 22 positions: 617,460…11,364,740) (sea lamprey scaf_00012: 1,160,196…12,993,068). We used the website Enrichr (https://amp.pharm.mssm.edu/Enrichr/) (Kuleshov et al., 2016) to gain insights into the potential function of these genes via both the manifested phenotypes in mammals (i.e., MGI Mammalian Phenotype Level 4 2019) and in fishes (FishEnrichr; Phenotype AutoRIF Predicted Z score).

2.2 | Association testing

Genotyping-in-thousands by sequencing (GT-seq, Campbell, Harmon, & Narum, 2015) was employed to genotype 308 genetic markers for the association testing analyses. The GT-seq 308 loci

were 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 detail in Appendix S1. Locus selection began with a group of 457 total SNP loci considered in round 1, which included 120 for which TaqMan assays had already been designed (Hess et al., 2015). Final optimization left 308 loci that worked best in GT-seq genotyping. For all data sets used below in the association testing we filtered out individuals missing >10% of genotypes at the 308 loci. Excluding the four species diagnostic loci and two duplicated loci provided 302 unique loci for association tests.

There were six data sets, five comprised of adults (JDD, S_BON, T_BON, WFA♀, and WFA♂) and one comprised of larvae (GAR), with which we performed association testing (Table 1). Adult data sets were from the following three locations: males (WFA♂, $N = 136$) and females (WFA♀, $N = 133$) from Willamette Falls collected in 2016 (Willamette River, Oregon City, OR; 205.6 Rkm upstream from the Columbia River mouth), two data sets (S_BON, $N = 295$ and T_BON, $N = 883$) from Bonneville Dam in 2014 (235.1 Rkm upstream from the Columbia River mouth), and one data set (JDD, $N = 656$) from John Day Dam in 2014 and 2015 (346.9 Rkm upstream from the Columbia River mouth). The following five adult traits were measured on all adult data sets: ordinal "day" of collection (timing of migration to the sample point), girth (mm), total "length" (mm), weight (g), and distance between dorsal fins ("interdorsal", mm). Interdorsal measurements have been suggested to serve as an indicator of maturation status in Pacific lamprey because the distance tends to decrease with maturation (Clemens, van de Wetering, Kaufman, Holt, & Schreck, 2009). We measured an additional migration trait for three adult data sets (S_BON, T_BON, and JDD) via a combination of passive integrated transponder (PIT) and radio tagging of individual fish and observing their furthest upstream detection from the release location ("Rkm"). Further, since the males and females collected at Willamette Falls (WFA♂ and WFA♀) were being harvested, we were able to measure gonad weight as a proxy for maturity in those data sets. Finally, a subset of the adult data set from Bonneville Dam (S_BON) was used in a swim trial experiment within a flume (Kirk, Caudill, Tonina, & Syms., 2016), in which the following three swimming behavioral traits were measured: "approached" experiment, passed challenge ("pass"), and passed challenge without fallback ("passrep"). Details of these swimming performance experiments can be found in Kirk et al. (2016) and Appendix S1.

A single group of larvae were artificially propagated using adults captured at Bonneville Dam. These larvae were reared in a common garden experiment to generate early larval growth ("GAR") rate data ($N = 337$). All larvae were spawned in the spring of 2015 and allowed to rear from 30 to 163 days after hatching at which point they were measured for growth. Growth rate was measured as length/time ("growth"), and also corrected growth rate ("growth rate_b"; [length − 4 mm]/time) to correct for length at hatch (~4 mm).

Intercorrelation among all measured traits in these six data sets (i.e., JDD, S_BON, T_BON, WFA♀, WFA♂, and GAR) was examined (based on Pearson's r) to avoid excessive redundancy of predictor variables ($|r| > .95$), and p-values were calculated (SAS Institute & Inc., 2000). We performed univariate analyses using a general linear model (GLM) and a mixed linear model (MLM) with TASSEL v. 5.1.0 (Bradbury et al., 2007). The GLM is a fixed effects linear model that is used in TASSEL to identify significant associations between phenotypes and genotypes. TASSEL takes population structure into account by using genetic principal coordinate axes as covariates in the model. The MLM is similar to GLM but includes both fixed effects (e.g., population structure, and genetic marker) and random effects (i.e., relationships among individuals) and can thus account for both population structure and kinship to reduce false positive associations (Yu et al., 2006). Details on the covariates and ways in which loci were used taking population structure and relatedness into account in the GLM and MLM tests are provided in the Appendix S1. 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). Critical values were calculated using the function p.adjust within the R package stats (R Core Team, 2019).

2.3 | Association mapping

The 308 SNP loci in the GT-seq panel were aligned to reference genomes using BOWTIE2. There were 306 that were each assigned to a single location on the Pacific lamprey male genome (99.4%), covering 70 different chromosomes with an average of 4.4 loci per chromosome (range 1–22). Marker locations were based on the alignments of marker sequences to the Pacific lamprey male and female genomes, homologous scaffolds of the sea lamprey genome, and positions on the previously published Pacific lamprey linkage map (Smith et al., 2018).

Adjusted p-values from the association testing described above were log transformed (−LOG10) and plotted by consensus genome position on the Pacific lamprey male genome. We tested correlation of association tests −LOG10(P) with F_{ST} from the range-wide divergence to understand whether trait associations may explain the high divergence observed at the range-wide scale for the subset of markers shared between data sets. Among the 308 SNPs, there were 230 neutral SNPs, 41 adaptive markers SNPs, and a set of 31 "intermediate" SNPs that did not fit definitions of putatively neutral and putatively adaptive (divergence mapping). Finally, four loci were species-diagnostic, and 2 loci were duplicated. Therefore, there were 302 unique markers available for these association analyses. These markers included 38 SNPs that were mostly adaptive loci that were categorized into the following 4 groups of statistically linked loci: A ($N = 10$), B ($N = 13$), C ($N = 7$), and D ($N = 8$, Hess et al., 2013).

2.4 | Extrapolation of spatiotemporal phenotypic distributions

We characterized candidate SNP genotypic distributions across time and space to better understand the ecological conditions that may favour particular life history traits. These spatiotemporal distributions were characterized using the candidate SNPs with the most robust associations with body size (chromosome 02) and sexual maturity (chromosome 01). Distributions of representative SNPs of the other adaptive chromosome regions (chromosomes 04 and 22) were also characterized (Figures S1 and S2).

We used three independent data sets to characterize spatial and temporal distributions of genetic variation (Table 1). These data sets were independent of each other and separate from the association testing data sets, and they were optimally suited for these characterizations. For the spatial data set, we primarily used collections of larvae and juveniles (95% of data set of $N = 3,435$) but included some adult collections that were distributed widely across the species' range. Larvae and juveniles were the ideal life stage to represent genotypic distributions of individuals that successfully spawned at discrete locations throughout the range. Adult collections were used to fill in portions of the range where larval samples were not available. Genotyping was partially conducted with a TaqMan assay panel (Hess et al., 2015), which overlapped the GT-seq panel by 85 SNPs they had in common. COLONY v. 2.0.6.5 (Jones & Wang, 2010) was used to reconstruct full-sibling families (Wang, 2004) using the 85 shared SNPs on each of the 70 collections. We analysed all collections together as one using the following parameter settings: polygamous mating for males and females without inbreeding, full-likelihood, medium length of run, no allele updating, and no sibship priors. Only one collection out of the seven adult collections had full siblings ($N = 13$, Stamp River, B.C.) which were maintained to accurately represent this small spawning segment. We excluded duplicate genotypes, 797 full siblings, and collections with fewer than five individuals, resulting in a final set of 57 collections consisting of a total of 2,581 individuals each representing a unique family (Table S1). This data set was then used to calculate allele frequencies across collections for the representative candidate SNPs Etr_464 and Etr_5317 within the adaptive regions on chromosomes 01 and 02, respectively.

For the temporal data sets, we used individuals collected from two successive spawning runs at Willamette Falls (2014–2015; N of 868 and 581, respectively) over which it was possible to randomly sample the majority of the annual adult migration of Pacific lamprey (typically February–August) in weekly strata. A daily abundance estimate (Whitlock, Deweber, & Peterson, 2019) was used to expand candidate SNP genotypic proportions in the weekly strata. We estimated the abundance and 95% confidence intervals of the candidate genotypes using a bootstrapping method (Steinhorst, Copeland, Ackerman, Schrader, & Anderson, 2017) that was automated in R for a broad set of fisheries applications that require stratified sampling (Thomas Delomas, PSMFC/IDFG, https://github.com/delomast/fishCompTools)". One biological complexity was that a portion of the adults encountered before May probably overwintered and experienced shrinkage in body size due to advanced maturation (Beamish, 1980). Therefore, in addition to characterizing allele frequencies of candidate SNPs Etr_464 and Etr_5317, we categorized fish by body length to provide insight into the transition between overwintered fish and newly-arrived migrants.

