

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Species diversity, evolutionary relationships, and genomic diversity of lampreys in California (Entosphenus spp. and Occidentis spp.)

Permalink

<https://escholarship.org/uc/item/0826d533>

ISBN

9798193265986

Author

Auringer, Grace

Publication Date

2026-09-08

Supplemental Material

<https://escholarship.org/uc/item/0826d533#supplemental>

Peer reviewed|Thesis/dissertation

Species Diversity, Evolutionary Relationships, and Genomic Diversity of Lampreys in
California (*Entosphenus* spp. and *Occidentis* spp.)

By

GRACE AURINGER
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Andrea Schreier, Chair

Graham Coop

Matthew A. Campbell

Committee in Charge

2026

Acknowledgements

I want to thank everyone who made this study possible, especially Dr. Amanda J. Finger, Dr. Andrea Schreier, Pascale Goertler, and Dr. Mac Campbell. Thank you so much to every person who volunteered to handle, measure, photograph and collect tissue samples from lamprey – I do not have the names of every crew member, but I do want to individually thank Ed Rible and Michelle Workman, East Bay Municipal Utility District; Brian Sak, San Francisco Public Utilities Commission; Jason Guignard, Tyler Pilger, Tara Lamb, FISHBIO; Dave Contreras, Jason Julienne, Michael Dege, Maggie Massie, Morgan Knechtle, Janine Phillips, California Department of Fish and Wildlife; Jonathan Koehler, Napa County Resource Conservation District; Chris Cook, California Department of Water Resources; Jacob Dickey and Melanie Davis, Oregon State University; Oliver ‘Towns’ Burgess, Bureau of Reclamation; Kassie Hickey, Logan Day, Cory Starr, Pacific States Marine Fisheries Commission; Bill Poytress, U.S. Fish and Wildlife Service; Jeremy Notch, National Oceanic and Atmospheric Administration (NOAA). Additionally, I want to thank those who contributed funds to this project, especially California Sea Grant/NOAA, and the Department of Water Resources. Last, but certainly not least, thank you to my family and friends for all your encouragement and support.

Abstract

Lampreys on the West Coast of North America are an understudied group of native fishes. Many West Coast lamprey populations are in decline, but the lack of historical data and poor fundamental scientific understanding of these fishes obstruct conservation efforts. Taxonomic uncertainty is prevalent across the entire lamprey lineage, in part due to patterns of life history variation (i.e. migration, parasitism) and associated morphological diversity (i.e. body size, oral disc dentition). In some cases, different life history variants are accurately classified as distinct species, but in other cases, these life history variants co-occur within the same species. As genetic tools were developed, some taxonomic uncertainty has been resolved; however, previous research has been limited to few genetic markers and included few samples from California. This study utilizes mitochondrial phylogenetics and molecular species delimitation methods in tandem with phylogenomics and population genomic analyses to investigate species diversity, evolutionary relationships, and genomic diversity of lampreys in California (*Entosphenus* spp. and *Occidentis* spp.). Study results show relatively shallow divergences between *Entosphenus* species, deep divergences between *Occidentis* species, and high species richness in *Occidentis* in California, including two novel species found in the Napa River and Alameda Creek. In *Entosphenus*, results largely support current taxonomy, except the unnamed Goose Lake lamprey, which is not genetically distinct from *E. lethophagus*. Genomic differentiation across *E. tridentatus* individuals is isolated to a putative chromosomal inversion region, which is associated with variation in reproductive maturity timing. Study results advance our understanding of lampreys and inform conservation management.

Chapter 1: Lampreys in California (*Lampetra* spp. and *Entosphenus* spp.): Mitochondrial phylogenetic analysis reveals previously unrecognized lamprey diversity



Received: 13 April 2023 | Revised: 11 September 2023 | Accepted: 18 September 2023
DOI: 10.1002/nafm.10959

SPECIAL SECTION

Native Lampreys: Research and Conservation of Ancient Fishes

Lampreys in California (*Lampetra* spp. and *Entosphenus* spp.): Mitochondrial phylogenetic analysis reveals previously unrecognized lamprey diversity

Grace Auringer¹ | Matthew A. Campbell^{1,2} | Pascale A. L. Goertler³ |
Amanda J. Finger¹

¹Department of Animal Science,
University of California, Davis, Davis,
California, USA

²University of Alaska Museum of the
North, Fishes and Marine Invertebrates,
Fairbanks, Alaska, USA

³Delta Stewardship Council, Delta
Science Program, Sacramento,
California, USA

Correspondence

Grace Auringer
Email: gmauringer@ucdavis.edu

Abstract

Objective: Protecting biodiversity is important for preserving ecosystem functions and services, and understanding the diversity present in a system is necessary for effective conservation. Lampreys display extensive diversity in morphology and life history strategy. The extent of this diversity and the underlying genetic patterns have not yet been fully characterized. Uncertainty about species boundaries and operational taxonomic units hinders lamprey management and conservation. Limited data have been collected on California lampreys (*Lampetra* spp. and *Entosphenus* spp.), but evidence suggests widespread population decline across the west coast of North America. Study objectives were to identify which nominal species are present at study sites across northern and central California, determine whether current taxonomic classifications accurately and sufficiently describe lamprey diversity in California, and characterize the biogeographic distribution of genetically distinct lamprey lineages across the study area.

Methods: To achieve these objectives, this study utilized DNA barcoding, phylogenetic analysis, and species delimitation analysis. Lamprey individuals ($N=87$) from 19 sites in the Sacramento–San Joaquin River basin, San Francisco Bay, and Klamath River basin were sequenced for the mitochondrial cytochrome *b* (*cyt b*) gene, and the data were combined with publicly available lamprey *cyt b* sequences for analysis.

Result: Results showed relatively deep phylogenetic divergence between *Lampetra* and *Entosphenus*. Distinct and genetically divergent lineages were observed within *Lampetra*, while distinct but genetically similar lineages were observed within *Entosphenus*. This study revealed novel *Lampetra* lineages in the Napa River and Alameda Creek, and results suggest that the biogeographic distribution of lamprey diversity may follow patterns observed in other native fishes. Species delimitation results indicated that at least seven separate candidate species of *Lampetra* can be found in California, suggesting that California holds more species-level diversity than the expected three nominal *Lampetra* species known to occur in the state.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *North American Journal of Fisheries Management* published by Wiley Periodicals LLC on behalf of American Fisheries Society.

North American Journal of Fisheries Management. 2023;43:1511–1530.

wileyonlinelibrary.com/journal/nafm | 1511

Conclusion: These results highlight the underestimated diversity of lampreys in California and the need for further assessment of taxonomic classifications and operational taxonomic unit designations of California lampreys.

KEYWORDS

California, conservation, *Entosphenus*, *Lampetra*, lamprey, molecular species delimitation, phylogenetics

INTRODUCTION

Lampreys are jawless, boneless fish characterized by eel-like bodies with seven gill pores on each lateral side, oral discs with keratinized plates, and the absence of paired fins. The lamprey life cycle consists of a prolonged larval phase that is variable in duration (~3–9 years), followed by a period of metamorphosis, after which the life history types diverge (Renaud 2011; Lucas et al. 2021). Despite high similarity across all species during the larval phase, substantial diversity is observed in adult lampreys, especially in migration type (anadromous and freshwater resident), feeding strategy (parasitic and nonparasitic), oral disc dentition patterns, body size, and pigmentation (Renaud 2011; Potter et al. 2015). Although lampreys are part of an ancient lineage, contemporary lamprey diversity (41 species; Page et al. 2013) has a relatively recent origin; the diversification of extant lamprey species has occurred primarily in the last 20 million years of the over 450-million-year evolutionary history (Brownstein and Near 2023). Twenty-five percent of extant lamprey species are considered to be at risk of extinction, and this is likely an underestimate due to the lack of data regarding distributions, population trends, and overall species diversity and distribution (Endangered and Threatened Wildlife and Plants 2004; Maitland et al. 2015; Docker and Hume 2019; Lucas et al. 2021).

Across the globe, native lamprey species provide substantial benefits to freshwater systems and the surrounding communities. During the filter-feeding larval phase, all lampreys improve water quality, maintain streambed conditions, and cycle nutrients, thus boosting primary productivity (Shirakawa et al. 2013; Boeker and Geist 2016). Juvenile and adult lampreys are valuable prey items for many birds, fishes, and aquatic mammals; additionally, anadromous adults deposit abundant marine-derived nutrients far upstream upon completion of their spawning migration, contributing valuable resources to stream food webs (Close et al. 2002; Dunkle et al. 2020). Anadromous Pacific Lampreys *Entosphenus tridentatus* are a dietary staple and cultural resource for many Indigenous peoples along the west coast (Close et al. 2002). These fish have been highly valued, monitored, and managed by Indigenous nations since time immemorial, but

Impact statement

Lampreys are ecologically and culturally significant. Populations are thought to be declining across the west coast of North America, but a lack of research impedes monitoring and conservation of lampreys in California. This study investigated lampreys in California using DNA sequences and found more diversity than expected. By determining how many genetically different types of lampreys are present and where these various lampreys are found across the study area, this research paves the way for conservation and protection of lampreys and the numerous benefits they contribute.

population declines over the past 50–75 years jeopardize Indigenous culture and food sovereignty (Sheoships 2014; Wicks-Arshack et al. 2018). Evidence suggests that the decline in lamprey populations across North America is due to loss of larval rearing and spawning habitat, pollution, damming and channelization of rivers, and alteration of natural flow regimes (Renaud 1997; Maitland et al. 2015; Lucas et al. 2021). However, a lack of historical data and the geographically fragmented nature of contemporary data prevent a robust rangewide assessment of population declines (Lucas et al. 2021; Parker 2022). Data collection is further complicated by challenges in species identification due to the lack of distinguishing features between larval lampreys of different species (Docker 2009; Parker 2022). These barriers to lamprey conservation are compounded by a lack of basic biological knowledge and taxonomic uncertainty (Moser et al. 2007; Maitland et al. 2015; Docker et al. 2016; Lucas et al. 2021).

Conservation actions based on inaccurate taxonomy can have unintended negative consequences for the target taxa and could reduce biodiversity (McNeely 2002; Ely et al. 2017). In most lamprey genera, “paired” or “stem-satellite” species occur, in which one (or more) freshwater-resident, nonparasitic lamprey species is derived from a migratory, parasitic species, and the resultant species have overlapping ranges and are closely related (Docker 2009).

Current taxonomy, primarily based on morphological evidence, recognizes paired lamprey species as distinct species (Kneibelsberger et al. 2015). However, taxonomic uncertainty is prevalent throughout the entire lamprey lineage, in part due to this confusion surrounding paired species (Docker 2009).

Recent studies have demonstrated that these lamprey pairs do not always form reciprocally monophyletic groups (Schreiber and Engelhorn 1998; Espanhol et al. 2007; Docker 2009; Boguski et al. 2012) and may represent alternative life history strategies within one freely interbreeding species (Rougemont et al. 2021; Carim et al. 2023). However, other pairs are considered correctly classified as distinct species under the biological and phylogenetic species concepts (Docker 2009; Clemens and Wade 2023). Molecular phylogenetic analysis and molecular species delimitation analysis can help to resolve taxonomic relationships and inform operational taxonomic unit designation (Rannala and Yang 2020; Young et al. 2021). In molecular phylogenetic analysis, DNA sequences are analyzed to infer evolutionary relationships between related individuals, which can help to determine taxonomic classifications (Yang and Rannala 2012). The goal of molecular species delimitation is to sort sequences from related individuals into groups that reflect species boundaries, and these methods typically base delimitations on differences between DNA sequences (distance-based methods) or the topology of phylogenetic trees (tree-based methods; Zhang et al. 2013).

Knowledge gaps about west coast lampreys are prevalent, but a surge of new research on the Pacific Lamprey

has been driven by the formation of the Pacific Lamprey Conservation Initiative in 2007 (Wang and Schaller 2015). However, most of the new research focuses on a single species, and lamprey populations in California watersheds are particularly understudied (Moyle et al. 2015). Current taxonomy describes eight lamprey species (*Lampetra* spp.: $N=3$; *Entosphenus* spp.: $N=5$) present in California watersheds (Table 1), and all are listed as California species of special concern (Moyle et al. 2015). Moyle (2002) described six major zoogeographic provinces with 22 sub-provinces, and five of the eight nominal lamprey species found in California have geographically restricted ranges that correlate to these zoogeographic subprovinces. This suggests that the biogeographic distribution of lamprey diversity may follow the same patterns observed in other native fish taxa. Although the abundance and distribution of lamprey species have not been systematically assessed throughout California, evidence strongly suggests that both range contractions (Goertler et al. 2019) and population declines (Moyle et al. 2009) have occurred. No previous studies have explored genetic diversity across all nominal species of California lamprey, and most prior research has had a very limited geographic scope within California.

This study utilized mitochondrial DNA (mtDNA) sequence data to assess the species composition and genetic diversity of lamprey populations throughout the Sacramento–San Joaquin River basin, San Francisco Bay, and the Klamath River basin of California. The three main objectives of this study were to (1) identify which nominal species were present at each sampled

TABLE 1 California lamprey species names, life history types, and status in California. “Other” migratory type indicates that the species is considered migratory but migration behavior is not well understood. The taxonomic classification of the Western Brook Lamprey *Lampetra richardsoni* as a distinct species from the Western River Lamprey *L. ayresii* is currently in doubt based on phylogenetic evidence from this study and others (Boguski et al. 2012; Carim et al. 2023).

Species	Migratory			Status in California
	Anadromous	Other	Parasitic	(Moyle et al. 2015)
Pacific Lamprey <i>Entosphenus tridentatus</i> ^{a,b}	✓		✓	Moderate concern
Klamath Lamprey <i>E. similis</i> ^b		✓	✓	Moderate concern
Pit-Klamath Brook Lamprey <i>E. lethophagus</i> ^b				Moderate concern
Northern California Brook Lamprey <i>E. folletti</i> ^b				High concern
Goose Lake Lamprey <i>Entosphenus</i> sp.		✓	✓	High concern
Western River Lamprey <i>Lampetra ayresii</i> ^{a,b}		✓	✓	Moderate concern
Western Brook Lamprey <i>L. richardsoni</i> ^{a,b}				Moderate concern
Kern Brook Lamprey <i>L. hubbsi</i> ^{a,b,c}				High concern

^aIncluded in an Endangered Species Act petition to list as endangered in 2003.

^bRecognized by the American Fisheries Society (Page et al. 2013).

^cEndemic to California.

location using genetic species identification, (2) determine whether current taxonomic classifications accurately and sufficiently describe California's lamprey diversity through molecular phylogenetic and species delimitation analyses, and (3) compare the biogeographic distribution of lamprey diversity in California (based on the results of objectives 1 and 2) to the patterns observed for other native fish taxa. Based on species range descriptions for the eight nominal species known to occur in California (i.e., Table 1; Moyle 2002), we expected that all eight species would be present in our study area. Previous studies have demonstrated that lamprey systematics are complex and that taxonomic classifications can be discordant with molecular phylogenetic relationships (Docker et al. 2012; Rougemont et al. 2021), so we expected to observe similar contradictions across the *Lampetra* and *Entosphenus* genera in California. Due to the unique hydrology and rugged topography of the state (Moyle 2002; Ball et al. 2013), we predicted that the biogeographic patterns of lamprey diversity would be consistent with those of other native fish taxa. New sequencing data from lamprey individuals were integrated with publicly available data to clarify higher level relationships within and between nominal lamprey taxa. This study provides a foundation for understanding the diversity of California's lamprey populations, which is a necessary step in protecting lampreys and their ecosystem functions.

METHODS

Study system

In California's Central Valley, adult lampreys are thought to spawn throughout the Sacramento and San Joaquin rivers and their tributaries (Moyle 2002). These two rivers join to form a freshwater delta and exit toward the ocean through a series of large bays (Suisun, San Pablo, and San Francisco bays), hereafter collectively termed "the San Francisco Estuary" (Figure 1). Additionally, lampreys are widely distributed in the lower Klamath River basin and anadromous Pacific Lampreys have been documented in the upper Klamath River basin (above Iron Gate Dam; Hamilton et al. 2005). California is one of the most hydrologically altered regions in the world; extensive water infrastructure was designed to maintain a consistent water supply to the state's multi-billion-dollar (1 billion = 10^9) agricultural industry and meet the needs of municipal, residential, and other industrial water uses, but the resultant altered flows and habitat reduction are known to be detrimental to native fish populations

(Brown and Bauer 2010; Zimmerman et al. 2018). Historically, the range of anadromous species encompassed a drainage area equal to roughly two-thirds of the state of California (Whipple et al. 2012). Currently, impassible dams block 72% of upstream historical habitat (Yoshiyama et al. 2001) from access by anadromous forms of returning lamprey. California's watersheds form the southernmost boundaries of several lamprey species' ranges, and evidence of southern range contractions (Reid et al. 2016) underscores the need to evaluate lamprey populations in California.

Sample collection and DNA sequencing

Tissue samples (fin clips) were collected opportunistically from lamprey individuals ($N=87$) by collaborators at fish monitoring programs (Table S.1 available in the Supplement in the online version of this article) from 19 sites across the San Francisco Estuary and the Sacramento, San Joaquin, and Klamath River basins in 2018–2020 (Figure 1). Sampling covered 15 subbasins (eight-digit hydrologic unit code [HUC8]) and occurred at least once in all nominal species' ranges to increase the likelihood that all species diversity was captured in this study. The majority (70%) of samples were from larval individuals, and 50% of samples were unidentified, even to the genus level (Table S.2). Additionally, reference samples ($N=6$) were included in sequence data collection for this study. These archived tissue samples had previously been collected from adult lampreys (Pacific Lamprey and Western Brook Lamprey) with confident species identifications and included sampling locations from Oregon, Washington, and British Columbia (Table S.2). Genomic DNA was extracted from tissue samples ($N=95$) using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. Samples were prepared for Sanger sequencing of the cytochrome *b* (*cyt b*) gene, a mitochondrial marker previously used for lamprey species identification and phylogenetic analysis (Boguski et al. 2012). Cytochrome *b* sequences were amplified using forward primer Cytb-196F (5'-GCCTTYTCTTCAGTTATACA CATTG-3') and reverse primer Pro-R (5'-TAATTTAA TGTTAAGATRCTAGCTTTGG-3'; Boguski et al. 2012). The thermocycling profile used was as follows: 95°C for 2 min; followed by 45 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 2 min; and then a final hold at 72°C for 5 min. Chromatograms were manually inspected to verify base calls, and forward and reverse sequencing reads were trimmed and aligned into consensus sequences (hereafter, "contig sequences") using Geneious version 2023.0.4 (<https://www.geneious.com>).

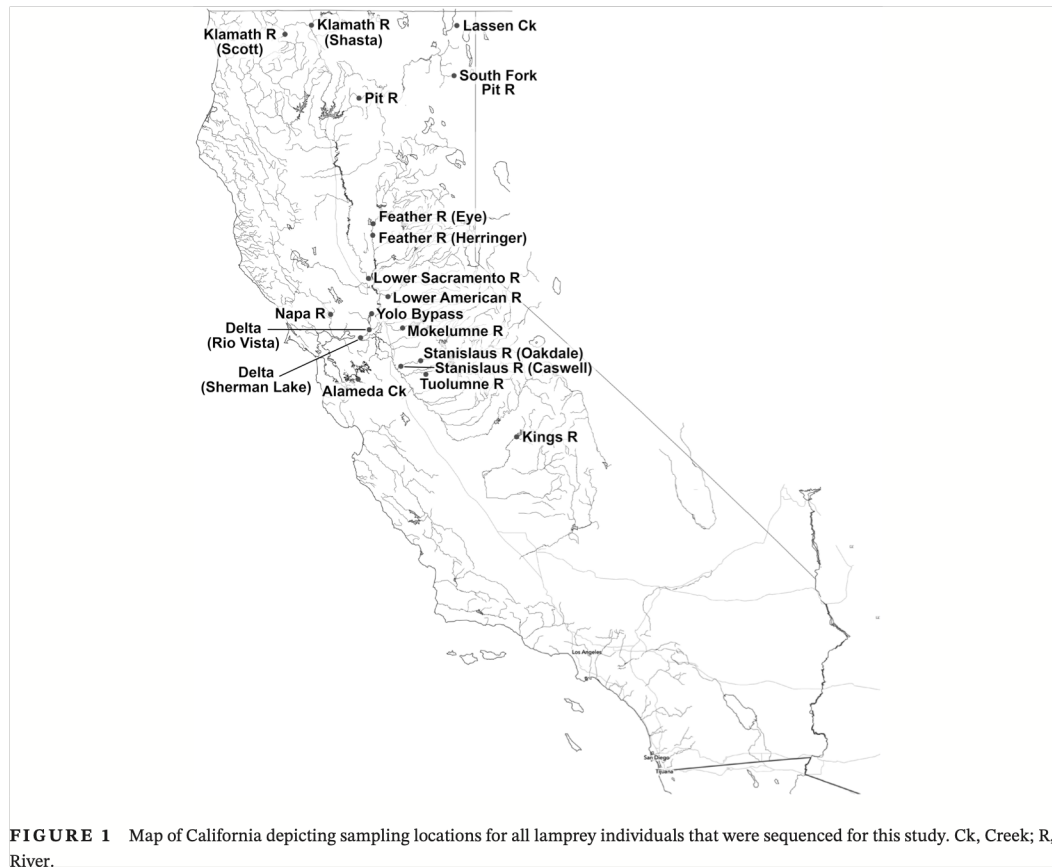


FIGURE 1 Map of California depicting sampling locations for all lamprey individuals that were sequenced for this study. Ck, Creek; R, River.

Alignment

Cytochrome *b* contig sequences from all individuals were assembled and aligned using Geneious. To enable comparisons to sequences from voucher specimens and to make comparisons across a broader geographic range, *cyt b* sequences ($N=145$) produced by Boguski et al. (2012) were downloaded from the GenBank database (Accession Numbers [GU120726](#) – [GU120870](#)) and incorporated into the alignment, resulting in 930-base-pair (bp) sequences for downstream analyses. After the addition of publicly available sequences, all individual sequences ($N=238$) were eligible for downstream analyses (Figure S.1 shows alignment in FASTA format [available in the Supplement in the online version of this article]). Notably, Boguski et al. (2012) included many samples outside of California; since this study is primarily focused on lamprey diversity within California, many subsequent analyses utilized specific subsets of the total eligible samples.

Objective 1: Identify nominal lamprey species present at study sites

To verify morphologically based field species identifications and provide species (or at least genus) identification for unknown samples, an mtDNA barcoding approach was utilized. Database searches were performed to determine which sequences in the database most closely matched the sequences from study samples. Sequences for each sample collected in California for this study ($N=87$) were compared to all sequences in the National Center for Biotechnology Information Nucleotide collection database using the Basic Local Alignment Search Tool (BLAST; Altschul et al. 1990) to assess statistical confidence of sequence similarity. The “BLASTN” program option was selected with default parameters. Species identifications were considered reliable if database records had top match statistics for lamprey *cyt b* sequences (i.e., >99.5% sequence identity, >99.5% complete query coverage, and Expect (E) values < 0.005; E-values

are the BLASTN program's estimation of the statistical significance of a match; Meiklejohn et al. 2019). To prevent potential misidentifications due to errors in the database, we validated all genetic species identifications for unknown samples by ensuring that each unknown sequence was a sequence match to at least one reference sequence with a known species identification. Reference samples for this study were (1) voucher specimens from a previous study (Boguski et al. 2012), (2) adults collected for this study with confident species identification, or (3) archived samples with confident species identification that were shared for this study ($N=6$; mentioned above).

Objective 2: Determine whether current taxonomic classifications accurately and sufficiently describe California's lamprey diversity

Phylogenetic analysis

Maximum likelihood phylogenetic analysis of the partial *cyt b* gene was conducted in RAxML version 8.2.10 (Stamatakis 2014). Specifically, rapid bootstrap analyses (option = -f a) were run with 1000 bootstrap replicates under a general time-reversible (GTR) model with a gamma distribution for incorporating rate variation among sites specified with four discrete rate categories (option = GTRGAMMA). Maximum likelihood phylogenies were estimated for (1) all *Lampetra* individuals collected within California ($N=65$), including publicly available sequences (Boguski et al. 2012); (2) all *Entosphenus* individuals collected within California ($N=62$), including publicly available sequences (Boguski et al. 2012); (3) all samples (*Entosphenus* and *Lampetra*) collected within California as well as out-of-state reference samples for the Western Brook Lamprey, Pacific Brook Lamprey *L. pacifica*, and Pacific Lamprey ($N=136$), including publicly available sequences from Boguski et al. (2012); and (4) unique haplotypes from all samples (*Entosphenus* and *Lampetra*) collected along the west coast of North America ($N=75$), including publicly available sequences from Boguski et al. (2012). See Table S.3 for additional information about the publicly available sequences used in this study. The resultant trees were visualized using the R package ggtree (Yu et al. 2017). Nodes with bootstrap proportions less than 50% were collapsed, and all nodes with greater than 70% support were displayed.

Genetic distance

All samples collected for this study within California ($N=87$) were included in the analysis of genetic distance.

Pairwise genetic distances were calculated for all pairwise comparisons of all individual sequences using the *dist.dna* function in the R package ape (Paradis and Schliep 2019) under the Kimura two-parameter (K2P) model of base substitution (also frequently called the Kimura 80 [K80] model). To calculate average pairwise K2P distances within and between two lineages, genetic distance values for all individual pairwise comparisons made between lineages were averaged (by taking the arithmetic mean). For example, to obtain the average pairwise K2P distance between Western River Lamprey and Pacific Lamprey, all pairwise genetic distance values produced from pairwise comparisons between Western River Lamprey and Pacific Lamprey individuals were averaged. Typically considered standard in barcoding analyses, the widespread use of K2P distances has faced criticism, primarily for a lack of justification of model choice (Srivathsan and Meier 2012). However, when comparing results from the best model determined by rigorous model selection to results from the K2P model, genetic distance estimates are highly similar (Collins et al. 2012). Thus, the consequences of using the K2P model, even when the model is least supported, are very minimal.

Molecular species delimitation

To delimit individuals into candidate species groupings based on *cyt b* sequences, this study utilized a distance-based method (Assemble Species by Automatic Partitioning [ASAP]; Puillandre et al. 2020) and a tree-based method (Bayesian Poisson Tree Processes [bPTP] model; Zhang et al. 2013). For these analyses, "candidate species" are defined as "mitochondrial lineages delimited as separate species by statistical methods." Each method was applied to sequence data from all individuals in the *Lampetra* genus collected in California ($N=66$), including publicly available sequences (Boguski et al. 2012; see Table S.3). This analysis was also run for the *Entosphenus* genus, but the methods performed poorly for those data, likely due to high sequence similarity across the genus. Therefore, haplotype network analysis was conducted (described below) to investigate relationships between *cyt b* haplotypes for *Entosphenus*. The ASAP program implements a hierarchical clustering algorithm that does not rely on phylogenetic reconstruction or on prior biological insight; instead, it only considers genetic distances to identify a threshold between small intraspecific differences and large interspecific differences (Puillandre et al. 2020). The other approach, bPTP, estimates branching rates from an input phylogenetic tree, determining which portions

of the tree (with certain branching rates) best fit a speciation model versus a coalescent model, and species are partitioned by maximizing the likelihood of a shift between the two categories of branching rates (Zhang et al. 2013). Both maximum likelihood and Bayesian approaches for statistical inference of species delimitations were performed separately within bPTP. To eliminate meaningless intraspecific branching events in the tree, sequences were reduced before bPTP analysis such that each unique haplotype was only represented once ($N=23$). The underlying assumption in the bPTP model is that the number of substitutions between species is significantly larger statistically than the number within species. Both programs are designed for single-locus bar code data, such as the *cyt b* sequences generated in this study, but each approach has limitations. Empirical evidence indicates that tree-based methods (i.e., bPTP) tend to oversplit species, while distance-based methods (i.e., ASAP) often undersplit species, which highlights the importance of employing multiple approaches (Luo et al. 2018).

Haplotype network

Genetic relationships within the genus *Entosphenus* were explored using a haplotype network constructed from sequences of all *Entosphenus* individuals collected for this study within California ($N=62$) via the R package pegas (Population and Evolutionary Genetics Analysis System; Paradis 2010). Sequences from Boguski et al. (2012) were not included in this analysis because that study focused on *Lampetra* and had very limited sampling of *Entosphenus*. Unique haplotypes were identified, and the number of individuals belonging to each haplotype was calculated. The network is displayed such that each uniquely colored circle represents a unique haplotype, the size of the circle is proportional to the number of individuals with that haplotype, and the length of the line between circles depends on how closely related the haplotypes are.

RESULTS

The alignment of *Lampetra* sequences had 132 variable sites and 120 parsimony-informative sites and contained 65 sequences from two nominal species. Overall nucleotide diversity at *cyt b* was 0.032 for *Lampetra* sequences. Among *Lampetra* sequences from individuals collected within California ($N=65$), 23 total unique haplotypes were identified, 13 of which were from previously published data (Boguski et al. 2012) and 10 of which were from novel

sequencing for this study. The alignment of *Entosphenus* sequences was 930bp long, had 18 variable sites and 11 parsimony-informative sites, and contained 64 sequences from three nominal species. Nucleotide diversity at *cyt b* was 0.002 for *Entosphenus* sequences. Across all *Entosphenus* sequences from individuals collected within California ($N=62$), 14 unique haplotypes were identified: one from previously published data (Boguski et al. 2012) and 13 from novel sequencing for this study.

Objective 1: Identify nominal lamprey species present at study sites

In *Lampetra*, two nominal species (Western River Lamprey and Kern Brook Lamprey) were found within their described ranges (Figure 2). Individuals identified in the field as Western River Lampreys were verified by genetic identification, which revealed that only two individual Western River Lamprey had been correctly identified in the field; five out of seven samples that were morphologically identified as Western River Lamprey were genetically identified as Pacific Lamprey (Table S.4). All individuals from the Kings River ($N=5$) were a sequence match to Kern Brook Lamprey sequences in the database. The sequences from the Napa River and Alameda Creek samples were not a match to any publicly available sequences and were most similar to *Lampetra* samples of unknown species, confirming their placement in the genus *Lampetra*.

In *Entosphenus*, three nominal species (Pacific Lamprey, Pit-Klamath Brook Lamprey, and Klamath Lamprey) were found within their described ranges (Figure 3). Genetic verification of field species identification showed that 31 out of 34 individuals identified as Pacific Lampreys in the field were also genetically identified as Pacific Lampreys (Table S.4). The *cyt b* sequences from the adult Pit-Klamath Brook Lamprey and other individuals from the Pit River, the South Fork Pit River, and Lassen Creek best matched Pit-Klamath Brook Lamprey sequences in the database but were also highly similar ($\geq 99.68\%$) to Pacific Lamprey sequences. Similarly, *cyt b* sequences from individuals identified as Klamath Lamprey in the field matched Klamath Lamprey sequences in the database but also showed high similarity ($\geq 99.5\%$) to both Pacific Lamprey and Pit-Klamath Brook Lamprey sequences.

In total, the mitochondrial *cyt b* sequence identity BLAST search confirmed 82% of the total field species identifications ($N=37$), contradicted 17% of the field species identifications ($N=8$), confirmed genus-level identity for all genus-level field identifications ($N=12$), and provided genus- or species-level identity for all unidentified samples ($N=30$; Table S.4).