3 | RESULTS

3.1 | Divergence mapping

Outlier analyses identified 311 (4.0%) SNPs as candidates for positive selection (out of a total of 7,716 SNPs; $p > .995$). LOSITAN was also used to identify neutral loci, which we defined using a conservative threshold range of probabilities between 0.10 and 0.90. There were 350 (4.5%) and 4 (<0.1%) SNPs below and above this range, respectively (i.e. candidates for balancing and positive selection, respectively), and 7,051 neutral loci (91.4% of 7,716 loci) that fell within these probability levels.

A total of 7,385 out of 7,716 loci (95.7%) and 7,366 out of 7,716 loci (95.4%) aligned to the Pacific Lamprey female and male genome assemblies, respectively, and 4,916 out of 7,716 loci (63.7%) aligned to the male gametic sea lamprey genome. The alignment to the Pacific lamprey male genome was used to order the loci by scaffold position, and in cases in which only alignments to the other assemblies were available we interpolated values to estimate relative positions. Manhattan plots were used to visualize the distribution of the outlier SNPs in both the Pacific lamprey and sea lamprey male genome assemblies (Figure 1), and alignments were generated between Pacific lamprey male and female genomes (Figure S3). These results illustrated that genomic divergence is highly concentrated as demonstrated by the fact that 65% of the outlier loci are localized to each of the following four chromosomes:01, 02, 04, and 22; which share homology with sea lamprey scaffolds (scaf_00003, scaf_00006, scaf_00005, and scaf_00012, respectively; Figure 1). The patterns of synteny within these four chromosomes indicated large regions of inversions that overlapped with concentrations of outlier SNPs (e.g., chromosomes 01 and 02, Figure S3) and may be polymorphic within Pacific lamprey given the differences between male and female genome assemblies. The same inversion patterns on chromosomes 01 and 02 were present between species (male assemblies, Figure S3).

3.2 | Association testing

Examination of the intercorrelation of predictor variables indicated that many of the morphological variables related to body size attributes were highly correlated, however, none had a significant correlation above 0.95 (Tables S2–S7), and therefore all were retained for association analysis. Significant intercorrelations among traits were consistent across data sets and using these correlations

we categorized traits into the following four main groups: (a) body size; (b) female sexual maturity; (c) larval growth; and (d) swimming ability. The "body size" category included the body metrics of length, weight, girth, and interdorsal, which were all significantly positively correlated across data sets (Tables S2–S7). This trait category also included migration distance which was positively correlated with increasing body size metrics, and migration timing which was negatively correlated with these body size metrics. The male gonad mass was intended to serve as a proxy for male sexual maturity, but it also was significantly positively correlated with other body size metrics. However, the female gonad metric was not correlated with the other body size metrics and was potentially an accurate proxy for "female sexual maturity". The "larval growth" category was populated by the only two measures of growth in the common garden experiment. The "swimming ability" category contained the three metrics of swimming performance which were all significantly positively correlated, and included migration day, which was positively correlated.

We examined the relative strength of associations of the total 302 SNPs with the traits within each of the four trait categories (Table S8) but were primarily interested in associations of SNPs on the four chromosomes 01, 02, 04 and 22, where evidence for range-wide adaptive divergence was concentrated.

3.2.1 | Body size traits

Highly significant associations (adjusted $p < .001$) were observed for body size traits and SNPs on the four adaptive regions across data sets (WFA♀, WFA♂, T_BON, S_BON, and JDD; Table 2). The strongest associations between SNPs and body size traits (−LOG(10P) > 30) were with length (Figure 2), weight, and girth and SNPs on chromosome 02 for the T_BON data set; however, the S_BON and JDD data sets also showed significant associations for the same SNPs and traits (−LOG(10P) > 2). This result was similar to the findings of Hess et al. (2014) where they show that a SNP (Etr_5317), herein

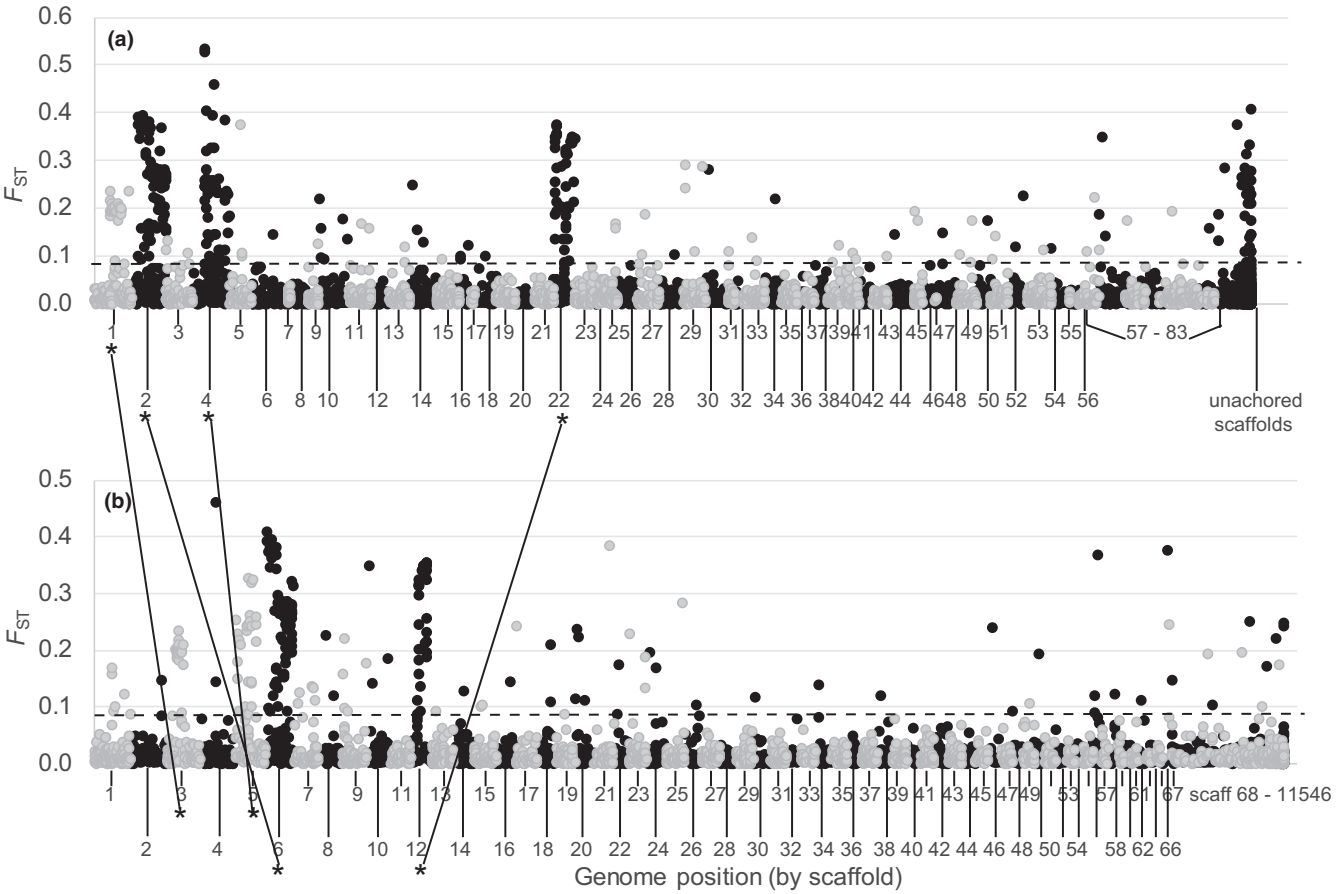

FIGURE 1 Manhattan plot of the Pacific lamprey SNPs aligned to the (a) Pacific lamprey male gametic ($N = 7{,}366$; 95.5% of total loci) and the (b) sea lamprey ($N = 4{,}916$; 63.7% of total loci) genome. F_{ST} values were generated in LOSITAN using 7,716 SNPs and the 16 collections with >20 individuals ($N = 518$ individuals). The critical F_{ST} value of 0.07 is indicated by the dashed line (SNPs above this line were considered *outliers* $p > .995$). SNPs are shown by alternating odd (black) and even (grey) linkage groups (a total of 83 chromosomes have been characterized in Pacific lamprey based on a linkage map and shown in numbers on this figure; the sea lamprey genome assembly is comprised mostly of super scaffolds (≥1 Mb) anchored to 94 linkage groups (chromosomes), Smith et al. (2018). The first 94 numbered sea lamprey scaffolds (scaff) represent defined linkage groups and the larger numbers are unanchored scaffolds. The asterisks indicate the synteny between Pacific lamprey and sea lamprey for concentrations of outlier loci