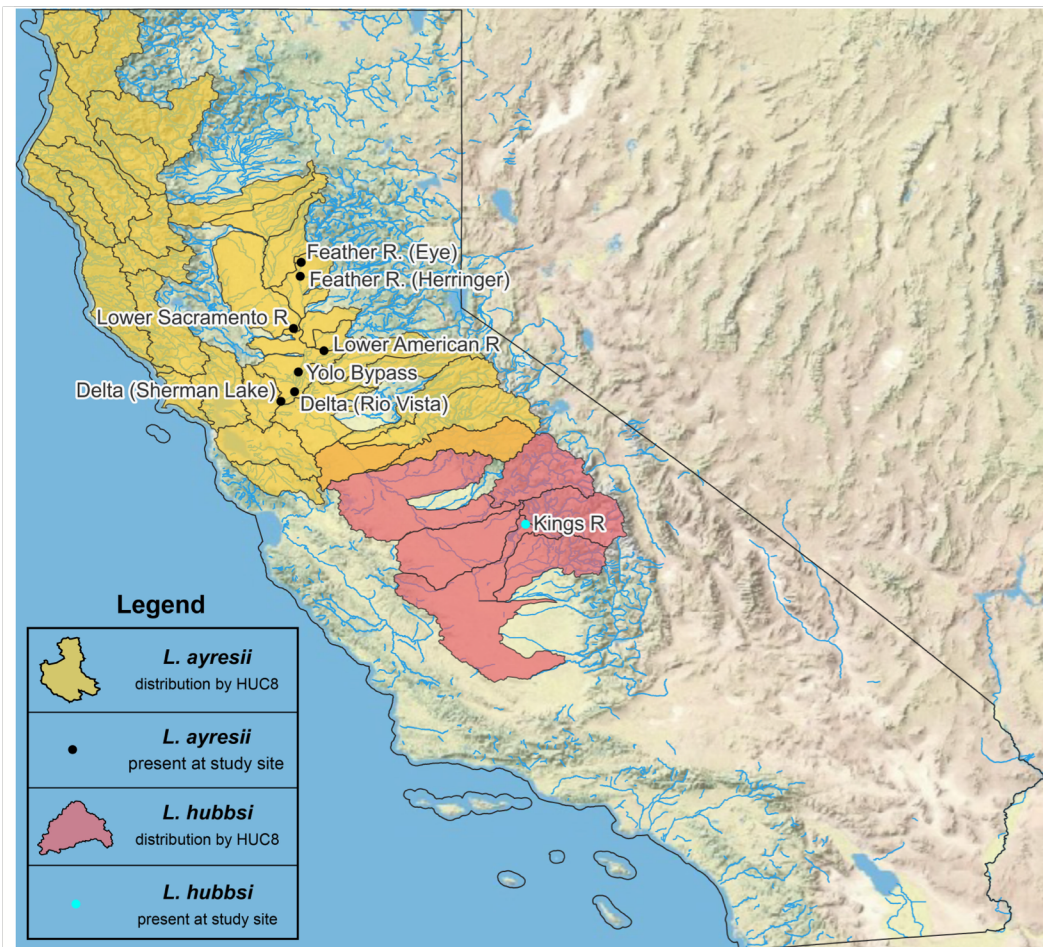


FIGURE 2 Map of California displaying all ranges for nominal species within *Lampetra* and study sites where each nominal species was found. Eight-digit hydrologic unit code (HUC8) range data are from the Western Native Fishes Committee ([databasin.org](https://data.basin.org)). R, River.

Objective 2: Determine whether current taxonomic classifications accurately and sufficiently describe California's lamprey diversity

Phylogenetic analysis

The maximum likelihood phylogeny supported both the *Lampetra* and *Entosphenus* genera as monophyletic groups with maximal (100%) bootstrap support (Figure S.2). The genera were then split, and trees were reconstructed for each genus separately (Figures 4 and 5). All Western River Lamprey in this study were part of the same clade (89% bootstrap support) that included

previously sequenced individuals from the San Francisco Bay and the Feather River (Boguski et al. 2012). The previously sequenced *Lampetra* sp. individuals from the Klamath River, Redwood Creek, and the Navarro River formed a monophyletic clade (100% bootstrap support) that was sister to Western River Lamprey collected from San Francisco Bay, the Sacramento River, and the Feather River. Kern Brook Lamprey from the Kings River and previously sequenced Kern Brook Lamprey from the Merced River formed a monophyletic group (100% bootstrap support) that was sister to one *Lampetra* sp. individual from Paynes Creek in the upper Sacramento River basin. *Lampetra* spp. from Alameda Creek, the Napa River, and the Russian River, which were previously sequenced by

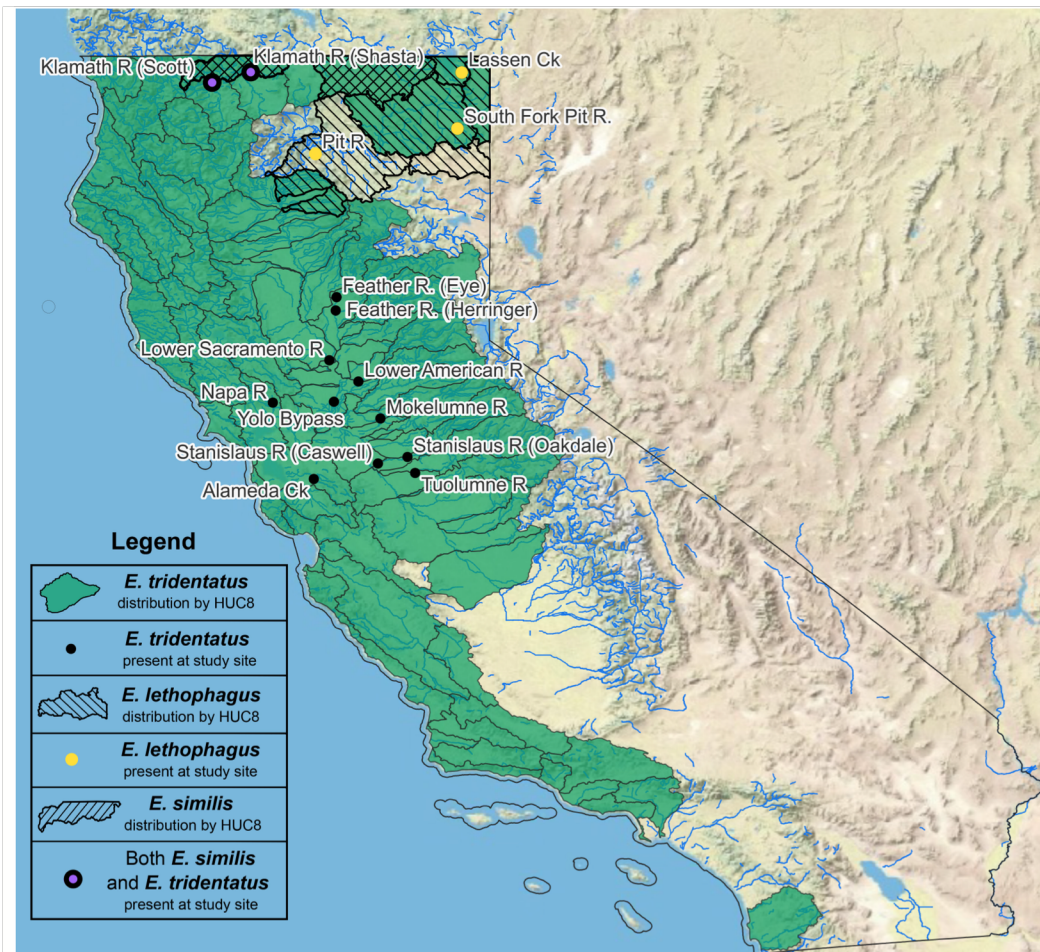


FIGURE 3 Map of California displaying all ranges for nominal species within *Entosphenus* and study sites where each species was found. Eight-digit hydrologic unit code (HUC8) range data are from the Western Native Fishes Committee (databasin.org). Ck, Creek; R, River.

Boguski et al. (2012), were part of a distinct clade (100% bootstrap support). Previously sequenced *Lampetra* sp. from Kelsey Creek in the Clear Lake basin (Boguski et al. 2012) formed a distinct monophyletic clade that was sister to all other *Lampetra* lineages in California (100% bootstrap support). Klamath Lamprey and Pit-Klamath Brook Lamprey both formed monophyletic groups (each 100% bootstrap support); within the Pit-Klamath Brook Lamprey, the individuals from the main-stem Pit River formed a distinct clade (100% bootstrap support). Several polytomies were observed throughout the *Entosphenus* phylogenetic tree, and most nodes were collapsed due to having less than 50% bootstrap support.

Genetic distance

Within the genus *Lampetra*, the average K2P distance was 3.25% (Figure 6), an order of magnitude higher than that within the genus *Entosphenus*. Genetic distances between lineages in *Lampetra* ranged from 2.95% to 5.54%, much higher than interspecific values in *Entosphenus*. The smallest average K2P distance between distinct lineages within *Lampetra* was between the Napa River and Alameda Creek lineages (2.95%), and the largest difference was between the Napa River lineage and the Western River Lamprey (5.54%; Figure 6). Average genetic distance within each species in the genus *Lampetra* ranged from

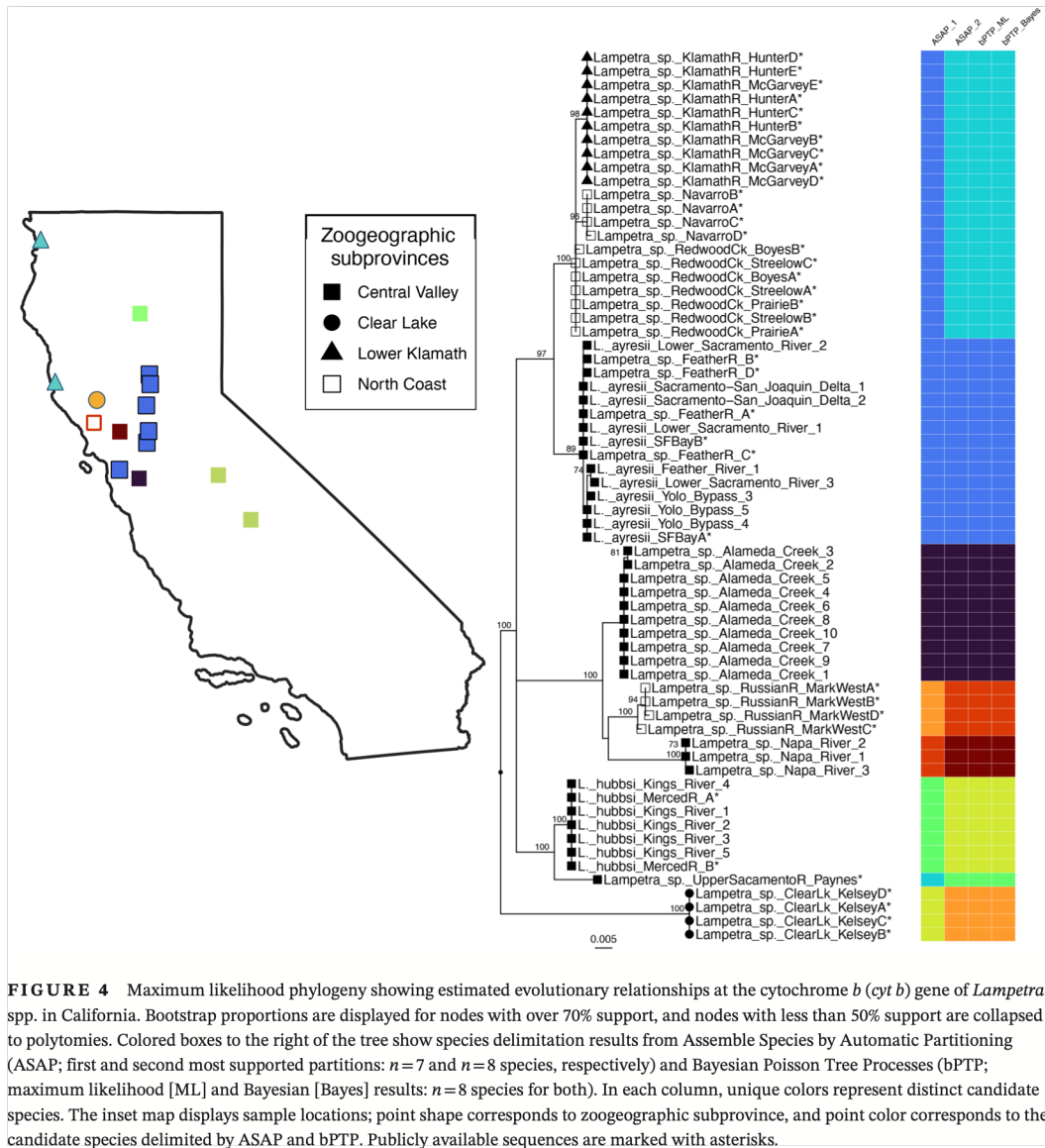


FIGURE 4 Maximum likelihood phylogeny showing estimated evolutionary relationships at the cytochrome *b* (*cyt b*) gene of *Lampetra* spp. in California. Bootstrap proportions are displayed for nodes with over 70% support, and nodes with less than 50% support are collapsed to polytomies. Colored boxes to the right of the tree show species delimitation results from Assemble Species by Automatic Partitioning (ASAP; first and second most supported partitions: $n = 7$ and $n = 8$ species, respectively) and Bayesian Poisson Tree Processes (bPTP; maximum likelihood [ML] and Bayesian [Bayes] results: $n = 8$ species for both). In each column, unique colors represent distinct candidate species. The inset map displays sample locations; point shape corresponds to zoogeographic subprovince, and point color corresponds to the candidate species delimited by ASAP and bPTP. Publicly available sequences are marked with asterisks.

0.00% (Kern Brook Lamprey) to 0.17% (Western River Lamprey).

Within *Entosphenus*, the average K2P distance was 0.21% (Figure 6). The smallest average K2P distance between lineages was between the Klamath Lamprey and Pacific Lamprey (0.30%), while the largest difference was between the Klamath Lamprey and Pit-Klamath Brook Lamprey (0.49%). Average genetic distance within each species in the genus *Entosphenus* ranged from

0.00% (Klamath Lamprey) to 0.18% (Pit-Klamath Brook Lamprey; Figure 6).

Lampetra species delimitation

All *Lampetra* sequences from individuals collected within California for this study and previously sequenced individuals collected in California from the Boguski et al. (2012)

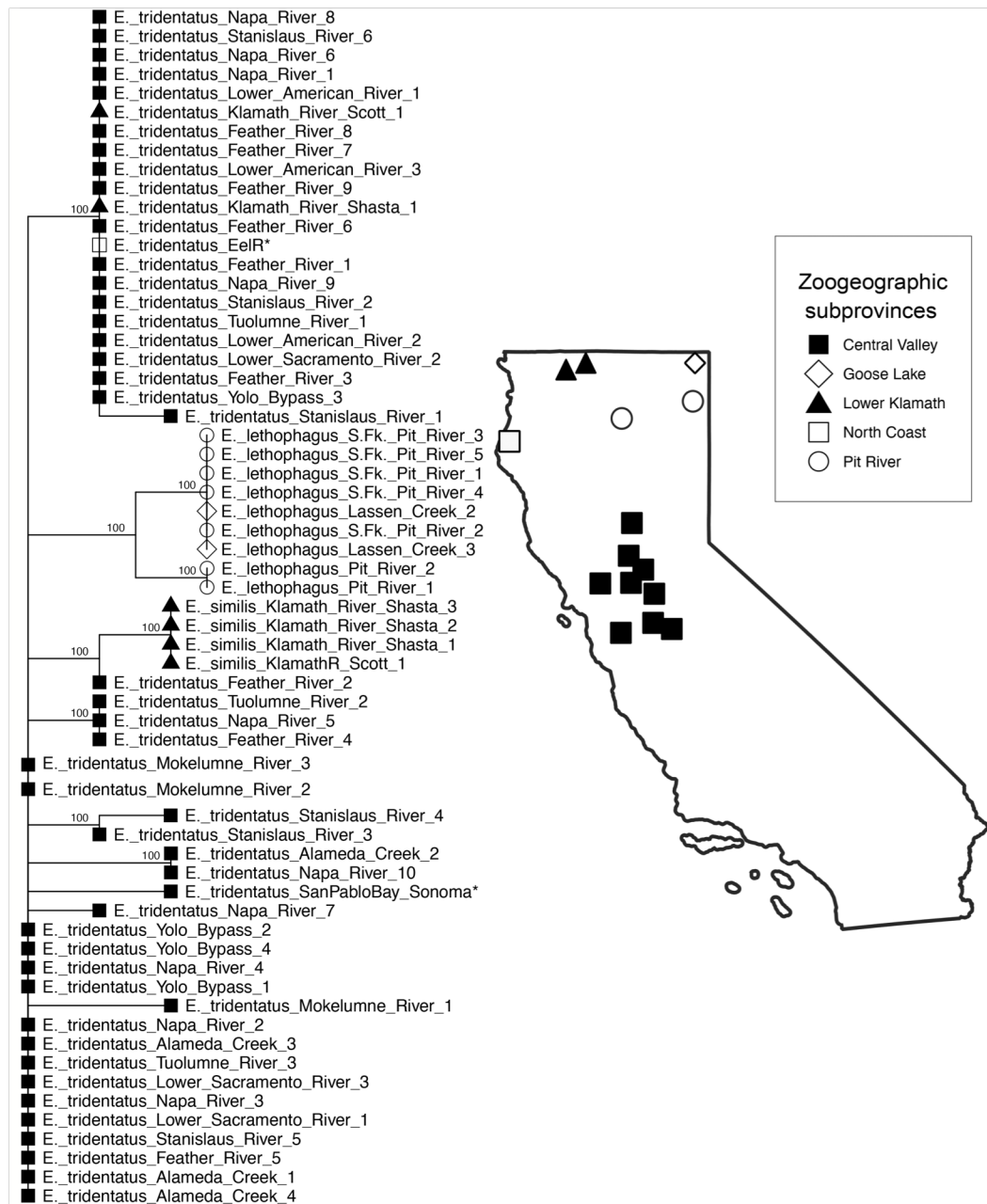


FIGURE 5 Maximum likelihood phylogeny showing estimated evolutionary relationships at the cytochrome b (*cyt b*) gene of *Entosphenus* spp. in California. The shape of each tree tip point reflects the zoogeographic subprovince from which the sample was collected. Bootstrap proportions are displayed for nodes with over 70% support, and nodes with less than 50% support are collapsed to polytomies. The inset map displays sample locations; point shape and fill indicate the zoogeographic subprovince that each sampling location falls under. Publicly available sequences are marked with asterisks.