mapped to chromosome 02, had strongest association to body size traits at Bonneville Dam. In contrast, the WFA♀ and WFA♂ data sets had fewer total significant associations with body size traits across the four adaptive chromosome regions. The WFA♀ and WFA♂ data sets significant associations to body size traits were concentrated on chromosome 02 and chromosome 04, and of these chromosomes, SNPs on chromosome 04 appeared to have the strongest associations with body size traits, and primarily with length and weight. For WFA♂, similar to the length and weight traits, male gonad size was also associated with chromosome 04 and chromosome 02 SNPs, which is probably due to the high intercorrelation observed among these traits. Overall, the results support strong association of SNPs with body size traits, primarily length and weight, compared to other intercorrelated traits (e.g., migration timing and migration distance). The chromosome that showed the highest association with body size was chromosome 02 and these associations were greatest in the data sets from furthest upstream locations (T_BON, S_BON, and JDD). This chromosome 02 association with body size was consistent among data sets located both within and outside the Columbia River basin (Parker et al., 2019).

The genotypes in the chromosome 02 adaptive region (SNP Etr_5317) in the T_BON data set were also predictive of average lengths, such that the average size of homozygotes for large body size alleles "AA", heterozygotes "AC", and homozygotes for small body size alleles "CC" were 677, 627, and 592 mm, respectively (Figure 3a). Although the average body sizes differed across sites and sexes (WFA♀ were larger on average than WFA♂), the trends were consistent between data sets (Figure 3a). Further, similar genotype and average length associations have been detected in Pacific lamprey collected from the Klamath River (634, 602, and 557 mm for the AA, AC, and CC genotypes at Etr_5317, respectively; Parker et al., 2019).

3.2.2 | Female sexual maturity

Female gonad size was significantly associated with chromosome 01 and none of the other three adaptive regions in the WFA♀ collection (Table 2, Figure 2). This finding is concordant with Parker et al. (2019) who also resolved a significant association between female gonad size and chromosome 01. The average gonad sizes associated with genotypes at candidate SNP Etr_464 (chromosome 01) were 25, 20, and 18 g for the AA, AC, and CC genotypes, respectively (Figure 3b). While gonad mass was less in the Klamath River collection, categorizations by genotype were consistent (average egg mass of 13, 7, and 6 g for the AA, AC, and CC genotypes at Etr_464, respectively; Parker et al., 2019).

3.2.3 | Swimming ability and larval growth rate

No significant associations were observed on the four adaptive chromosome regions or any of the other chromosomes for the short-term swimming performance trials or the larval growth rates (Table 2).

3.3 | Genotypic prediction of phenotypic traits and potential gene-interaction effects

In most of the trait associations, the representative candidate SNPs chosen on the four adaptive chromosome regions (Etr_464, Etr_5317, Etr_1806, and Etr_4281 on chromosomes 01, 02, 04, and 22, respectively) represented above average genotype-by-phenotype associations of all 34 of the significant SNPs on these four chromosomes. In many cases, these four SNPs lie at the extreme end of the range of observed *p*-values (Table S8).

Parker et al. (2019) found evidence for epistatic interactions that involved loci on chromosomes 01 and 04 (referred to previously as linkage groups D and B, respectively), which were found to represent the model with highest predictive ability for the female maturity trait (or ocean- and river-maturing ecotypes). We used single SNP locus representatives for each chromosome and conducted gene interaction tests for the maturity and body-size candidate loci following Parker et al. (2019, Appendix S1). For both traits, the model with highest support was a single locus model such that Etr_464 (chromosome 01) and Etr_5317 (chromosome 02) were the loci with highest predictive ability for the female maturity and adult total length traits, respectively (Table S9).

3.4 | Overlap of genomic divergence and association mapping

The range-wide F_{ST} values that were mapped to the male Pacific lamprey genome were plotted by genomic position with the subset of SNPs used in the association testing for the two traits with consistent strong associations (i.e., adult total body length as measured at Bonneville Dam "T_BON" and female gonad size as measured at Willamette Falls "WFA♀"; Figure 2a,b). For chromosomes 01 and 02, the adjusted −LOG10(P) values from the association tests were highly correlated with the genomic divergence as measured by F_{ST} for the female gonad size and adult body size traits, respectively (Figure 2a, b). We quantified the overlap of genomic divergence and trait association by regressing range-wide F_{ST} and the adjusted −LOG10(P) values for the 302 SNPs in the GT-seq panel. These 302 SNPs show positive linear trends for both traits, but the linear trends with highest slope and R^2 were observed for SNPs on chromosomes 01 and 02 for the gonad size and body length traits, respectively (Figure S4). These results suggest that the high geographic divergence exhibited by SNPs on chromosomes 01 and 02 may be related to selection on the traits with which these same SNPs are highly associated.

3.5 | Functions enriched within adaptive chromosome regions

The mammalian phenotype terms showed some significant tests for enrichment based on the overlap of annotated genes from the four chromosome regions (Table S10). The top three terms with lowest

TABLE 2 The number of significant association tests on four evolutionarily important chromosomes for traits and datasets analyzed in this study

			Chromosome 01				Chromosome 02				Chromosome 04				Chromosome 22			
Category	Trait	Dataset	*N*	P_N	Avg.	Range	*N*	P_N	Avg.	Range	*N*	P_N	Avg.	Range	*N*	P_N	Avg.	Range
Body size	Length	JDD	13	0			12	5	3.0	2.0–3.7	22	4	1.4	1.4–1.4	7	2	2.3	1.9–2.6
		S_BON	13	0			12	8	6.7	2.3–13.1	22	7	2.1	1.5–2.3	7	5	4.4	2.1–5.3
		T_BON	13	9	6.6	6.4–6.8	12	8	29.4	15.5–35.2	22	12	10.0	1.4–15.5	7	5	6.6	5.4–8.2
		WFA♀	13	0			12	1	1.5	1.5–1.5	22	8	2.2	2–2.7	7	0		
		WFA♂	13	0			12	0			22	9	1.7	1.4–2.2	7	0		
	Weight	JDD	13	0			12	5	2.4	1.9–3.0	22	0			7	0		
		S_BON	13	0			12	8	7.3	1.9–12.0	22	7	2.1	1.7–2.4	7	3	1.6	1.4–1.7
		T_BON	13	9	5.9	5.7–6.2	12	8	31.9	17.0–38.5	22	12	8.5	1.4–13.6	7	4	2.3	1.5–2.8
		WFA♀	13	0			12	7	2.1	1.9–2.5	22	10	2.8	1.5–5.2	7	0		
		WFA♂	13	0			12	4	1.4	1.3–1.4	22	3	1.9	1.8–2	7	0		
	Girth	JDD	13	0			12	5	2.4	1.9–3.0	22	0			7	0		
		S_BON	13	0			12	7	6.6	5.0–10.0	22	5	1.6	1.5–1.6	7	4	1.9	1.4–2.3
		T_BON	13	9	4.4	4.0–4.5	12	8	29.2	15.1–34.9	22	11	6.6	1.4–9.4	7	4	2.1	1.6–2.8
		WFA♀	13	0			12	0			22	0			7	0		
		WFA♂	13	0			12	0			22	0			7	0		
	Dorsal	JDD	13	0			12	0			22	0			7	0		
		S_BON	13	9	1.5	1.4–1.6	12	8	3.1	2.4–3.8	22	0			7	0		
		T_BON	13	9	4.1	3.9–4.3	12	8	5.4	3.5–6.5	22	10	3.5	1.7–4.1	7	0		
		WFA♀	13	0			12	7	2.1	1.8–2.6	22	4	1.8	1.3–2.1	7	0		
		WFA♂	13	0			12	0			22	0			7	0		
	Day	JDD	13	0			12	0			22	0			7	0		
		S_BON	13	0			12	0			22	0			7	0		
		T_BON	13	9	1.8	1.4–2.2	12	0			22	0			7	0		
		WFA♀	13	0			12	0			22	0			7	0		
		WFA♂	13	0			12	0			22	0			7	0		
	Rkm	JDD	13	0			12	0			22	0			7	0		
		S_BON	13	0			12	0			22	0			7	0		
		T_BON	13	9	5.1	4.8–5.3	12	7	4.4	3.3–5.0	22	5	1.4	1.4–1.6	7	1	1.5	1.5–1.5
	Gonad	WFA♂	13	0			12	0			22	0			7	0		
Larval Growth	Growth	GAR	13	0			12	0			22	0			7	0		
	Growth_b	GAR	13	0			12	0			22	0			7	0		