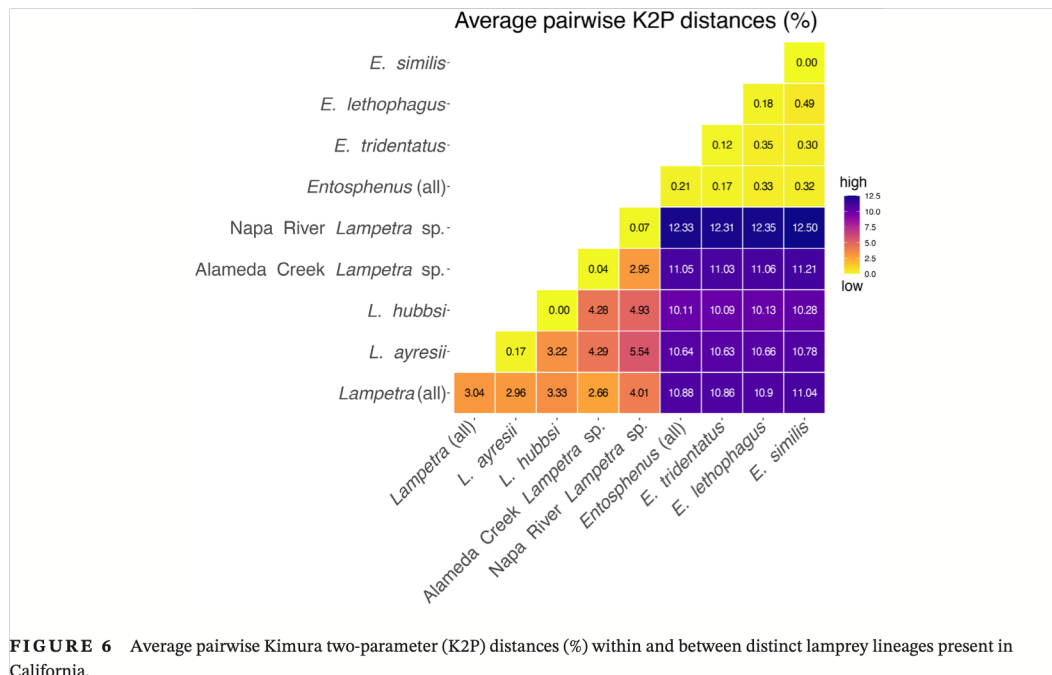


FIGURE 6 Average pairwise Kimura two-parameter (K2P) distances (%) within and between distinct lamprey lineages present in California.

study were included in this analysis ($N=65$). The highest supported result from ASAP delimited seven candidate species in *Lampetra* from California (Figure 4): (1) the Western River Lamprey from the lower Sacramento River, Feather River, and Yolo Bypass and *Lampetra* sp. from the Klamath River, Navarro River, and Redwood Creek; (2) *Lampetra* sp. from Alameda Creek; (3) *Lampetra* sp. from Mark West Creek in the Russian River basin; (4) *Lampetra* sp. from the Napa River; (5) the Kern Brook Lamprey from the middle and upper San Joaquin River basin (Merced and Kings rivers); (6) *Lampetra* sp. from Paynes Creek in the upper Sacramento River basin; and (7) *Lampetra* sp. from Kelsey Creek in the Clear Lake basin. The second most supported result from ASAP and both maximum likelihood and Bayesian results from bPTP delimited eight species (Figure 4): (1) the Western River Lamprey from the lower Sacramento River, Feather River, and Yolo Bypass; (2) *Lampetra* sp. from the Klamath River, Navarro River, and Redwood Creek; (3) *Lampetra* sp. from Alameda Creek; (4) *Lampetra* sp. from Mark West Creek in the Russian River basin; (5) *Lampetra* sp. from the Napa River; (6) the Kern Brook Lamprey from the middle and upper San Joaquin River basin (Merced and Kings rivers); (7) *Lampetra* sp. from Paynes Creek in the upper Sacramento River basin; and (8) *Lampetra* sp. from Kelsey Creek in the Clear Lake basin. The difference between the top supported and second most supported results from ASAP was the delimitation of

(1) the Western River Lamprey from the lower Sacramento River, Feather River, and Yolo Bypass and (2) *Lampetra* sp. from the Klamath River, Navarro River, and Redwood Creek as either two separate species or a single species.

Entosphenus haplotype network

All *Entosphenus* sequences from individuals collected in California for this study and publicly available *Entosphenus* sequences collected in California from the Boguski et al. (2012) study were included in this analysis ($N=62$). Fourteen unique haplotypes were identified, with two common haplotypes found in 60.9% ($N=39$) of individuals from 15 different sampling locations (Figure 7). These two common haplotypes were separated by only one mutational step. Eleven haplotypes were considered rare (i.e., found in <5% of individuals). Seven of those haplotypes were unique to one individual. All individuals from the North Coast zoogeographic province shared a single haplotype. Similarly, all individuals from the Goose Lake zoogeographic subprovince shared one haplotype (Figure 7). The Pit-Klamath Brook Lamprey contained two haplotypes: one was unique to individuals from the main-stem Pit River and the other was shared by individuals from the South Fork Pit River and Lassen Creek (Figure 7).

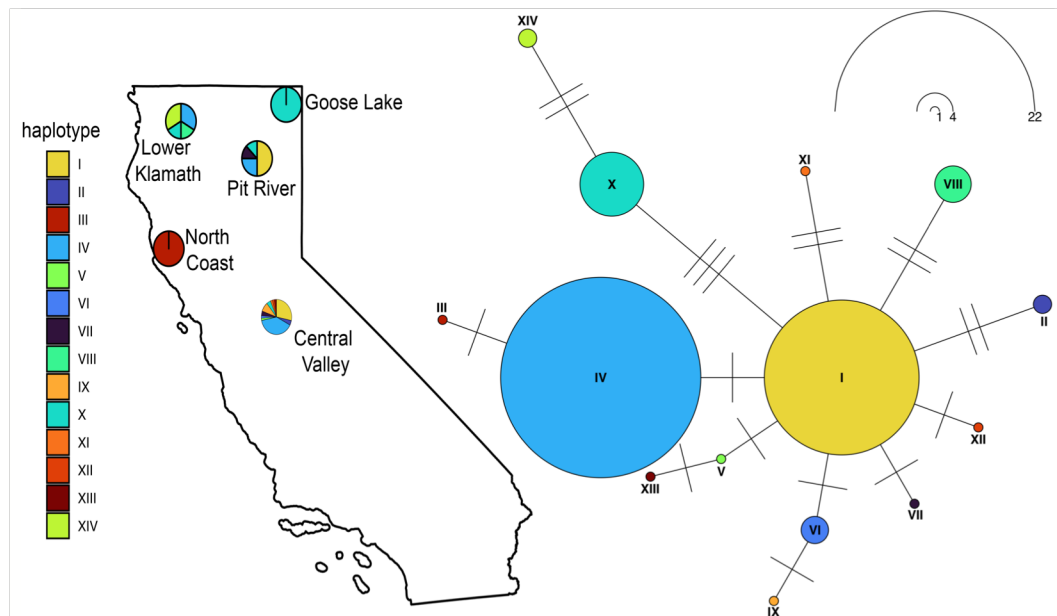


FIGURE 7 Haplotype network displaying relationships between lamprey individuals in the *Entosphenus* genus. Each uniquely colored circle represents a unique haplotype, and the size of each circle in the network is proportional to the number of individuals with each haplotype, as indicated by the scale of different-sized circles in the top right legend. Tick marks on the lines connecting circles represent mutational steps between haplotypes. The inset map shows how haplotypes are distributed across zoogeographic subprovinces in California; the circles on the map represent zoogeographic subprovinces, and the different colored segments of each circle correspond to unique haplotypes found within that zoogeographic subprovince.

DISCUSSION

Results from this study add resolution to the existing knowledge of lamprey diversity. Of the eight nominal species that are thought to occur in California (Table 1), five were observed in their described ranges (Figures 2 and 3): the Western River Lamprey, Kern Brook Lamprey, Pacific Lamprey, Pit-Klamath Brook Lamprey, and Klamath Lamprey. Results supported previous findings (Goodman et al. 2009), which demonstrated that mtDNA barcoding using the *cyt b* gene can reliably distinguish between the genera *Lampetra* and *Entosphenus* and can identify Kern Brook Lampreys to the species level. Phylogenetic and species delimitation results indicated that current taxonomic classifications likely both under- and overrepresent the genetically distinct lamprey lineages present in California. For example, results suggested that several distinct *Lampetra* lineages (i.e., those in the Napa River, Alameda Creek, Clear Lake, and the Russian River) may warrant novel species designations. Conversely, the low levels of genetic divergence between the Western River Lamprey and Western Brook Lamprey as well as the genetic similarity among

the Pacific Lamprey, Pit-Klamath Brook Lamprey, and Klamath Lamprey indicated that these nominal species might be incorrectly classified as separate species and may instead be more accurately described as ecotypes or subspecies. The biogeographic distribution of nominal species and distinct lamprey lineages is similar to patterns of other native fish taxa, which highlights the importance of biodiversity conservation in regions with high rates of endemism. This study provides a valuable initial assessment of the genetic diversity of lampreys in the study area; further research is required to reach a more complete understanding of these lineages and to determine taxonomic classifications.

Objective 1: Identify nominal lamprey species present at study sites

Results corroborated previous findings that mtDNA barcoding using the *cyt b* gene can reliably distinguish between the *Entosphenus* and *Lampetra* genera and can identify the Kern Brook Lamprey (Goodman et al. 2009). However, mtDNA barcoding using the *cyt b* gene does

not provide adequate resolution to discern between all nominal species within each lamprey genus. For example, the high *cyt b* sequence similarity among the Pacific Lamprey, Pit-Klamath Brook Lamprey, and Klamath Lamprey indicated that these three nominal species are very closely related. This points to the need for further research to determine whether these are three distinct species or represent life history variants within one species. To distinguish between these species (or ecotypes) using molecular methods, additional gene sequences or higher resolution genetic markers would need to be utilized.

Although this study was not designed to quantify abundance or distribution, patterns were observed that can inform these topics. Our results confirmed the presence of nominal species within their described ranges (Figures 2 and 3). The most abundant and wide-ranging species in our data set was the Pacific Lamprey, which made up about 50% of all individuals in this study and was found at 14 of 19 sites (Table S.2). The Western River Lamprey was the next most widespread species and was found at 5 of the 19 collection sites but was not nearly as abundant as the Pacific Lamprey, making up less than 10% of all samples (Table S.2). Results indicated that other lineages have relatively restricted geographic ranges and likely have low relative abundance compared to the two anadromous species (Pacific Lamprey and Western River Lamprey).

Objective 2: Determine whether current taxonomic classifications accurately and sufficiently describe California's lamprey diversity

We confirmed the presence of two main mitochondrial lineages of lamprey in California, divided between two genera. The distribution of diversity observed in mitochondrial sequence data from *Lampetra* and *Entosphenus* was distinctive. Within *Lampetra*, geographically restricted species-level lineages were phylogenetically structured (Figure 4) and exhibited deep genetic divergences (average K2P distance = 3.252%; Figure 6). *Entosphenus* was characterized by a lack of phylogenetic structure (Figure 5) and by shallow mtDNA divergences across a broad geographic scale, with some evidence of isolation of nominal species (average K2P distance = 0.205%; Figure 6).

Lampetra

Previous phylogenetic studies have determined that the genus *Lampetra* is polyphyletic (Brownstein and Near 2023; Carim et al. 2023). The *Lampetra* species

originating from Europe, Asia, and eastern North America form a lineage that is sister to *Eudontomyzon* and representative of the original description of this genus (Pereira et al. 2021; Brownstein and Near 2023; Carim et al. 2023). Another clade of *Lampetra* is present in western North America that diverged from a combined lineage of *Lethenteron*, *Eudontomyzon*, and other *Lampetra* prior to the Neogene and that merits recognition as a separate genus. In California, the western North American "*Lampetra*" genus (hereafter, *Lampetra*) was characterized by high levels of *cyt b* sequence differentiation across unique lineages and nominal species, except between the Western River Lamprey and Western Brook Lamprey. However, within the larger grouping of Western River Lamprey and Western Brook Lamprey, there were numerous deep divergences, and within California there were two clear divisions. With Western River Lamprey collected in California, little divergence was observed within major clades such that the Sacramento River, San Joaquin River, and San Francisco Estuary samples were highly similar in mtDNA sequences to each other, as were the northern California samples originating from Redwood Creek, the Navarro River, and the Klamath River (Figure 4). Like the Pacific Lamprey, there may be a lack of strong philopatry in this species (Moyle et al. 2009), but further investigation is required to understand the migratory behavior of Western River Lamprey. Rangewide studies have shown polyphyly of Western River Lamprey and Western Brook Lamprey individuals across multiple phylogenetic lineages (Boguski et al. 2012; Carim et al. 2023), indicating a lack of support for their current taxonomic classification as separate species. One possible resolution is to consider the Western Brook Lamprey as a synonymy of the Western River Lamprey. Another is the redescription of the Western Brook Lamprey to be representative of a northern river lamprey species. The Western River Lamprey appears to occupy disjunct distributions with six differentiated clades (Boguski et al. 2012), and we sampled two of these: Sacramento–San Joaquin rivers/San Francisco Bay and northern California. Additional Western River Lampreys are found in the Yaquina River, the Columbia River, and northward (Moyle 2002). Without fully representing Western River Lamprey across the full species range, this study does provide some evidence that supports splitting Western River Lamprey into two distinct species (Figure 4).

None of the samples collected in our study was genetically identified as the Western Brook Lamprey. Given that the Western Brook Lamprey likely represents a genetically similar life history variant of the Western River Lamprey, it is reasonable that it was not detected in this study. Although both species are thought to occur widely across the west coast, the type localities for these species

are at opposite ends of their mostly overlapping ranges. The original species description for the Western Brook Lamprey was based on specimens from Smith Creek in the Fraser River basin of British Columbia (Vladykov and Follett 1965), while the type locality for the Western River Lamprey is in the San Francisco Bay (Günther 1870). Sequences from the original type specimens were unavailable, but representative samples collected near the type localities were included in this study. Phylogenetic analysis indicates that Western Brook Lamprey specimens from the Fraser River, British Columbia, and the Columbia River are in the same clade as *Lampetra* sp. from the north coast of California (the Klamath River, Redwood Creek, and the Navarro River; Figure S.3). The Navarro River was reported to have spawning Western Brook Lamprey, and life history variation may exist within Western River Lamprey on the northern California coast, representing a freshwater-resident lamprey life history. The question, then, is “In the San Francisco Bay and Sacramento–San Joaquin River system, are there life history variants (i.e., freshwater-resident lampreys) within the Western River Lamprey that may be identified as Western Brook Lamprey based on morphology and current taxonomy but genetically identified as Western River Lampreys due to high genetic similarity between life history variants?” Watersheds where Western Brook Lamprey are thought to occur in this region include Kelsey Creek in the Clear Lake drainage (Sacramento River drainage; Moyle 2002), which clearly holds a distinct *Lampetra* lineage (Figure 4). Southern San Francisco Bay brook lamprey (species unknown) (Hubbs 1925) are represented by a population in Alameda Creek, which is also a distinct *Lampetra* lineage (Figure 4). The Russian River has also been indicated as a location where Western Brook Lamprey occur (Moyle 2002), and a distinctive *Lampetra* lineage occupies this system (Figure 4). The eastern San Joaquin Valley has already been demonstrated to contain a separate *Lampetra* species, the Kern Brook Lamprey. Instead of a close genetic relationship between life history variants, as observed between the Western River Lamprey and Western Brook Lamprey in more northerly regions of the west coast, perhaps the southern lineage of Western River Lamprey in the San Francisco Bay and associated drainages has a more distant genetic relationship with multiple freshwater-resident lamprey lineages that have previously been assumed to be Western Brook Lamprey. Such a conclusion would require further investigation of adult individuals' morphology, genotypes, and life history phenotypes; however, study results are consistent with the hypothesis that multiple freshwater-resident *Lampetra* lineages occur in the San Francisco Bay area and Sacramento–San Joaquin River drainages. We identified five candidate freshwater-resident lamprey species from mtDNA sequence data (Figure 4), and our findings

supported the recognition of the Kern Brook Lamprey, for a total of six freshwater-resident lamprey species in the *Lampetra* genus.