(Continues)

TABLE 2 (Continued)

Category	Trait	Dataset	Chromosome 01				Chromosome 02				Chromosome 04				Chromosome 22			
			N	P_N	Avg.	Range	*N*	P_N	Avg.	Range	*N*	P_N	Avg.	Range	*N*	P_N	Avg.	Range
Maturity	Gonad	WFA♀	13	8	1.8	1.6–1.9	12	0			22	0			7	0		
Swimming	Approach	S_BON	13	0			12	0			22	0			7	0		
	Pass	S_BON	13	0			12	0			22	0			7	0		
	Passrep	S_BON	13	0			12	0			22	0			7	0		

Note: p-values were adjusted for multiple testing using the Benjamini and Hochberg (1995) false discovery rate and then transformed with −Log10(P). The −Log10(P) values were shaded light to dark to indicate critical values of 1.3, 2.0, and 3.0 corresponding to alpha levels of 0.050, 0.010, and 0.001, respectively. Included in this table are the total number of loci (*N*) genotyped on each chromosome, the number of loci with significant adjusted *p*-values (P N), and the "average" and "range" of −Log10 (P) values across loci with significant adjusted *p*-values. See Methods text for trait definitions. A more detailed Table is available in the Appendix S1.

p-value from our list of 98 candidate genes on chromosome 01 (which is associated with maturity) were abnormal social investigation, abnormal blood urea nitrogen level, and induced hyperactivity. The three top terms output for our list of 260 genes on chromosome 02 (which is associated with body size traits) were short tibia, decreased brown fat cell lipid droplet size, and Purkinje cell degeneration. The phenotypes in fishes did not show significant Fisher's exact tests for enrichment of terms, although for chromosome 01, the top three ranked phenotypic terms (based on combined score, Chen et al., 2013) may be relevant to the associations we observed in this study. These terms were reproductive behavior, response to absence of light, and entrainment of circadian clock by photoperiod (Table S11).

3.6 | Extrapolation of spatial and temporal distribution of phenotypic traits based on candidate SNPs

One general pattern observed consistently across candidate SNPs (e.g., Etr_464 on chromosome 01 and Etr_5317 on chromosome 02) was a divergence between coastal and interior collections (Figure 4). This pattern was most evident within the Columbia River basin, where both candidate SNPs increased in one allelic variant with increasing distance from the river mouth (Figure S5). The most dramatic increase was with the frequency of the allelic variant of Etr_5317 associated with large adult body-size ("A" allele) from ~10% at the river mouth to near fixation (~98%) upstream of river kilometer 644. A more moderate increase was observed for the allelic variation of Etr_464 associated with small gonad size (40%–95% shift in "C" allele from the river mouth to river kilometer 644). Similar clines were observed within the Willamette River subbasin of the Columbia River (Figure S5), such that strong linear trends were observed with a change in frequencies 53% to 77% (Etr_464, R^2 = .42) and 22% to 61% (Etr_5317, R^2 = .87) over the span of 180 river kilometers. We classified all collections of the spatial data set into putative "Mature" and "Premature" forms based on the whether the Etr_464 mature allele frequency was ≥50% or <50%, respectively (Table S1). We also classified all collections of the spatial data set into putative "Small" and "Large" body-size forms based on the whether the Etr_5317 large body-size allele frequency was <50% or ≥50%, respectively (Table S1).

Intra-annual temporal heterogeneity was observed at Willamette Falls among the adult Pacific lamprey returning in run years 2014 and 2015. The abundance of the AA genotype of Etr_464 (chromosome 01) associated with large gonad size arrived earlier than the CC genotype associated with small gonad size (Figure 5a, b). When the abundance of each run year was divided into equal halves, we estimated that the AA genotype decreased by 3× and 2× between the first and second halves of the run in 2014 and 2015, respectively (Figure 5a, b). This decrease was significant based on nonoverlapping 95% confidence intervals (C.I.) of the relative abundances estimated for individuals with the AA genotype in the first and second halves of the run in 2014 (i.e., 26.8% AA fish comprised the early run, 22.2%–31.5 95% C.I.; and 7.9% of AA

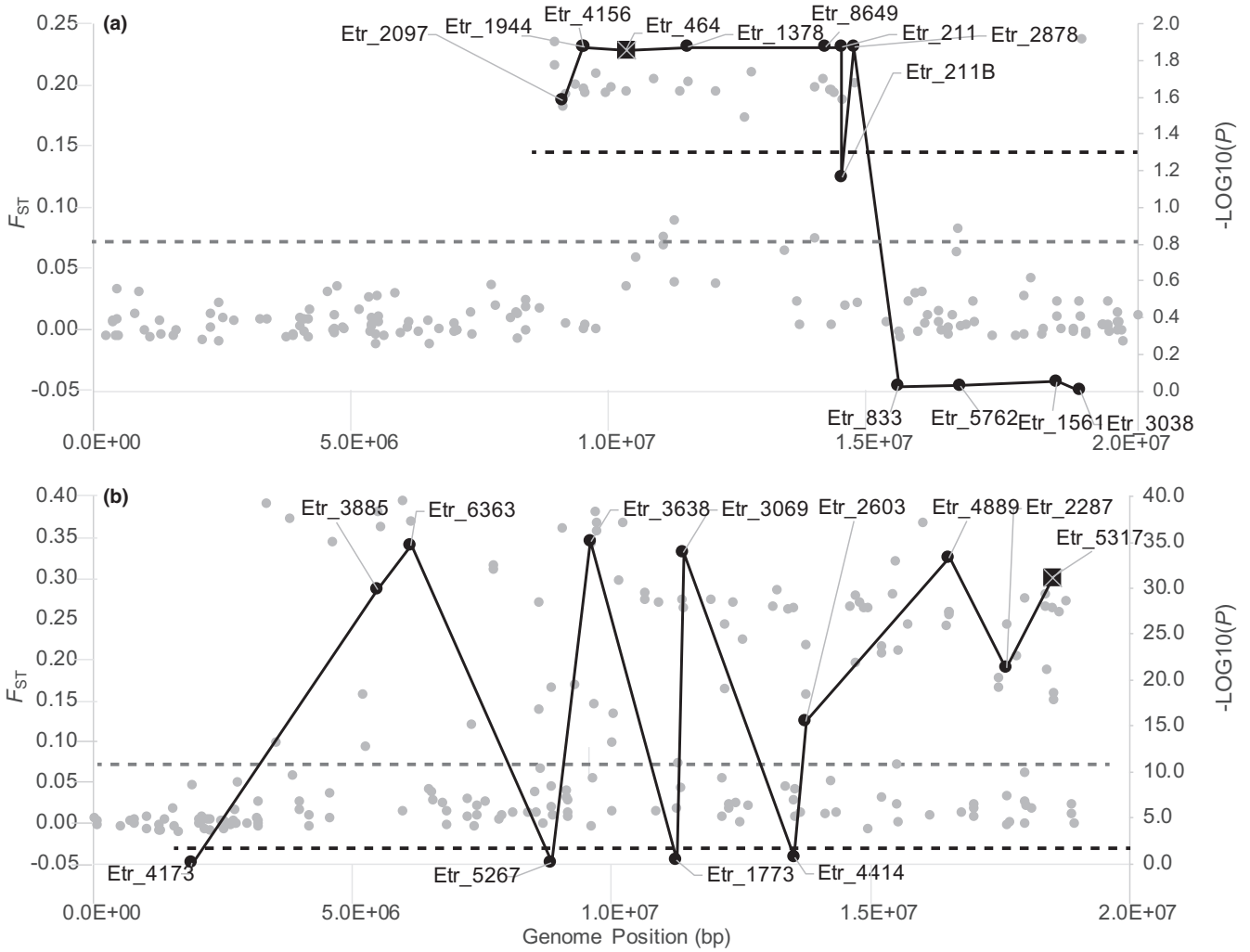

FIGURE 2 Manhattan plots of SNP positions on the male Pacific lamprey gametic genome of (a) chromosome 01 and (b) chromosome 02. F_{ST} values among range-wide collections generated in LOSITAN are indicated on the primary y-axis (grey). The critical F_{ST} value of 0.07 is indicated by the grey dashed line (SNPs above this line were considered *outliers* $p > .995$). The −LOG10(P) values from association testing are indicated on the secondary y-axis (black), and show values from (a) testing gonad size in females for the WFA♀ data set and (b) testing adult body size for the T_BON data set. The solid black line was used to distinguish the group of SNPs with labeled names used for association testing from the larger number of SNPs used for range-wide divergence mapping (i.e., grey dots). Further, the representative candidate SNPs used in subsequent analyses (Figures 3–5) are indicated with a square. The critical value of 1.3 −LOG10(P) indicates the adjusted *p*-values using the Benjamini and Hochberg (1995) false discovery rate for alpha = 0.05

fish comprised the late run, 4.5%–11.8 95% C.I.) and 2015 (i.e., 26.8% AA fish comprised the early run, 22.2%–31.5 95% C.I.; and 7.9% of AA fish comprised the late run, 4.5%–11.8 95% C.I.).

For the genotypes at Etr_5317 (chromosome 02) associated with adult body-size we did not observe consistent intra-annual trends across years (Figure S6). Genotype proportions at Etr_5317 were similar for both halves of the runs in 2014 and 2015; further, all of these relative proportions did not show significant changes between halves of the run in 2015 based on overlap of the 95% C.I. (Figure S7). However, when we paired phenotypic body-size with the genotypes at Etr_5317, we observed a relatively large and consistent trend of a decrease in proportions of AA and AC genotypes that exhibited phenotypic small body-sizes ("ShortAA/AC") across the runs in 2014 and 2015 (Figure 5c,d). We estimated that the ShortAA/AC fish decreased by >2.5× between the first and second halves of the run (Figure 5a, b). This large decrease of ShortAA/AC fish was significant based on nonoverlapping 95% CI of the estimated proportions for this category of individuals (Figure S7) in the first and second halves of the run in 2014 (i.e., 46.8% ShortAA/AC fish comprised the early run, 41.1%–52.8 95% CI; and 16.2% of ShortAA/AC fish comprised the late run, 10.9%–21.7 95% CI) and 2015 (i.e., 56.7% ShortAA/AC fish comprised the early run, 49.1%–64.2 95% CI; and 20.8% of ShortAA/AC fish comprised the late run, 13.7%–28.5 95% CI). On the basis of multiyear observations of migration patterns, we infer that this category of AA and AC Etr_5317 genotype with phenotypic small body size is a proxy for fish that exhibit advanced maturity. Association testing conducted on pre-mature adults with genotypes AC and AA at Etr_5317 (e.g., at Bonneville Dam) demonstrated strong association with intermediate to large adult body-size, respectively. However, Willamette Falls data sets contained

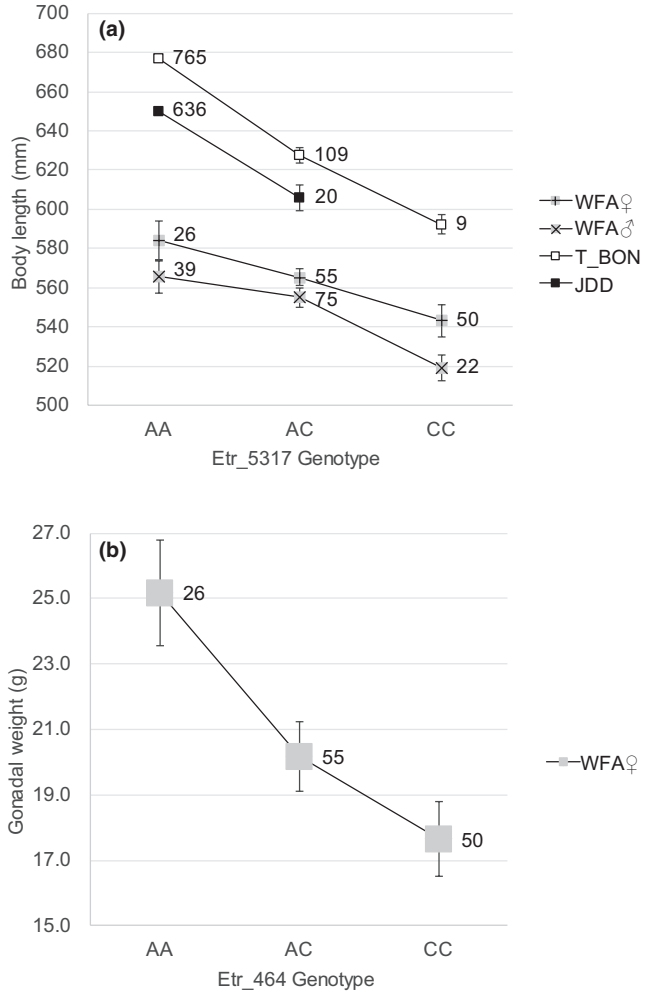

FIGURE 3 Average total body length (a) and average gonad weight (b) for genotypes at the representative candidate SNPs on chromosome 02 and chromosome 01, respectively. For each point, the sample size and standard error bars are shown. Data set abbreviations as in Table 1

mixtures of fish in varying states of maturity. Since these fish tend to shrink in body size as they reach maturity, these AC and AA genotypes can also be found at Willamette Falls in adults with relatively small-body size (i.e., total length in the lower 50% of the length distribution). Our proxy for fish with advanced maturity (AA and AC Etr_5317 genotype with phenotypic small body-size) made it possible to demonstrate that these fish arrive shortly before spawning as compared to premature fish that spend several months in freshwater prior to spawning (Figure 5c, d).

4 | DISCUSSION

4.1 | Phenotypic trait associations explain existence of genomic islands of divergence on chromosomes

Trait associations with adult body-size metrics and female gonad size (a proxy for maturity) appear to explain the presence of high levels of genomic divergence on two of the four major adaptive chromosome regions in Pacific lamprey (i.e., chromosomes 01 and 02). Genotype-phenotype association testing across multiple data sets from the Columbia and Klamath River basins consistently had strong association of body size with chromosome 02. Using samples from the Klamath River in California, Parker et al. (2019) associated the maturity trait ("ocean" and "river-maturing" ecotypes) with markers we have now mapped on chromosome 01. In this study, the association has been extended geographically to include Willamette Falls, in Oregon City, OR. We have evidence, particularly on chromosome 01, that the divergent alleles on these chromosomes are tightly linked across extensive genomic regions because they are captured within inversions that are polymorphic in the species. This concentrated genomic architecture could be key to the landscape genetics and apparent local adaptation for this highly dispersive and near panmictic species. The genotype-by-phenotype associations of candidate markers were exploited for their predictive ability to extrapolate putative distributions of the phenotypes across the species' range. These predicted phenotypic distributions provide insight into how these traits may be heterogeneously distributed in space: ocean-mature and small-bodied lamprey appear concentrated in coastal streams, whereas stream-mature and large-bodied lamprey are concentrated in interior streams. These phenotypes also appear temporally heterogeneous based on their arrival at Willamette Falls where stream-mature fish return to freshwater long before spawning in contrast to ocean-mature fish that arrive shortly before spawning. This predicted heterogeneity of the spatiotemporal distribution of maturity and body size traits provides a basis for understanding what combinations of traits may be optimally suited for particular freshwater habitats across the species' range.

4.2 | Genetic architecture of Pacific lamprey body forms

The genetic architecture underlying these traits related to body size and maturity is highly concentrated. However, this result is somewhat expected given the high degree of gene flow exhibited in Pacific lamprey (Spice, Goodman, Reid, & Docker, 2012). When natural selection is strongly acting on a particular trait in the face of high gene-flow, the genes involved in the trait tend to become highly concentrated and physically linked within the genome. Concentrated genetic architecture (i.e., few quantitative trait loci, QTL, of large effect) has been predicted to evolve under a set of conditions that include, among other factors, higher rates of gene flow between diverging populations compared to conditions leading to more diffuse genetic architecture (i.e., many QTL of small effect, Yeaman & Whitlock, 2011). We have previously found that the adaptive genetic markers were statistically linked (i.e. exhibited linkage disequilibrium within populations) and that allowed categorization of these markers into four groups of linked loci (groups A, B, C, and D; Hess et al., 2013). Now we can confirm that groups previously characterized as A, B, C, and D loci (Hess et al., 2015; Parker et al., 2019)