Novel distinct *Lampetra* lineages were identified in this study from the north and south San Francisco Bay regions. The *cyt b* sequences from *Lampetra* sp. from the Napa River and Alameda Creek were 3% different than any publicly available sequences, which is substantial considering the short length of the sequence (930 bp) and the fact that these sequences represent a highly conserved mitochondrial protein-coding gene. For reference, the Western River Lamprey and Pacific Lamprey showed an approximately 10% sequence difference at this locus, and these species belong to different genera that share a most recent common ancestor from about 60 million years ago (Brownstein and Near 2023).

The most highly supported partitions from bPTP and ASAP delimited populations from the Napa River and Alameda Creek as distinct species. This, combined with the high genetic distance from other California species and the lack of any similar publicly available *cyt b* sequences, suggests that these populations represent previously undescribed lamprey diversity. One caveat of these analyses relates to the low-resolution representation of phylogenetic relationships provided by sequences from a single gene (Degnan and Rosenberg 2009; Waters et al. 2010). The genetic patterns observed at the *cyt b* gene represent a very small fraction of the genome-wide patterns that species trees aim to portray. Another drawback of species delimitation based on mitochondrial sequence data is the maternal inheritance of mitochondrial genes, which can bias results due to differences in male and female dispersal (Petzold and Hassanin 2020; Lukic et al. 2021). Putative species delimited in these analyses can be used as an initial hypothesis that should be further investigated. An integrative taxonomic framework must incorporate these results with genome-wide data as well as morphological, biological, and ecological evidence to thoroughly assess species boundaries in *Lampetra* (Fujita et al. 2012).

Boguski et al. (2012) identified a unique mitochondrial lineage from the Clear Lake basin in central California. Sequences in the present study included new diversity but did not recover any individuals that were closely related to the unique *Lampetra* sp. population found in Clear Lake, thus further supporting the distinctiveness of the Clear Lake lineage. Phylogenetic results also substantiated this lineage as the earliest branching *Lampetra* in California (Figure 4). The Clear Lake basin, a region of high endemism, is thought to have been one of the first to become isolated in California (late Pleistocene–Holocene), promoting speciation and adaptation to Clear Lake's complex ecosystem (Hopkirk 1988). In more recent history, flooding, fires,

land use changes, mining, nutrient loading, nonnative species introductions, and high levels of contaminants have all threatened the multitude of endemic species in the Clear Lake basin (Suchanek et al. 2002). Human dimensions, especially tourism and agriculture, play an important role in Clear Lake management (Suchanek et al. 2002), but the distinct lamprey lineage, accumulation of endemic species, and numerous natural and environmental stressors in this ecosystem emphasize the need to prioritize protection of the rich native biodiversity that is present.

An unexpected finding from Boguski et al. (2012) was the close phylogenetic relationship between Kern Brook Lampreys from the Merced River and one *Lampetra* sp. individual from Paynes Creek in the upper Sacramento River basin. Phylogenetic analysis in our study also found that the *Lampetra* sp. individual from Paynes Creek was sister to Kern Brook Lampreys from the Kings and Merced rivers. Boguski et al. (2012) suggested that the range of the Kern Brook Lamprey extended into the upper Sacramento River; however, our additional sequencing of fish from the Sacramento River did not recover any sequences similar to those of the Paynes Creek sample. Additionally, species delimitation analysis identified the Paynes Creek sample as a species lineage, which suggests that this may be a distinct lineage warranting its own species designation. Considering that this pattern has only been observed based on mitochondrial sequences from a single individual, further sampling of the Paynes Creek region and other tributaries of the Sacramento River will be required to resolve the relationship and taxonomic classification of this Paynes Creek lineage.

Entosphenus

Contrary to the deep genetic divergence observed in *Lampetra*, high levels of *cyt b* sequence similarity were observed within and across nominal species in *Entosphenus* within California. The absence of strong genetic differentiation among the Pacific Lamprey, Klamath Lamprey, and Pit-Klamath Brook Lamprey despite stark morphological and ecological differences may be the result of ongoing gene flow between these groups, may represent a case of high phenotypic plasticity, or may indicate a relatively recent divergence. The Pacific Lamprey and Pit-Klamath Brook Lamprey, a presumed species pair, showed high similarity at the *cyt b* gene, but higher resolution genomic analyses and population genetic sampling would help to clarify the relationship between these groups. Within the Pacific Lamprey, very low to no genetic structure was observed across geographic sites, which supports the hypothesized lack of strong philopatry in this species (Hatch

and Whiteaker 2009; Spice et al. 2012; Hess et al. 2020). Although this study is focused on California, prior research also found high levels of gene flow among Pacific Lamprey populations across the west coast of North America. However, at this larger geographic scale, genetic structure and local adaptation were observed (Spice et al. 2012; Hess et al. 2020). In the present study, no Pacific Lamprey individuals were found in the Pit River basin or the Goose Lake basin, suggesting that there is an impassable barrier to upstream migration between the lower Sacramento River and the Pit River. This finding is consistent with the lack of connectivity between the Pit River and lower Sacramento River for other fish populations, likely due to the construction of the Shasta Dam in 1945 (Schick and Lindley 2007).

Objective 3: Compare biogeographic distribution of lamprey diversity to patterns observed in other native taxa

At a broad scale, species diversity occurs on a latitudinal cline, being highest at the equator and declining toward the polar regions (Hillebrand 2004; van Humboldt 1828). Viewing western North American lamprey diversity in this framework, a higher amount of species diversity should occur in the southern range (i.e., California) of the group, declining to the north. Our work and other studies substantiate this trend (Boguski et al. 2012). Overall high rates of endemism occur in California; over 80% of native fishes in the state are endemic to California or watersheds shared with neighboring states (Leidy and Moyle 2021). This richness of endemic species may be further explained by other aspects of California's geography and climate. The prevalence of intermittent streams, which frequently cause fluctuating stream fragmentation, may increase the likelihood of fish populations becoming isolated, leading to high rates of endemism (Ball et al. 2013). Furthermore, the state's rugged topography results in many isolated watersheds (Moyle 2002).

The distribution of lamprey diversity across the landscape was concordant with broader zoogeographic patterns of California native fish diversity, suggesting that the diversification of extant lampreys was shaped by biogeographic characteristics similar to other fish lineages. Moyle (2002) described six major zoogeographic provinces in California: Klamath, Sacramento–San Joaquin, North Coast, Great Basin, South Coast, and Colorado River, with various subprovinces in each (22 total subprovinces). Within the Sacramento–San Joaquin province, nominal lamprey species with restricted ranges were found in the upper Kern River subprovince (Kern Brook Lamprey) and the Pit River

and Goose Lake subprovinces (Pit-Klamath Brook Lamprey). Additionally, the range-restricted Klamath Lamprey was found in the Upper and Lower Klamath subprovince. Beyond nominal species, distinct lamprey populations were found in the Clear Lake subprovince (*Lampetra* sp. from Kelsey Creek; Boguski et al. 2012), the Central Valley subprovince (*Lampetra* sp. from Alameda Creek and the Napa River), and the North Coast province (*Lampetra* sp. from Mark West Creek; Boguski et al. 2012). Additionally, species delimitation results suggested that *Lampetra* sp. from the North Coast province may be distinct from Western River Lampreys in the Sacramento–San Joaquin province. These pockets of distinct lampreys and other native fishes highlight the importance of protecting ecosystems in regions of California with high biodiversity and endemism.

CONCLUSIONS AND MANAGEMENT IMPLICATIONS

In conclusion, results of these analyses showed that the genus *Lampetra* harbors high diversity in California and contains very divergent lineages in comparison to the genus *Entosphenus*. Previously unrecognized lamprey diversity within *Lampetra* was identified in the Napa River and Alameda Creek; these populations are distinct from the novel unique lineages found in Mark West, Kelsey, and Paynes creeks by Boguski et al. (2012). Additional lines of evidence are required to determine the appropriate taxonomic classification for these populations, but the level of genetic differentiation at *cyt b* suggests that these lineages should at least be treated as independent management units. The HUCs, part of a standard system developed by the U.S. Geological Survey for defining hydrologic areas (Seaber et al. 1987), can be a helpful tool in the initial stages of management unit delineation. In this study, lamprey populations often showed genetic differentiation between basins (six-digit HUCs [HUC6]), but little to no genetic differentiation was observed between subbasins (HUC8), suggesting that basins (HUC6) are the more appropriate hydrological unit for management of lamprey populations in the study area. Notably, results indicated that multiple species frequently co-occur within basins (HUC6), so accurate species identification will be especially important in efforts to characterize the abundance and distribution of lampreys. Since lampreys are most frequently sampled in the field during the larval phase, when they may lack distinguishing morphological features, genetic species identification would improve the accuracy of data collected and would help to advance our understanding

of species distributions (Docker et al. 2016; Clemens et al. 2022). The genetic marker used for DNA barcoding in this study (*cyt b*) could not distinguish between all nominal lamprey species in California, but molecular identification of different lamprey species (or ecotypes) is likely feasible with higher resolution genomic sequencing techniques. Although it also may be possible to distinguish between species by using a combination of several genetic markers for DNA barcoding, genomic sequencing provides higher resolution data and may be more cost effective due to advances in technology. Goals for addressing long-term stability of native species are complicated by the compounding impacts of human modifications, resource use, and climate warming. However, maintaining lamprey species complexity and fostering resilience cannot begin without an understanding of their underlying genetic diversity.

ACKNOWLEDGMENTS

We thank all of our collaborators across the state who voluntarily collected samples in order to make this project possible: Ed Rible (East Bay Municipal Utility District); Brian Sak, Randall Renn, and all Natural Resources and Lands Management biologists with the San Francisco Public Utilities Commission; Jason Guignard (FISHBIO); Dave Contreras (California Department of Fish and Wildlife [CDFW]); Jonathan Koehler (Napa County Resource Conservation District); Jason Julienne (CDFW, North Central Region [Region 2]); Chris Cook (California Department of Water Resources [CDWR], Division of Integrated Science and Engineering, Feather River Program); and all team members who assisted in lamprey sample collection for this study. Additionally, we thank Jon Hess (Columbia River Inter-Tribal Fish Commission) for generously donating archived tissue samples for this study. Support for this project also came from colleagues in the Genomic Variation Laboratory, University of California, Davis, especially the lab director, Andrea Schreier. This publication was prepared by G. Auringer under National Oceanic and Atmospheric Administration (NOAA) Grant NA22OAR4170106, California Sea Grant College Program Project R/HCE-26F, through the NOAA National Sea Grant College Program, U.S. Department of Commerce. Additional funding was provided by CDWR. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of California Sea Grant, NOAA, the U.S. Department of Commerce, or CDWR.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest regarding the publication of this paper.

ETHICS STATEMENT

This research study was conducted in compliance with ethical principles and guidelines. Tissue samples for this project were obtained from individuals listed in the acknowledgements. Sample collections for this study followed institutional guidelines and permitting. Previously published DNA sequence data included in this study was accessed on a publicly available database.

DATA AVAILABILITY STATEMENT

All data (genetic sequences) generated as part of this study are publicly available through GenBank (Accession Numbers [OQ818491–OQ818585](#)).

ORCID

Grace Auringer  <https://orcid.org/0000-0002-7639-3766>

Matthew A. Campbell  <https://orcid.org/0000-0002-5826-0329>

Pascale A. L. Goertler  <https://orcid.org/0000-0001-6259-5108>

Amanda J. Finger  <https://orcid.org/0000-0003-3850-3685>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Ball, J. E., Bêche, L. A., Mendez, P. K., & Resh, V. H. (2013). Biodiversity in Mediterranean-climate streams of California. *Hydrobiologia*, 719, 187–213. <https://doi.org/10.1007/s10750-012-1368-6>
- Boeker, C., & Geist, J. (2016). Lampreys as ecosystem engineers: Burrows of *Eudontomyzon* sp. and their impact on physical, chemical, and microbial properties in freshwater substrates. *Hydrobiologia*, 777, 171–181. <https://doi.org/10.1007/s10750-016-2774-y>
- Boguski, D. A., Reid, S. B., Goodman, D. H., & Docker, M. F. (2012). Genetic diversity, endemism and phylogeny of lampreys within the genus *Lampetra sensu stricto* (Petromyzontiformes: Petromyzontidae) in western North America. *Journal of Fish Biology*, 81(6), 1891–1914. <https://doi.org/10.1111/j.1095-8649.2012.03417.x>
- Brown, L. R., & Bauer, M. L. (2010). Effects of hydrologic infrastructure on flow regimes of California's Central Valley rivers: Implications for fish populations. *River Research and Applications*, 26(6), 751–765.
- Brownstein, C. D., & Near, T. J. (2023). Phylogenetics and the Cenozoic radiation of lampreys. *Current Biology*, 33(2), 397–404. <https://doi.org/10.1016/j.cub.2022.12.018>
- Carim, K. J., Larson, D. C., Helstab, J. M., Young, M. K., & Docker, M. F. (2023). A revised taxonomy and estimate of species diversity for western North American *Lampetra*. *Environmental Biology of Fishes*, 106, 817–836.
- Clemens, B. J., Harris, J. E., Starcevich, S. J., Evans, T. M., Skalicky, J. J., Neave, F., & Lampman, R. T. (2022). Sampling methods and survey designs for larval lampreys. *North American Journal of Fisheries Management*, 42(2), 455–474. <https://doi.org/10.1002/nafm.10762>
- Clemens, B. J., & Wade, J. (2023). Conservation biology of the *Lampetra* satellite species of western North America, with a focus on Western Brook Lamprey (*L. richardsoni*). *Canadian Manuscript Report of Fisheries and Aquatic Sciences* 3258. Fisheries and Oceans Canada.
- 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](https://doi.org/10.1577/1548-8446(2002)027<0019:TEACIO>2.0.CO;2)
- Collins, R. A., Boykin, L. M., Cruickshank, R. H., & Armstrong, K. F. (2012). Barcoding's next top model: An evaluation of nucleotide substitution models for specimen identification. *Methods in Ecology and Evolution*, 3(3), 457–465. <https://doi.org/10.1111/j.2041-210X.2011.00176.x>
- Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24(6), 332–340. <https://doi.org/10.1016/j.tree.2009.01.009>
- Docker, M. F. (2009). A review of the evolution of nonparasitism in lampreys and an update of the paired species concept. In L. R. Brown, S. Chase, M. Mesa, R. Beamish, & P. Moyle (Eds.), *Biology, management, and conservation of lampreys in North America* (Symposium 72, pp. 71–114). American Fisheries Society. <https://doi.org/10.47886/9781934874134>
- Docker, M. F., & Hume, J. B. (2019). There and back again: Lampreys in the 21st century and beyond. In M. F. Docker (Ed.), *Lampreys: Biology, conservation and control* (Fish & Fisheries 38, pp. 527–570). Springer. https://doi.org/10.1007/978-94-024-1684-8_7
- Docker, M. F., Mandrak, N. E., & Heath, D. D. (2012). Contemporary gene flow between “paired” Silver (*Ichthyomyzon unicuspis*) and Northern Brook (*I. fossor*) lampreys: Implications for conservation. *Conservation Genetics*, 13, 823–835. <https://doi.org/10.1007/s10592-012-0332-3>
- Docker, M. F., Silver, G. S., Jolley, J. C., & Spice, E. K. (2016). Simple genetic assay distinguishes lamprey genera *Entosphenus* and *Lampetra*: Comparison with existing genetic and morphological identification methods. *North American Journal of Fisheries Management*, 36(4), 780–787. <https://doi.org/10.1080/02755947.2016.1167146>
- Dunkle, M. R., Lampman, R. T., Jackson, A. D., & Caudill, C. C. (2020). Factors affecting the fate of Pacific Lamprey carcasses and resource transport to riparian and stream macrohabitats. *Freshwater Biology*, 65(8), 1429–1439. <https://doi.org/10.1111/fwb.13510>
- Ely, C. V., de Loreto Bordignon, S. A., Trevisan, R., & Boldrini, I. I. (2017). Implications of poor taxonomy in conservation. *Journal for Nature Conservation*, 36, 10–13.
- Endangered and Threatened Wildlife and Plants; 90-Day Finding on a Petition to List Three Species of Lampreys as Threatened or Endangered, 50 C.F.R. § 17 (2004). <https://www.govinfo.gov/content/pkg/FR-2004-12-27/pdf/04-28167.pdf>
- Espanhol, R., Almeida, P. R., & Alves, M. J. (2007). Evolutionary history of lamprey paired species *Lampetra fluviatilis* (L.) and *Lampetra planeri* (Bloch) as inferred from mitochondrial DNA variation. *Molecular Ecology*, 16(9), 1909–1924. <https://doi.org/10.1111/j.1365-294X.2007.03279.x>
- Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27(9), 480–488. <https://doi.org/10.1016/j.tree.2012.04.012>
- Goertler, P. A. L., Shakya, A. W., Seesholtz, A. M., Schreier, B. M., Matica, S. Z., & Holley, K. S. (2019). Lamprey (*Entosphenus* sp. and *Lampetra* sp.) estuarine occupancy is regionally variable