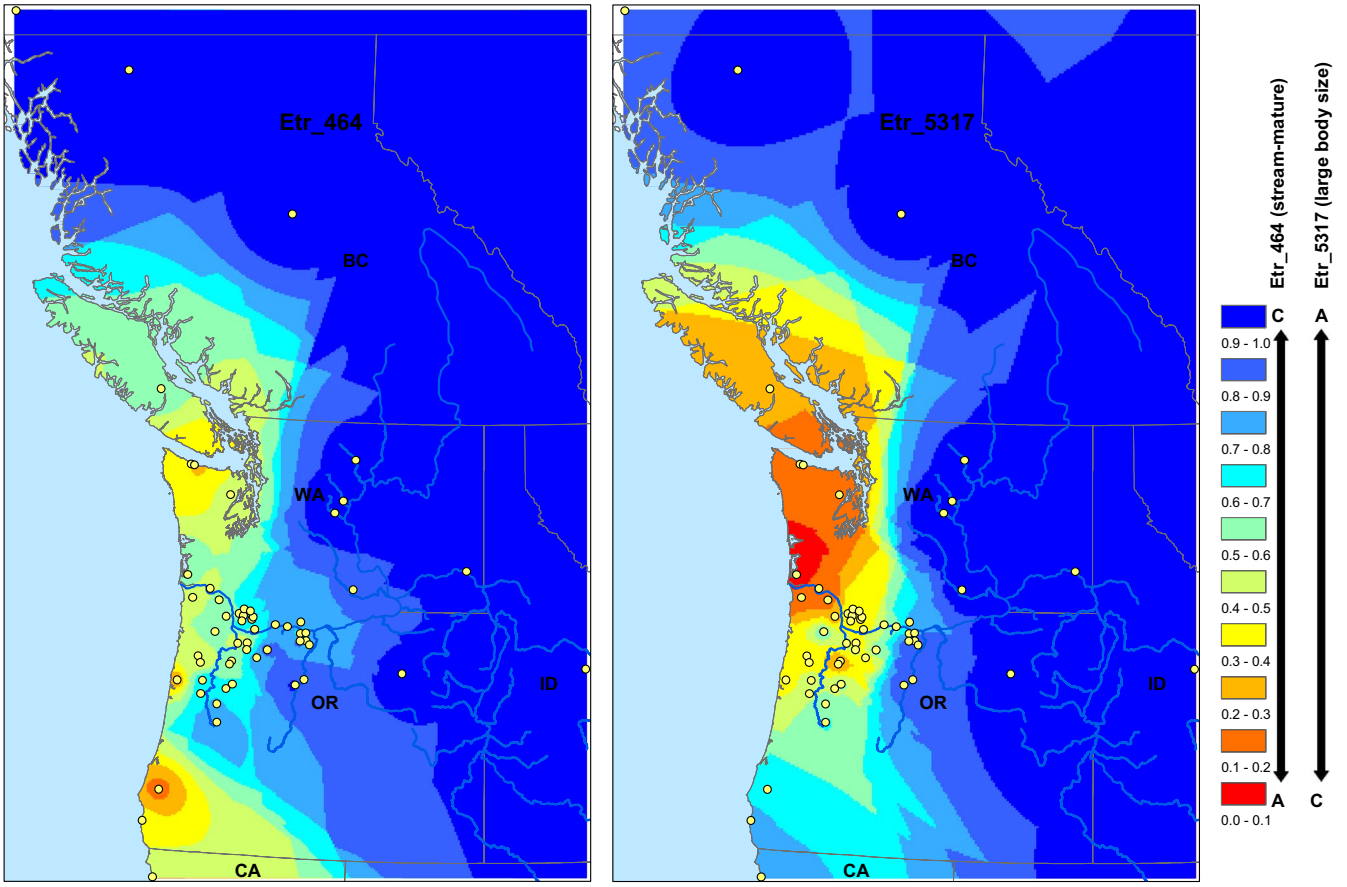

FIGURE 4 Interpolated candidate gene distribution map for prediction of maturity (Etr_464, left) and body size (Etr_5317, right) phenotypic trait distributions. Dots indicate the locations of the 57 collections used for this spatial data set (Table S1). Allele proportions are colour coded from low to high proportions of the C allele for Etr_464 and the A allele for Etr_5317 which are associated with premature ("stream-mature") female gonad and large adult body size, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

localize to chromosomes 02, 04, 22, and 01, respectively, out of a total of 83 chromosomes characterized based on a linkage map (Smith et al., 2018). Further, we observed that several of these chromosomes have one or more inversion alleles that distinguish the species from a noninverted state (based on sea lamprey) and appear polymorphic within the species (based on alignments between male and female Pacific lamprey genome assemblies). These inversions appear to coincide with the adaptive SNPs identified as F_{ST} outliers, particularly for the cases of chromosome 01 and 02, which suggests that polymorphic inversions may play an important role in the adaptation of Pacific lamprey to local environments. Since recombination is highly reduced, the fitness conferred by these inverted haplotypes can help maintain them as a polymorphism in a population through both forces of balancing and divergent selection (Faria et al., 2019; Pearse et al., 2019; Wellenreuther & Bernatchez, 2018).

4.3 | Targets of selection

Although the phenotypes that we measured and observed in association with candidate SNPs cannot be concluded to be the causal variants that are the actual targets of selection, we can use the genotype-by-phenotype associations to guide future research aimed to identify these targets. Further, we can also narrow down some traits and life stages that do not appear relevant to any observed adaptive variation. For example, the adaptive genetic variation had no predictive ability for the short-term swimming ability of migration-phase adult lamprey at Bonneville Dam or the growth differences among young-of-year larvae. These two cases of failure to reject a null hypothesis help narrow down the search for a mechanism that manifests in large body size adults that tend to travel further upstream to spawn (a trait highly associated with genes on chromosome 2). These genes apparently do not confer adult swimming endurances, at least for the short timeframe that could be tested in swim trials at Bonneville Dam. Further, these adult body size differences do not translate to faster growth in young of year larvae. However, these adult body size differences could be influenced by differential growth at older life stages; prey selection, length of time in the ocean, or ocean distribution probably affect growth (Clemens et al., 2019). It will require further investigation of multiple life stages in both freshwater and the ocean to understand Pacific lamprey life history strategies. For the maturity trait, gene ontology could suggest other traits to examine as potential targets of selection including circadian rhythm.

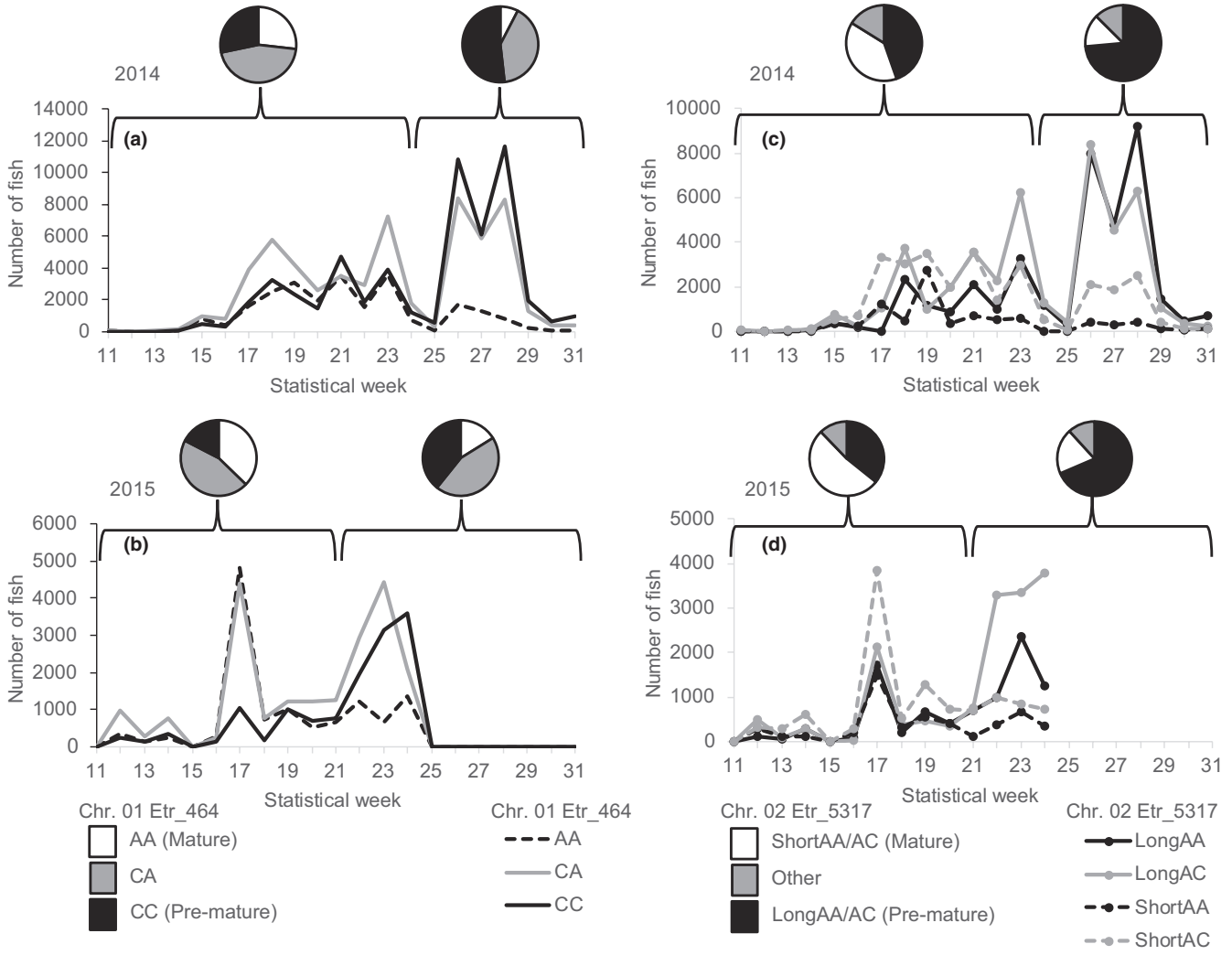

FIGURE 5 Temporal distributions of candidate SNPs for characterizing phenotypes of the adult Pacific lamprey migration at Willamette Falls. The timing of AA versus CC genotypes at Etr_464 demonstrates relative migration of the "mature" and "premature" ecotype for both (a) 2014 and (b) 2015 runs. The line charts (a–b) indicate the relative abundance of genotypes at Etr_464 across statistical weeks and the pies indicate the relative proportions of those genotypes for first and second halves of the run. The migration timing of the AA and AC genotypes at Etr_5317 of fish that were phenotypically small-bodied "ShortAA/AC" versus phenotypically large-bodied "LongAA/AC" was relatively early and late, respectively, for both (c) 2014 and (d) 2015 runs. Late stages of maturation cause body size shrinkage (Clemens et al., 2009) and so even genotypes associated with large and intermediate body size (i.e., genotypes AA and AC at Etr_5317) can exhibit relatively short body lengths in the lower 50% distribution of length ("ShortAA/AC") when they mature. The line charts (c–d) indicate relative abundance of the "Short" and "Long" phenotypes of the Etr_5317 genotypes and the pies indicate the relative proportions of those phenotype-genotype groups for the first and second halves of the run

4.4 | Trait comparisons with other anadromous fishes

Despite strong differences in philopatry and population genetic structure, there are similarities between the Pacific lamprey ecotypes and those of steelhead trout (anadromous *Oncorhynchus mykiss*) and Chinook salmon (*O. tshawytscha*), as described by Parker et al. (2019). In this study, we observed even greater similarities with steelhead of these ecotypes in the Willamette River than were apparent in the Klamath River Pacific lamprey. Notably, like steelhead (Hess, Zendt, Matala, & Narum, 2016), the Pacific lamprey ocean- and river-maturing ecotypes exhibit seasonal separation where premature fish return to freshwater long before spawning in contrast to mature fish that arrive shortly before spawning. Also similar to steelhead (Micheletti, Hess, Zendt, & Narum, 2018), the Pacific lamprey ocean-maturing form is only distributed in coastal regions and the river-maturing ecotype is distributed further inland. However, it is unknown whether inland migrating Pacific lamprey exhibit both early and late arrival to spawning grounds as observed for inland migrating steelhead (Micheletti et al., 2018) and Chinook salmon (Narum, Di Genova, Micheletti, & Maass, 2018). Finally, although we found no evidence that the homologous genes were conserved with salmonids, Pacific lamprey ecotypes were associated with a single locus of major effect as shown in steelhead (Hess

et al., 2016; Micheletti et al., 2018) and Chinook salmon (Prince et al. 2017; Narum et al., 2018).

The adult body size trait that was associated with chromosome 02 genes in Pacific lamprey may share similarities with the age-at-maturity trait described in Pacific salmon (McKinney et al., 2019), steelhead (Copeland, Ackerman, Wright, & Byrne, 2017), and Atlantic salmon (*Salmo salar*, Barson et al., 2015). In salmonids, the number of consecutive years spent in the ocean before returning to freshwater as a mature adult is highly correlated to body size (e.g., Chinook salmon, Lewis, Grant, Brenner, & Hamazaki, 2015). It has also been shown that larger, older Chinook and Sockeye salmon (*O. nerka*) tend to arrive earlier at Bonneville Dam compared to the smaller 1-ocean-age adults (Anderson & Beer, 2009). Similarly, larger Pacific lamprey arrive earlier at Bonneville Dam than the smaller bodied forms (Keefer et al., 2009, 2013), which may be related to life history decisions in seasonal environments. The primary growth of Pacific lamprey occurs during the ocean phase of the lamprey's parasitic life cycle and so bigger lamprey may also be older in ocean age. There is not yet an accurate way to measure the total age or ocean age of lampreys since they lack bony structures, but this hypothesis could be tested once an aging method is developed (e.g., statolith microstructure). Collectively, the convergence of traits in salmonids and lamprey suggest strong tradeoffs between allocation of resources in capital breeding fishes, whereby long distance migration constrains maturation schedules and in at least some cases (e.g., lamprey) body size.

4.5 | Hypothesis testing framework

Pacific lamprey genetic traits and their associated phenotypes appear to be inherited independently and may occur in combinations that manifest as different life history strategies to fit unique ecological conditions throughout the species' range. For example, in the Willamette River basin there is relatively high diversity of traits and nearly equal portions of genetic variants associated with alternate forms of small and large-bodied adults and ocean- versus river-maturing ecotypes. The following multiple strategies appear to be represented: (a) stream-maturing small- and large-bodied fish that have overwintered below the Falls; (b) ocean-maturing small- and large-bodied fish that arrive shortly before spawning above the Falls; and (c) stream-maturing small- and large-bodied fish that arrive after June and ascend the Falls. The fact that there are four separate chromosomes with important adaptive genes (some with undetermined trait associations) provides for the possibility that various combinations of adaptive variants at these four chromosomes could underpin a multitude of life history strategies. Patterns in the occurrence of the two phenotypic traits we emphasized in this study suggest that Pacific lamprey life history traits may exhibit differential fitness across the range. For example, extrapolation predicts a predominance of large-bodied, stream-maturing forms in northern B.C. and the interior Columbia River, small-bodied ocean- and stream-maturing forms in Puget Sound, intermediate-bodied ocean- and stream-maturing forms in the lower Columbia, and large-bodied ocean- and stream-maturing forms in the southern coastal range. This provides a hypothesis testing framework to examine the incidence likelihoods of life history traits across the range, and understand factors driving optimization of these traits.

In conclusion, we identified four chromosomes that appear to contain adaptive regions and represent genomic islands of divergence across the species' range. Although there may be other genomic islands of divergence that we have not yet identified, it is unlikely that other regions will harbor similar magnitudes of divergence at broad geographic scales. Characterization of adaptive genomic divergence in general terms is relatively easy compared to the efforts involved in characterizing phenotypic variation. It is the latter effort that is sorely needed for future insights into the evolutionarily important traits for this species. Of the phenotypes we anticipate would be difficult to characterize but especially illuminating for testing associations with the four genomic islands would be spawn timing, marine duration, marine distribution, and prey preference. Some of these traits may be the true target phenotypes of selection, for which the traits we found to be most important in this study (i.e., body size and female gonad size) probably serve as proxies.

Nonetheless, we have provided information on how at least two traits related to body size and maturity may be optimally selected for suites of ecological conditions across the species' range. This type of trait information addresses some of the critical uncertainties that have needed further clarification to increase the success of restoration activities that are currently being implemented in the Columbia River basin and other regions. Despite the fact that this species seems to maintain high rates of gene flow across broad geographic regions and should be resilient to loss of genetic diversity, these results underscore the importance in maintaining life history diversity and the genes that underlie it to allow this species to occupy and persist in its range effectively.

ACKNOWLEDGEMENTS

This work was funded by Bonneville Power Administration, and we appreciate the contributed sampling efforts by Jeffery Jolley, Luke Schultz, Andrew Wildbill, Matt Fox, Ralph Lampman, and the Army Corp of Engineers, as well as the laboratory and coordination efforts by Travis Jacobson, Stephanie Harmon, Brian McIlraith, and Jeff Stephenson. The edits suggested by three anonymous reviewers were extremely helpful for improving the original draft of this manuscript.