- and constrained by temperature. *Journal of Fish Biology*, 96(2), 527–532. <https://doi.org/10.1111/jfb.14143>
- Goodman, D. H., Kinziger, A. P., Reid, S. B., & Docker, M. F. (2009). Morphological diagnosis of *Entosphenus* and *Lampetra* ammocoetes (Petromyzontidae) in Washington, Oregon, and California. In L. R. Brown, S. Chase, M. Mesa, R. Beamish, & P. Moyle (Eds.), *Biology, management, and conservation of lampreys in North America* (Symposium 72, pp. 223–232). American Fisheries Society. <https://doi.org/10.47886/9781934874134>
- Günther, A. E. (1870). *Catalogue of the physostomi: Containing the families gymnotidae, symbranchidae, murænidae, pegasidae, and of the lophobranchii, plectognathi, dipnoi, ganoidei, chondropterygii, cyclostomata, leptocardi, in the British Museum* (Vol. 8). Taylor and Francis.
- Hamilton, J. B., Curtis, G. L., Snedaker, S. M., & White, D. K. (2005). Distribution of anadromous fishes in the upper Klamath River watershed prior to hydropower dams—A synthesis of the historical evidence. *Fisheries*, 30(4), 10–20. [https://doi.org/10.1577/1548-8446\(2005\)30\[10:DOAFIT\]2.0.CO;2](https://doi.org/10.1577/1548-8446(2005)30[10:DOAFIT]2.0.CO;2)
- Hatch, D. R., & Whiteaker, J. M. (2009). A field study to investigate repeat homing in Pacific Lampreys. In L. R. Brown, S. Chase, M. Mesa, R. Beamish & P. Moyle (Eds.), *Biology, management, and conservation of lampreys in North America* (Symposium 72, pp. 72–191). American Fisheries Society. <https://doi.org/10.47886/9781934874134>
- Hess, J. E., Smith, J. J., Timoshevskaya, N., Baker, C., Caudill, C. C., Graves, D., Keefer, M. L., Kinziger, A. P., Moser, M. L., Porter, L. L., Silver, G., Whitlock, S. L., & Narum, S. R. (2020). Genomic islands of divergence infer a phenotypic landscape in Pacific Lamprey. *Molecular Ecology*, 29(20), 3841–3856. <https://doi.org/10.1111/mec.15605>
- Hillebrand, H. (2004). Strength, slope and variability of marine latitudinal gradients. *Marine Ecology Progress Series*, 273, 251–267. <https://doi.org/10.3354/meps273251>
- Hopkirk, J. D. (1988). Fish evolution and the late Pleistocene and Holocene history of Clear Lake, California. *Geological Society of America Special Papers*, 214, 183–194.
- Hubbs, C. L. (1925). The life cycle and growth of lampreys. *Papers of the Michigan Academy of Science*, 4, 587–603.
- Kneibelsberger, T., Dunz, A. R., Neumann, D., & Geiger, M. F. (2015). Molecular diversity of Germany's freshwater fishes and lampreys assessed by DNA barcoding. *Molecular Ecology Resources*, 15(3), 562–572. <https://doi.org/10.1111/1755-0998.12322>
- Leidy, R. A., & Moyle, P. B. (2021). Keeping up with the status of freshwater fishes: A California (USA) perspective. *Conservation Science and Practice*, 3(8), Article e474. <https://doi.org/10.1111/csp2.474>
- Lucas, M. C., Hume, J. B., Almeida, P. R., Aronsuu, K., Habit, E., Silva, S., Wang, C. J., & Zampatti, B. (2021). Emerging conservation initiatives for lampreys: Research challenges and opportunities. *Journal of Great Lakes Research*, 47, S690–S703. <https://doi.org/10.1016/j.jglr.2020.06.004>
- Lukic, D., Eberle, J., Thormann, J., Holzschuh, C., & Ahrens, D. (2021). Excluding spatial sampling bias does not eliminate oversplitting in DNA-based species delimitation analyses. *Ecology and Evolution*, 11(15), 10327–10337. <https://doi.org/10.1002/ece3.7836>
- Luo, A., Ling, C., Ho, S. Y., & Zhu, C. D. (2018). Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology*, 67(5), 830–846. <https://doi.org/10.1093/sysbio/syy011>
- Maitland, P. S., Renaud, C. B., Quintella, B. R., Close, D. A., & Docker, M. F. (2015). Conservation of native lampreys. In M. F. Docker (Ed.), *Lampreys: Biology, conservation and control* (Fish & Fisheries 37, pp. 375–428). Springer. https://doi.org/10.1007/978-94-017-9306-3_8
- McNeely, J. A. (2002). The role of taxonomy in conserving biodiversity. *Journal for Nature Conservation*, 10(3), 145–153. <https://doi.org/10.1078/1617-1381-00015>
- Meiklejohn, K. A., Damaso, N., & Robertson, J. M. (2019). Assessment of BOLD and GenBank—Their accuracy and reliability for the identification of biological materials. *PLoS One*, 14(6), Article e0217084. <https://doi.org/10.1371/journal.pone.0217084>
- Moser, M. L., Butzerin, J. M., & Dey, D. B. (2007). Capture and collection of lampreys: The state of the science. *Reviews in Fish Biology and Fisheries*, 17, 45–56. <https://doi.org/10.1007/s11160-006-9037-3>
- Moyle, P. B. (2002). *Inland fishes of California: Revised and expanded*. University of California Press.
- Moyle, P. B., Brown, L. R., Chase, S. D., & Quiñones, R. M. (2009). Status and conservation of lampreys in California. In L. R. Brown, S. Chase, M. Mesa, R. Beamish, & P. Moyle (Eds.), *Biology, management, and conservation of lampreys in North America* (Symposium 72, pp. 279–292). American Fisheries Society. <https://doi.org/10.47886/9781934874134>
- Moyle, P. B., Quinones, R. M., Katz, J. V., & Weaver, J. (2015). *Fish species of special concern in California* (3rd ed.). California Department of Fish and Wildlife. <https://doi.org/10.47886/9781934874134>
- Page, L. M., Espinosa-Pérez, H., Findley, L. T., Gilbert, C. R., Lea, R. N., Mandrak, N. E., & Mayden, R. L. (2013). New seventh edition of *Common and Scientific Names of Fishes*: Changes include capitalization of common names. *Fisheries*, 38(4), 188–189. <https://doi.org/10.1080/03632415.2013.767244>
- Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Parker, C. M. (2022). *Spatial and temporal patterns of detection of native lampreys in Central Valley California: A starting point* [Master's thesis, Oregon State University].
- Pereira, A. M., Levy, A., Vukić, J., Šanda, R., Levin, B. A., Freyhof, J., Geiger, M., Choleva, L., Francisco, S. M., & Robalo, J. I. (2021). Putting European lampreys into perspective: A global-scale multilocus phylogeny with a proposal for a generic structure of the Petromyzontidae. *Journal of Zoological Systematics and Evolutionary Research*, 59(8), 1982–1993. <https://doi.org/10.1111/jzs.12522>
- Petzold, A., & Hassani, A. (2020). A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia, Cetartiodactyla). *PLoS One*, 15(2), Article e0217956. <https://doi.org/10.1371/journal.pone.0217956>
- Potter, I. C., Gill, H. S., Renaud, C. B., & Haoucher, D. (2015). The taxonomy, phylogeny, and distribution of lampreys. In M. F. Docker (Ed.), *Lampreys: Biology, conservation and control* (Fish & Fisheries 37, pp. 35–73). Springer. https://doi.org/10.1007/978-94-017-9306-3_2
- Puillandre, N., Brouillet, S., & Achaz, G. (2020). ASAP: Assemble species by automatic partitioning. *Molecular Ecology*

- Resources, 21(2), 609–620. <https://doi.org/10.1111/1755-0998.13281>
- Rannala, B., & Yang, Z. (2020). Species delimitation. In C. Scornavacca, F. Delsuc, & N. Galtier (Eds.), *Phylogenetics in the genomic era* (pp. 5.5:1–5.5:18). Self-published.
- Reid, S. B., & Goodman, D. H. (2016). Pacific Lamprey in coastal drainages of California: Occupancy patterns and contraction of the southern range. *Transactions of the American Fisheries Society*, 145(4), 703–711. <https://doi.org/10.1080/00028487.2016.1159615>
- Renaud, C. B. (1997). Conservation status of northern hemisphere lampreys (Petromyzontidae). *Journal of Applied Ichthyology*, 13(3), 143–148. <https://doi.org/10.1111/j.1439-0426.1997.tb00114.x>
- Renaud, C. B. (2011). *Lampreys of the world: An annotated and illustrated catalogue of lamprey species known to date* (Species Catalogue for Fishery Purposes No. 5). Food and Agriculture Organization of the United Nations.
- Rougemont, Q., Dolo, V., Oger, A., Besnard, A. L., Huteau, D., Coutellec, M. A., Perrier, C., Launey, S., & Evanno, G. (2021). Riverscape genetics in Brook Lamprey: Genetic diversity is less influenced by river fragmentation than by gene flow with the anadromous ecotype. *Heredity*, 126(2), 235–250. <https://doi.org/10.1038/s41437-020-00367-9>
- Schick, R. S., & Lindley, S. T. (2007). Directed connectivity among fish populations in a riverine network. *Journal of Applied Ecology*, 44(6), 1116–1126. <https://doi.org/10.1111/j.1365-2664.2007.01383.x>
- Schreiber, A., & Engelhorn, R. (1998). Population genetics of a cyclostome species pair, River Lamprey (*Lampetra fluviatilis* L.) and Brook Lamprey (*Lampetra planeri* Bloch). *Journal of Zoological Systematics and Evolutionary Research*, 36(1–2), 85–99.
- Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). *Hydrologic unit maps* (Water-Supply Paper 2294). U.S. Geological Survey.
- Sheoships, G. (2014). *Pacific Lamprey Entosphenus tridentatus: Integrating traditional ecological knowledge and contemporary values into conservation planning, and stream substrate associations with larval abundance in the Willamette River basin, Oregon, USA* [Master's thesis, Oregon State University].
- Shirakawa, H., Yanai, S., & Goto, A. (2013). Lamprey larvae as ecosystem engineers: Physical and geochemical impact on the streambed by their burrowing behavior. *Hydrobiologia*, 701, 313–322. <https://doi.org/10.1007/s10750-012-1293-8>
- 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>
- Srivathsan, A., & Meier, R. (2012). On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, 28(2), 190–194. <https://doi.org/10.1111/j.1096-0031.2011.00370.x>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Suchanek, T. H., Richerson, P. J., Nelson, D. C., Eagles-Smith, C. A., Anderson, D. W., Cech, J. J., Zierenberg, R., Schladow, G., Mount, J. F., McHatton, S. C., Slotten, D. G., Webber, L. B., Swisher, B. J., Bern, A. L., & Sexton, M. (2002). Evaluating and managing a multiply stressed ecosystem at Clear Lake, California: A holistic ecosystem approach. In D. J. Rapport, B. L. Lasley, D. E. Rolston, N. O. Nielsen, C. O. Qualset, & A. B. Dalmania (Eds.), *Managing for healthy ecosystems* (pp. 1239–1272). CRC Press.
- van Humboldt, A. (1828). *Über das Universum: Die Kosmosvorträge 1827/28 in der Berliner Singakademie [About the universe: The cosmos lectures 1827/28 in the Berlin Singing Academy]*. Insel Verlag.
- Vladykov, V. D., & Follett, W. I. (1965). *Lampetra richardsoni*, a new nonparasitic species of lamprey (Petromyzontidae) from western North America. *Journal of the Fisheries Board of Canada*, 22(1), 139–158. <https://doi.org/10.1139/f65-013>
- Wang, C., & Schaller, H. (2015). Conserving Pacific Lamprey through collaborative efforts. *Fisheries*, 40(2), 72–79. <https://doi.org/10.1080/03632415.2014.996871>
- Waters, J. M., Rowe, D. L., Burrige, C. P., & Wallis, G. P. (2010). Gene trees versus species trees: Reassessing life-history evolution in a freshwater fish radiation. *Systematic Biology*, 59(5), 504–517. <https://doi.org/10.1093/sysbio/syq031>
- Whipple, A. A., Grossinger, R. M., Rankin, D., Stanford, B., & Askevold, R. (2012). *Sacramento-San Joaquin Delta historical ecology investigation: Exploring pattern and process* (Contribution No. 672). San Francisco Estuary Institute.
- Wicks-Arshack, A., Dunkle, M., Matsaw, S., & Caudill, C. (2018). An ecological, cultural and legal review of Pacific Lamprey in the Columbia River basin. *Idaho Law Review*, 54, 45–91.
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics*, 13(5), 303–314. <https://doi.org/10.1038/nrg3186>
- Yoshiyama, R. M., Gerstung, E. R., Fisher, F. W., & Moyle, P. B. (2001). Historical and present distribution of Chinook Salmon in the Central Valley drainage of California. *Fish Bulletin*, 179(1), 71–176.
- Young, M. K., Smith, R., Pilgrim, K. L., & Schwartz, M. K. (2021). Molecular species delimitation refines the taxonomy of native and nonnative physine snails in North America. *Scientific Reports*, 11(1), Article 21739. <https://doi.org/10.1038/s41598-021-01197-3>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. <https://doi.org/10.1111/2041-210X.12628>
- Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22), 2869–2876. <https://doi.org/10.1093/bioinformatics/btt499>
- Zimmerman, J. K., Carlisle, D. M., May, J. T., Klausmeyer, K. R., Grantham, T. E., Brown, L. R., & Howard, J. K. (2018). Patterns and magnitude of flow alteration in California, USA. *Freshwater Biology*, 63(8), 859–873. <https://doi.org/10.1111/fwb.13058>

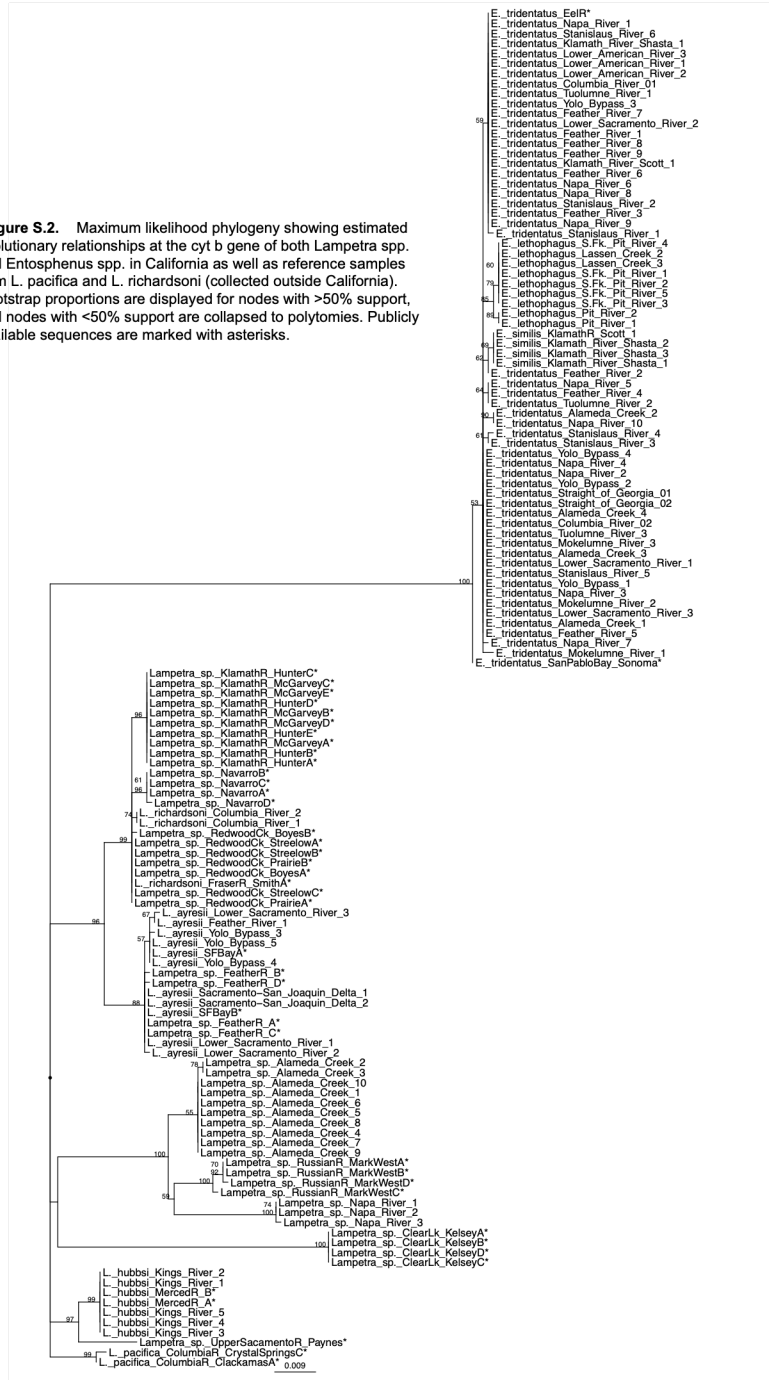
SUPPORTING INFORMATION

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

Chapter 1: Supplementary Files

Figure S.1. (attached separately as .fasta file) Multiple sequence alignment (FASTA format) of all *cyt b* sequences generated for this study and publicly available lamprey *cyt b* sequences from Boguski et al. 2012 included in this study.

Figure S.2. Maximum likelihood phylogeny showing estimated evolutionary relationships at the *cyt b* gene of both *Lampetra* spp. and *Entosphenus* spp. in California as well as reference samples from *L. pacifica* and *L. richardsoni* (collected outside California). Bootstrap proportions are displayed for nodes with >50% support, and nodes with <50% support are collapsed to polytomies. Publicly available sequences are marked with asterisks.



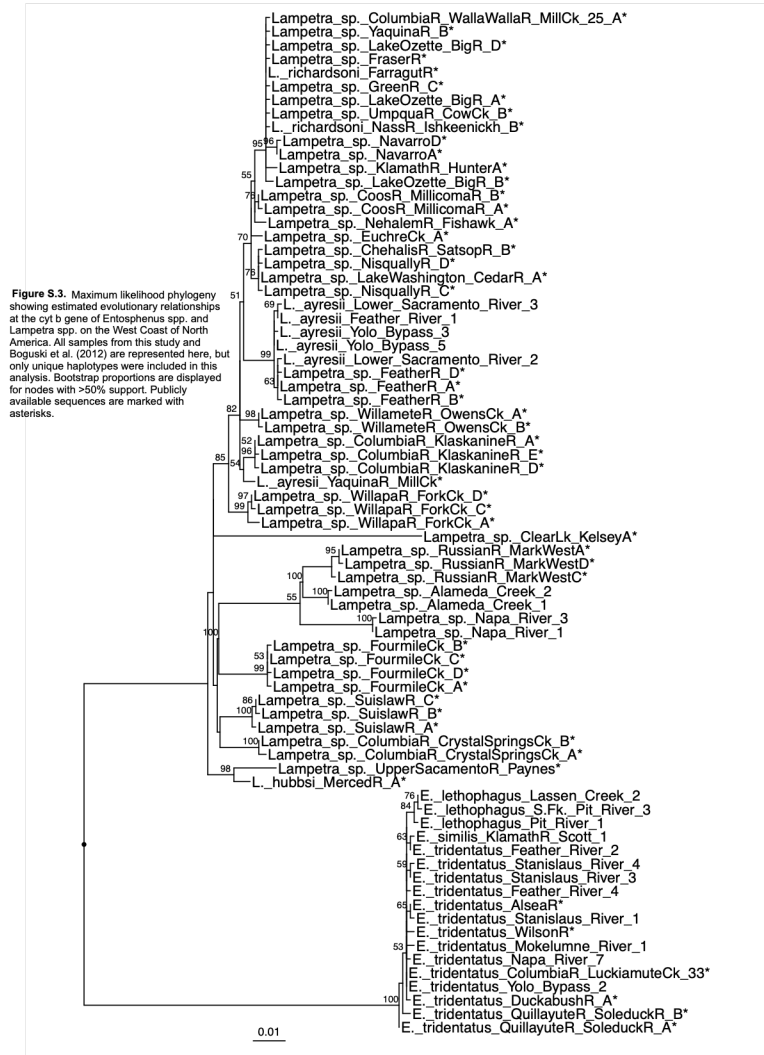


Table S.1. Sample collection information, including the names of agencies/entities that contributed lamprey tissue samples to this study and their respective Scientific Collecting Permit numbers.

Location	Sample Collecting Entity	Scientific Collecting Permit Number
Alameda Creek	San Francisco Public Utilities Commission	4808
American River	Pacific States Marine Fisheries Commission	13640
Delta	California Department of Water Resources	10842
Feather River	California Department of Water Resources	9611
Kings River	California Department of Fish and Wildlife	N/A
Klamath River	California Department of Fish and Wildlife	N/A
Lassen Creek	California Department of Fish and Wildlife	N/A
Mokelumne River	East Bay Municipal Utility District	2990
Napa River	Napa County Resource Conservation District	3495
Pit River	California Department of Fish and Wildlife	N/A
Sacramento River	California Department of Fish and Wildlife	N/A
Stanislaus River [Oakdale]	FISHBIO	2147
Stanislaus River [Caswell]	Pacific States Marine Fisheries Commission	13640
Tuolumne River	FISHBIO	2147
Yolo Bypass	California Department of Water Resources	10842

Table S.2. Sample information for all samples sequenced for this study, including HUC6 basin, GPS coordinates, field species identifications and genetic species identifications.

Genus	Species	Sample Name	Site Name	Basin (HUC6)	Collection Location		Length			Life Stage	Field ID
					Latitude	Longitude	Fork length (mm)	Tail length (mm)			
Lampetra	opreisi	Feather River, 1	Eye Side Channel	Lower Sacramento	39.45659	-121.61539	107			larval	unknown lamprey
	opreisi	Sacramento-San Joaquin, Delta, 1	Rio Vista	Sacramento-San Joaquin Delta	38.1547193	-121.6814495				31 larval	unknown lamprey
	opreisi	Sacramento-San Joaquin, Delta, 2	Shernan Lake	Sacramento-San Joaquin Delta	38.0581151	-121.8193469				141 larval	unknown lamprey
	opreisi	Lower Sacramento River, 1	Knight's Landing	Lower Sacramento	38.790841	-121.6921	122			juvenile	Lampetra ayeiell
	opreisi	Lower Sacramento River, 2	Knight's Landing	Lower Sacramento	38.790841	-121.6921	115			larval	Lampetra ayeiell
	opreisi	Lower Sacramento River, 3	Knight's Landing	Lower Sacramento	38.790841	-121.6921	100			larval	unknown lamprey
Lampetra	opreisi	Yolo Bypass, 3	Yolo Bypass	Lower Sacramento	38.353383	-121.643181	158			juvenile	unknown lamprey
	opreisi	Yolo Bypass, 4	Yolo Bypass	Lower Sacramento	38.353383	-121.643181	150			larval	unknown lamprey
	opreisi	Yolo Bypass, 5	Yolo Bypass	Lower Sacramento	38.353383	-121.643181	141			larval	unknown lamprey
	hubbsi	Kings River, 1	Kings River	Tulare-Buena Vista Lakes	36.821073	-119.354416	143			larval	unknown lamprey
Lampetra	hubbsi	Kings River, 2	Kings River	Tulare-Buena Vista Lakes	36.821073	-119.354416	145			larval	unknown lamprey
	hubbsi	Kings River, 3	Kings River	Tulare-Buena Vista Lakes	36.821073	-119.354416	121			larval	unknown lamprey
	hubbsi	Kings River, 4	Kings River	Tulare-Buena Vista Lakes	36.821073	-119.354416	121			larval	unknown lamprey
	hubbsi	Kings River, 5	Kings River	Tulare-Buena Vista Lakes	36.821073	-119.354416	96			larval	unknown lamprey
Lampetra	sp.	Alameda Creek, 1	Alameda Creek	San Francisco Bay	37.538924	-121.856242	82			larval	Lampetra sp.
	sp.	Alameda Creek, 2	Alameda Creek	San Francisco Bay	37.538924	-121.856242	135			adult	Lampetra sp.
	sp.	Alameda Creek, 3	Alameda Creek	San Francisco Bay	37.538924	-121.856242	102			adult	Lampetra sp.
	sp.	Alameda Creek, 4	Alameda Creek	San Francisco Bay	37.538924	-121.856242	87			larval	Lampetra sp.
	sp.	Alameda Creek, 5	Alameda Creek	San Francisco Bay	37.538924	-121.856242	114			adult	Lampetra sp.
	sp.	Alameda Creek, 6	Alameda Creek	San Francisco Bay	37.538924	-121.856242	112			adult	Lampetra sp.
	sp.	Alameda Creek, 7	Alameda Creek	San Francisco Bay	37.538924	-121.856242	100			adult	Lampetra sp.
	sp.	Alameda Creek, 8	Alameda Creek	San Francisco Bay	37.538924	-121.856242	122			adult	Lampetra sp.
	sp.	Alameda Creek, 9	Alameda Creek	San Francisco Bay	37.538924	-121.856242	107			adult	Lampetra sp.
	sp.	Alameda Creek, 10	Alameda Creek	San Francisco Bay	37.538924	-121.856242	104			larval	Entrophis indeterminate
Lampetra	sp.	Napa River, 1	Napa River	San Francisco Bay	38.343737	-122.287949	164			larval	Lampetra sp.
	sp.	Napa River, 2	Napa River	San Francisco Bay	38.343737	-122.287949	140			larval	Lampetra sp.
	sp.	Napa River, 3	Napa River	San Francisco Bay	38.343737	-122.287949	111			larval	Lampetra sp.
Reference samples shared from out-of-state collections -- only included in supplementary analyses shared by Jon Hess, CITECJ											
Lampetra	richardsoni	L. richardsoni, Columbia River, 1	Columbia River	Upper Columbia	45.74764	-121.5711	N/A			N/A	N/A
Lampetra	richardsoni	L. richardsoni, Columbia River, 2	Columbia River	Upper Columbia	45.74764	-121.5711	N/A			N/A	N/A

Table S.2 Sample information for all samples sequenced for this study, including HUC6 basin, GPS coordinates, field species identifications and genetic species identifications.

Genetic ID		Collection Location				Length		Life Stage		Field ID	
Genus	Species	Sample Name	Site Name	Basin (HUC6)	Latitude	Longitude	Fork length (mm)	Total length (mm)			
Entosphenus	similis	Klamath_R_Scott_1	Scott	Klamath	41.72653	-123.008364	N/A	N/A		Entosphenus tridentatus	
	similis	Klamath_River_Shasta_1	Shasta	Klamath	41.82955	-122.593161	N/A	N/A		Entosphenus tridentatus	
	similis	Klamath_River_Shasta_2	Shasta	Klamath	41.82955	-122.593161	N/A	N/A		Entosphenus similis	
Entosphenus	similis	Klamath_River_Shasta_3	Shasta	Klamath	41.82955	-122.593161	N/A	N/A		Lampropetia richardsoni	
	lethophagus	Lassen_Creek_2	Lassen Creek	Upper Sacramento	41.8259722	-120.2951036		105		larval	unknown lam prey
	lethophagus	Lassen_Creek_3	Lassen Creek	Upper Sacramento	41.8259722	-120.2951036		90		larval	unknown lam prey
	lethophagus	Pit_River_1	Pit River	Upper Sacramento	40.967425	-121.8384472		148		larval	Entosphenus lethophagus
	lethophagus	Pit_River_2	Pit River	Upper Sacramento	40.967425	-121.8384472		166		larval	Entosphenus lethophagus
	lethophagus	S_Fk_Pit_River_1	South Fork Pit River	Upper Sacramento	41.23433	-120.3407		112		larval	unknown lam prey
	lethophagus	S_Fk_Pit_River_2	South Fork Pit River	Upper Sacramento	41.23433	-120.3407		140		larval	unknown lam prey
	lethophagus	S_Fk_Pit_River_3	South Fork Pit River	Upper Sacramento	41.23433	-120.3407		121		larval	unknown lam prey
	lethophagus	S_Fk_Pit_River_4	South Fork Pit River	Upper Sacramento	41.23433	-120.3407		12		larval	unknown lam prey
	lethophagus	S_Fk_Pit_River_5	South Fork Pit River	Upper Sacramento	41.23433	-120.3407		125		larval	unknown lam prey
Entosphenus	tridentatus	Alameda_Creek_1	Alameda Creek	San Francisco Bay	37.538924	-121.856242		77		larval	Entosphenus tridentatus
	tridentatus	Alameda_Creek_2	Alameda Creek	San Francisco Bay	37.538924	-121.856242		77		larval	Entosphenus tridentatus
	tridentatus	Alameda_Creek_3	Alameda Creek	San Francisco Bay	37.538924	-121.856242		460		adult	Entosphenus tridentatus
	tridentatus	Alameda_Creek_4	Alameda Creek	San Francisco Bay	37.538924	-121.856242		107		larval	Entosphenus tridentatus
	tridentatus	Stanislaus_River_1	Stanislaus [Caswell]	San Joaquin	37.7020694	-121.1792389		133		larval	Entosphenus tridentatus
	tridentatus	Stanislaus_River_2	Stanislaus [Caswell]	San Joaquin	37.7020694	-121.1792389		112		larval	Entosphenus tridentatus
	tridentatus	Stanislaus_River_3	Stanislaus [Caswell]	San Joaquin	37.7020694	-121.1792389		130		larval	unknown lam prey
	tridentatus	Feather_River_1	Eye Side Channel	Lower Sacramento	39.45659	-121.61539		83		larval	Lampropetia ayresii
	tridentatus	Feather_River_2	Eye Side Channel	Lower Sacramento	39.45659	-121.61539		133		juvenile	Entosphenus tridentatus
	tridentatus	Feather_River_3	Eye Side Channel	Lower Sacramento	39.45659	-121.61539		145		juvenile	Entosphenus tridentatus
	tridentatus	Feather_River_4	Feather	Lower Sacramento	N/A	N/A	N/A			juvenile	Entosphenus tridentatus
	tridentatus	Feather_River_5	Feather	Lower Sacramento	N/A	N/A	N/A			larval	Entosphenus tridentatus
	tridentatus	Feather_River_6	Feather	Lower Sacramento	N/A	N/A	N/A			larval	unknown lam prey
	tridentatus	Feather_River_7	Herringer	Lower Sacramento	39.3159	-121.62462	N/A			larval	Entosphenus tridentatus
tridentatus	Feather_River_8	Herringer	Lower Sacramento	39.3159	-121.62462		92		larval	unknown lam prey	
tridentatus	Feather_River_9	Herringer	Lower Sacramento	39.3159	-121.62462		115		juvenile	Entosphenus tridentatus	

Table S.2. (continued) Sample information for all samples sequenced for this study, including HUC6 basin, GPS coordinates, field species identifications and genetic species identifications.