AUTHOR CONTRIBUTIONS

J.E.H. designed the study, analysed the data, wrote the manuscript; J.J.S., and N.T. analysed and compiled genome assemblies; C.B., C.C.C., M.L.K., M.L.M., L.L.P., G.S. directly sampled, contributed analysis, and coordinated data collections; D.G. performed spatial interpolations; S.L.W. performed daily abundance analysis; S.R.N. advised on the analysis and study design; and all authors contributed to improving drafts of the manuscript.

DATA ACCESSIBILITY

The population genomic divergence data set of 7,716 quality-filtered SNP loci, the six data sets used for association mapping with 302 unique SNP loci, the two temporal data sets of Willamette Falls (2014–2015) adults with 302 unique SNP loci, and the spatial data set of larvae/juveniles range-wide with 85 SNP loci are available as datafiles (Dryad repository: https://doi.org/10.5061/dryad.hx3ffbgc2).

ORCID

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

REFERENCES

Anderson, J. J., & Beer, W. N. (2009). Oceanic, riverine, and genetic influences on spring Chinook salmon migration timing. *Ecological Applications*, *19*(8), 1989–2003. https://doi.org/10.1890/08-0477.1

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a *Fst*-outlier method. *BMC Bioinformatics*, *9*(1), 323.

Asaduzzaman, M., Wahab, M. A., Rahman, M. J., Nahiduzzzaman, M., Dickson, M. W., Igarashi, Y., … Wong, L. L. (2019). Fine-scale population structure and ecotypes of anadromous Hilsa shad (Tenualosa ilisha) across complex aquatic ecosystems revealed by NextRAD genotyping. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-52465-2

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., … Kent, M. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*(7582), 405–408.

Beamish, R. J. (1980). Adult biology of the river lamprey (*Lampetra ayresi*) and the Pacific lamprey (*Lampetra tridentata*) 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*(1), 289–300.

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*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

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*(4), 855–867. https://doi.org/10.1111/1755-0998.12357

Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., … Brooks, L. D. (2007). Replicating genotype–phenotype associations. *Nature*, *447*(7145), 655.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., … Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. https://doi.org/10.1186/1471-2105-14-128

Clemens, B. J., van de Wetering, S. J., Kaufman, J., Holt, R. A., & Schreck, C. B. (2009). Do summer temperatures trigger spring maturation in adult Pacific lamprey, Entosphenus tridentatus? *Ecology of Freshwater Fish*, *18*, 418–426.

Clemens, B. J., Weitkamp, L., Siwicke, K., Wade, J., Harris, J., Hess, J., … Orlov, A. M. (2019). Marine biology of the pacific lamprey *Entosphenus tridentatus*. *Reviews in Fish Biology and Fisheries*, *29*, 767–788. https://doi.org/10.1007/s11160-019-09578-8

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*(7), 19–25. https://doi.org/10.1577/1548-8446(2002)027<0019:TEACIO>2.0.CO;2

Copeland, T., Ackerman, M. W., Wright, K. K., & Byrne, A. (2017). Life history diversity of Snake River steelhead populations between and within management categories. *North American Journal of Fisheries Management*, *37*(2), 395–404. https://doi.org/10.1080/02755947.2016.1264506

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology and Evolution*, *34*, 239–248. https://doi.org/10.1016/j.tree.2018.12.005

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*(11), 2898–2916. https://doi.org/10.1111/mec.12150

Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., … Wildbill, A. J. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, *15*(1), 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*(10), 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.

Jones, O., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, *10*, 551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x

Keefer, M. L., Caudill, C. C., Clabough, T. S., Jepson, M. A., Johnson, E. L., Peery, C. A., … Moser, M. L. (2013). Fishway passage bottleneck identification and prioritization: A case study of Pacific lamprey at Bonneville Dam. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*, 1551–1565. https://doi.org/10.1139/cjfas-2013-0164

Keefer, M. L., Moser, M. L., Boggs, C. T., Daigle, W. R., & Peery, C. A. (2009). Effects of body size and river environment on the upstream migration of adult Pacific lampreys. *North American Journal of Fisheries Management*, *29*(5), 1214–1224. https://doi.org/10.1577/M08-239.1

Kirk, M. A., Caudill, C. C., Tonina, D., & Syms, J. C. (2016). Effects of water velocity, turbulence and obstacle length on the swimming capabilities of adult Pacific lamprey. *Fisheries Management and Ecology*, *23*(5), 356–366. https://doi.org/10.1111/fme.12179

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., … McDermott, M. G. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/nar/gkw377

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357. https://doi.org/10.1038/nmeth.1923

Lewis, B., Grant, W. S., Brenner, R. E., & Hamazaki, T. (2015). Changes in size and age of Chinook Salmon Oncorhynchus tshawytscha returning to Alaska. *PLoS One*, *10*(6), e0130184. https://doi.org/10.1371/journal.pone.0130184

McKinney, G. J., Seeb, J. E., Pascal, C. E., Schindler, D. E., Gilk-Baumer, S. E., & Seeb, L. W. (2019). Y-chromosome haplotypes are associated with variation in size and age at maturity in male Chinook salmon. *Evolutionary Applications*. https://doi.org/10.1111/eva.13084

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.

Miller, A. D., Hoffmann, A. A., Tan, M. H., Young, M., Ahrens, C., Cocomazzo, M., … Sherman, C. D. (2019). Local and regional scale habitat heterogeneity contribute to genetic adaptation in a commercially important marine mollusc (Haliotis rubra) from southeastern Australia. *Molecular Ecology*, *28*(12), 3053–3072.

Narum, S. R. (2006). Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics*, *7*(5), 783–787. https://doi.org/10.1007/s10592-005-9056-y

Narum, S. R., Di Genova, A., 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*(1883), 20180935.

Orlov, A. M., Beamish, R. J., Vinnikov, A. V., & Pelenev, D. M. I. T. R. Y. (2009). Feeding and prey of Pacific Lamprey in coastal waters of the western North Pacific. In *American Fisheries Society Symposium* (Vol. *69*, pp. 875–877).

Parker, K. A., Hess, J. E., Narum, S. R., & Kinziger, A. P. (2019). 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. *Molecular Ecology*, *28*, 3171–3185.

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., … Moen, T. (2019). Sexdependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, *3*(12), 1731–1742.

Phair, N. L., Toonen, R. J., Knapp, I., & von der Heyden, S. (2019). Shared genomic outliers across two divergent population clusters of a highly threatened seagrass. *PeerJ*, *7*, e6806. https://doi.org/10.7717/peerj.6806

Powell, M., & MacGregor, J. (2011). A geographic test of species selection using planktonic foraminifera during the Cretaceous/Paleogene mass extinction. *Paleobiology*, *37*, 426–437. https://doi.org/10.2307/23014731

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*(8), e1603198.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2011. Retrieved from https://www.R-project.org

SAS Institute, Inc. (2000). *SAS/STAT User's Guide, version 8*. Cary, NCSAS Institute, Inc.

Smith, J. J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M. C., Parker, H. J., … Kaessmann, H. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nature Genetics*, *50*(2), 270. 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*(12), 2916–2930. https://doi.org/10.1111/j.1365-294X.2012.05585.x

Steinhorst, K., Copeland, T., Ackerman, M. W., Schrader, W. C., & Anderson, E. C. (2017). Abundance estimates and confidence intervals for the run composition of returning salmonids. *Fishery Bulletin*, *115*(1), 1–12. https://doi.org/10.7755/FB.115.1.1

Wang, J. (2004). Sibship reconstruction from genetic data with typing errors. *Genetics*, *166*(4), 1963–1979. https://doi.org/10.1534/genetics.166.4.1963

Ward, D. L., Clemens, B. J., Clugston, D., Jackson, A. D., Moser, M. L., Peery, C., & Statler, D. P. (2012). Translocating adult Pacific lamprey within the Columbia River Basin: State of the science. *Fisheries*, *37*(8), 351–361. https://doi.org/10.1080/03632415.2012.704818

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, *33*(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002

Whitlock, S. L., Deweber, J. T., & Peterson, J. T. (2019). *Assessment of Pacific Lamprey monitoring techniques at Willamette Falls, OR*. Draft Report to the Confederated Tribes of Warm Springs Reservation of Oregon.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution: International Journal of Organic Evolution*, *65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., … Kresovich, S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hess JE, Smith JJ, Timoshevskaya N, et al. Genomic islands of divergence infer a phenotypic landscape in Pacific lamprey. *Mol Ecol*. 2020;29:3841–3856. https://doi.org/10.1111/mec.15605