Genetic ID		Collection location				Length		Life Stage	Field ID
Genus	Species	Sample Name	Site Name	Basin (HUC6)	Latitude	Longitude	Fork length (mm)		
<i>Entosphenus</i>	<i>tridentatus</i>	Lower Sacramento River, 1	Knight's Landing	Lower Sacramento	38.790841	-121.6921	119		Juvenile
	<i>tridentatus</i>	Lower Sacramento River, 2	Knight's Landing	Lower Sacramento	38.790841	-121.6921	104		Juvenile
	<i>tridentatus</i>	Lower Sacramento River, 3	Knight's Landing	Lower Sacramento	38.790841	-121.6921	105		Juvenile
	<i>tridentatus</i>	Klamath River, Shasta, 1	Scott's	Klamath	41.726355	-123.008364	N/A		N/A
	<i>tridentatus</i>	Klamath River, Shasta, 1	Scott's	Klamath	41.726355	-123.008364	N/A		N/A
	<i>tridentatus</i>	Lower American River, 1	Lower American River	Lower Sacramento	38.567033	-121.385156	350		Juvenile
	<i>tridentatus</i>	Lower American River, 2	Lower American River	Lower Sacramento	38.567033	-121.385156	104		Juvenile
	<i>tridentatus</i>	Lower American River, 3	Lower American River	Lower Sacramento	38.567033	-121.385156	110		Juvenile
	<i>tridentatus</i>	Modokume River, 1	Modokume	San Joaquin	38.17678	-121.153095	147		Juvenile
	<i>tridentatus</i>	Modokume River, 2	Modokume	San Joaquin	38.17678	-121.153095	129		Juvenile
	<i>tridentatus</i>	Modokume River, 3	Modokume	San Joaquin	38.17678	-121.153095	119		Juvenile
	<i>tridentatus</i>	Napa River, 1	Napa River	San Francisco Bay	38.343737	-122.287949	79		Juvenile
	<i>tridentatus</i>	Napa River, 2	Napa River	Napa	38.343737	-122.287949	119		Juvenile
	<i>tridentatus</i>	Napa River, 3	Napa River	Napa	38.343737	-122.287949	91		Juvenile
	<i>tridentatus</i>	Napa River, 4	Napa River	Napa	38.343737	-122.287949	76		Juvenile
	<i>tridentatus</i>	Napa River, 5	Napa River	Napa	38.343737	-122.287949	84		Juvenile
	<i>tridentatus</i>	Napa River, 6	Napa River	Napa	38.343737	-122.287949	86		Juvenile
	<i>tridentatus</i>	Napa River, 7	Napa River	Napa	38.343737	-122.287949	105		Juvenile
	<i>tridentatus</i>	Napa River, 8	Napa River	Napa	38.343737	-122.287949	106		Juvenile
	<i>tridentatus</i>	Napa River, 9	Napa River	Napa	38.343737	-122.287949	81		Juvenile
	<i>tridentatus</i>	Napa River, 10	Napa River	Napa	38.343737	-122.287949	102		Juvenile
	<i>tridentatus</i>	Stanislaus River, 4	Oakdale	San Joaquin	37.770997	-120.867744	110		Juvenile
	<i>tridentatus</i>	Stanislaus River, 5	Oakdale	San Joaquin	37.770997	-120.867744	130		Juvenile
	<i>tridentatus</i>	Stanislaus River, 6	Oakdale	San Joaquin	37.770997	-120.867744	132		Juvenile
	<i>tridentatus</i>	Tuolumne River, 1	Tuolumne	San Joaquin	37.602248	-120.786582	129		Juvenile
	<i>tridentatus</i>	Tuolumne River, 2	Tuolumne	San Joaquin	37.602248	-120.786582	115		Juvenile
	<i>tridentatus</i>	Tuolumne River, 3	Tuolumne	San Joaquin	37.602248	-120.786582	120		Juvenile
	<i>tridentatus</i>	Volio Bypass, 1	Volio Bypass	Lower Sacramento	38.353383	-121.643181	148		Juvenile
	<i>tridentatus</i>	Volio Bypass, 2	Volio Bypass	Lower Sacramento	38.353383	-121.643181	128		Juvenile
	<i>tridentatus</i>	Volio Bypass, 3	Volio Bypass	Lower Sacramento	38.353383	-121.643181	113		Juvenile
<i>tridentatus</i>	Volio Bypass, 4	Volio Bypass	Lower Sacramento	38.353383	-121.643181	135		Juvenile	
Reference samples shared from out-of-state collections -- only included in supplementary analyses shared by Lori Hess, CMT/CJ									
<i>Entosphenus</i>	<i>tridentatus</i>	E., tridentatus, Columbia River	Lower Columbia	45.3524	-122.618	N/A	N/A	N/A	N/A
	<i>tridentatus</i>	E., tridentatus, Columbia River	Lower Columbia	45.3524	-122.618	N/A	N/A	N/A	N/A
	<i>tridentatus</i>	E., tridentatus, Strait of Geo Strait of Georgia	British Columbia	49.329	-123.624	N/A	N/A	N/A	N/A
	<i>tridentatus</i>	E., tridentatus, Strait of Geo Strait of Georgia	British Columbia	49.329	-123.624	N/A	N/A	N/A	N/A

Table S.3. Sample information for publicly available sequences from Boguski et al. (2012) that were included in analyses.

Species ID		Sample Name	GenBank Accession Number	Publication
Genus	Species			
<i>Entosphenus</i>	<i>tridentatus</i>	EelR*	GU120848.1	Boguski et al. 2012
<i>Entosphenus</i>	<i>tridentatus</i>	SanPabloBay_Sonoma*	GU120866.1	Boguski et al. 2012
<i>Lampetra</i>	<i>ayresii</i>	SFBayA*	GU120867.1	Boguski et al. 2012
<i>Lampetra</i>	<i>ayresii</i>	SFBayB*	GU120868.1	Boguski et al. 2012
<i>Lampetra</i>	<i>hubbsi</i>	MercedR_A*	GU120869.1	Boguski et al. 2012
<i>Lampetra</i>	<i>hubbsi</i>	MercedR_B*	GU120870.1	Boguski et al. 2012
<i>Lampetra</i>	<i>pacifica</i>	ColumbiaR_CrystalSpringsC*	GU120797.1	Boguski et al. 2012
<i>Lampetra</i>	<i>pacifica</i>	ColumbiaR_ClackamasA*	GU120799.1	Boguski et al. 2012
<i>Lampetra</i>	<i>richardsoni</i>	FraserR_SmithA*	GU120736.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_HunterA*	GU120831.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_HunterB*	GU120832.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_HunterC*	GU120833.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_HunterD*	GU120834.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_HunterE*	GU120835.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_McGarveyA*	GU120836.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_McGarveyB*	GU120837.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_McGarveyC*	GU120838.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_McGarveyD*	GU120839.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	KlamathR_McGarveyE*	GU120840.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RedwoodCk_PrairieA*	GU120841.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RedwoodCk_PrairieB*	GU120842.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RedwoodCk_BoyesA*	GU120843.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RedwoodCk_BoyesB*	GU120844.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RedwoodCk_StreelovA*	GU120845.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RedwoodCk_StreelovB*	GU120846.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RedwoodCk_StreelovC*	GU120847.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	UpperSacramentoR_Paynes*	GU120849.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	NavarroA*_	GU120850.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	NavarroB*_	GU120851.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	NavarroC*_	GU120852.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	NavarroD*_	GU120853.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	FeatherR_A*	GU120854.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	FeatherR_B*	GU120855.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	FeatherR_C*	GU120856.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	FeatherR_D*	GU120857.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	ClearLk_KelseyA*	GU120858.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	ClearLk_KelseyB*	GU120859.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	ClearLk_KelseyC*	GU120860.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	ClearLk_KelseyD*	GU120861.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RussianR_MarkWestA*	GU120862.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RussianR_MarkWestB*	GU120863.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RussianR_MarkWestC*	GU120864.1	Boguski et al. 2012
<i>Lampetra</i>	sp.	RussianR_MarkWestD*	GU120865.1	Boguski et al. 2012

Table S.4. Genetic species identification (BLAST) results for all samples sequenced in this study. The top three sequence matches for each sequence are reported here with collection locations, accession numbers, percent sequence identity, and E-value scores.

[illegible]

Chapter 2: Phylogenomics and Population Genomics of Lampreys (*Occidentis* spp.) in California's Central Valley

Abstract:

Biodiversity loss disproportionately impacts freshwater ecosystems, and lampreys are an important, yet often overlooked component of freshwater biodiversity. Lamprey (*Entosphenus* spp. and *Occidentis* spp.) populations in California are understudied, but evidence suggests many populations are in decline. Knowledge gaps, including taxonomic uncertainty, inhibit conservation action for lampreys in California. Though genomic tools have been utilized to study *Entosphenus tridentatus*, genome-wide patterns have not yet been characterized in *Occidentis*. To characterize the genomic diversity of *Occidentis* in California, we generated reduced-representation genomic sequence data for individuals in the *Occidentis* genus collected from the San Francisco Bay Delta and the Sacramento and San-Joaquin River basins (N=131). Phylogenetic analyses at genome-wide SNPs confirmed the same deep phylogenetic divergences within *Occidentis* originally observed using single-gene mitochondrial DNA analyses. Phylogenetic results and estimates of genomic differentiation support the designation of the two lineages from the Napa River and Alameda Creek as novel species. Population genomic analyses showed that the migratory (anadromous) species, *O. ayresii*, is genetically homogenous across sampling locations in the San Francisco Bay Delta and Sacramento River basin, supporting the hypothesized lack of strict natal homing in this species. Each of the other species (*O. hubbsi*, Napa *Occidentis* sp., and Alameda *Occidentis* sp.) were each only found at one location, strongly supporting that these three species are non-migratory and each have a small, geographically restricted range. Estimates of nucleotide diversity for *O. hubbsi*, Napa *Occidentis* sp., and Alameda *Occidentis* sp. were similar to estimates for *O. ayresii*, which indicates these non-migratory lineages have maintained genetic diversity despite being relatively small and isolated. The taxonomic clarification and identification of novel species

provided by this study are fundamental for future research, monitoring, and conservation action to protect lampreys in California and across the West Coast.

Introduction:

Freshwater ecosystems provide habitat to disproportionately high levels of biodiversity (Tickner et al. 2020), and biodiversity loss in freshwater habitats is occurring up to three times faster than in terrestrial and marine habitats (Birnie-Gauvin et al. 2023). These losses are accelerated in systems that are heavily exploited by humans (Woodward et al. 2010). Existing legislative frameworks to protect freshwater diversity are species-specific (i.e. Endangered Species Act) and require clearly defined units of biodiversity (i.e. species, subspecies, distinct population segments). Therefore, the current system for conservation is not effectively leveraged for lineages with taxonomic uncertainty, like lampreys on the West Coast of North America.

Once-abundant lamprey populations in freshwater systems across the West Coast of North America have dwindled over the last 50-75 years (Close et al. 2002, Wicks-Arshack et al. 2018). Lampreys in California inhabit the southernmost portion of their range on the West Coast, and populations at the edge of their natural species ranges are often at a higher risk of local extirpation or extinction (Wilson et al. 2005). Northward range contraction has already been documented for Pacific Lamprey (*Entosphenus tridentatus*; Reid and Goodman 2016), which highlights the urgent need to protect lampreys in California's freshwater systems, especially in the face of accelerating climate change, a major driver of range shifts. This need for lamprey conservation in California is increasingly recognized, but conservation action is inhibited, especially within the *Occidentis* genus (formerly *Lampetra*; Carim et al. 2024), by the lack of clearly defined taxonomic units.

Recent genetic studies on lamprey diversity across the globe have shown that taxonomic classifications based on morphology are often inaccurate, both over- and under-estimating species diversity (Spice et al. 2018, Docker and Potter 2019, Roguemont et al. 2021, Auringer

et al. 2023, Carim et al. 2024). Much of the taxonomic uncertainty within each genus is related to the life history variation in lampreys. Adult lampreys exhibit life history variation on two major axes: migratory type and feeding type. In many lamprey genera, there is one anadromous, parasitic species that gives rise to one or more closely related non-migratory, non-parasitic lineages. Original species descriptions based on morphology classified these non-migratory, non-parasitic lampreys as separate species given significant differences in adult body size, pigmentation, and oral disc dentition between these two life history types (Renaud 2011; Potter et al. 2015). Genetic data revealed that this classification is accurate in some cases (Docker and Potter 2019, Auringer et al. 2023), but in others, these migratory and non-migratory lineages maintain gene flow and should be classified as life history variants within the same species (i.e. European River Lamprey and European Brook Lamprey; Roguemont et al. 2021). Understanding the patterns of speciation in *Occidentis* will not only improve the accuracy of taxonomic classifications, but will also provide insight into the evolutionary processes that generate and maintain biodiversity.

Occidentis ayresii, an anadromous, parasitic species, is widely distributed in coastal watersheds across the West Coast from Alaska to the Sacramento-San Joaquin Delta (Moyle 2002). Until recently, the sympatric, non-migratory counterpart to *O. ayresii* was classified as a separate species, *O. richardsoni*. However, genetic studies comparing *O. ayresii* and *O. richardsoni* from the northern latitudes of the West Coast found no species-level differentiation and recommended classifying these life history variants as part of the same species (Docker et al. 2016, Carim et al. 2023). Interestingly, patterns of speciation in *Occidentis* seem to differ in the northern and southern portion of the *O. ayresii* range. The non-migratory counterpart (i.e. *O. richardsoni*) maintains gene flow with *O. ayresii* in the north, while further south, *O. ayresii* gives rise to several genetically divergent lineages (Auringer et al. 2023).

Recent studies have begun to close knowledge gaps in *Occidentis* taxonomy through mitochondrial DNA (mtDNA) phylogenetic and species delimitation analyses (Carim et al. 2023,

Auringer et al. 2023, Carim et al. 2024) and genetic monitoring using markers that distinguish between life history variants within *Occidentis ayresii* (Silver et al. 2025). Mitochondrial DNA is widely used for estimating phylogenetic relationships because of its advantages over nuclear DNA. Due to the higher mutation rate in mtDNA, a single mitochondrial gene is more informative of evolutionary relationships between closely related taxa than a nuclear gene of the same length (Rubinoff and Holland 2005). Additionally, mtDNA is maternally inherited and does not usually recombine, which allows new mutations to reach fixation more quickly in mtDNA than nuclear DNA. These properties of mtDNA facilitate direct tracing of maternal lineages. However, analyses based on mtDNA have several important limitations that can obscure phylogenetic relationships (Ballard and Whitlock 2004; Waters et al. 2010; Lukic et al. 2021). Because mtDNA only reflects the maternal lineage, hybridization is not accurately reflected in mtDNA analyses, and in cases of sex-biased dispersal, mtDNA markers are likely to overestimate divergences (Rubinoff and Holland 2005). Other potential issues include selection on mitochondrial genes, lack of resolution in deep divergences due to substitution saturation, and nuclear mitochondrial DNA segments (NUMTs) (Ballard and Whitlock 2004). In general, studies that rely on one or few genes (mitochondrial or nuclear) only capture a very small portion of the variation found across the entire genome and can misrepresent patterns of diversity such that a phylogenetic tree based on one gene does not match the overall species tree (Degnan and Rosenberg 2009).

Conducting phylogenetic and population genetic analyses on multiple markers across the nuclear genome can help overcome limitations of mtDNA, and no previous research has investigated *Occidentis* in California using genome-wide nuclear data. Determining accurate evolutionary relationships and taxonomic classifications within *Occidentis* is critical for future research because it is required to leverage legislative frameworks (i.e. Endangered Species Act) in order to guarantee protections for these native species. To fill this fundamental need, this study uses genome-wide SNP (single nucleotide polymorphism) data obtained from RADseq

(restriction-site associated DNA sequencing) to investigate inter- and intra-specific diversity, clarify evolutionary relationships, and contextualize the observed patterns of speciation within the *Occidentis* genus in California.

Methods:

Study area: California's freshwater systems are extensively interrupted by dams, and natural flows are diverted via major water diversions for agriculture and municipal demands. Canals and aqueducts often export significant proportions of a river's water, and in extreme cases, such as in the San Joaquin River, as much as 80% of total freshwater is extracted, causing the mainstem to frequently run dry for long stretches, up to 240 km (Lacán and Resh 2016). This study investigates *Occidentis* across the highly modified Sacramento and San Joaquin River basins and the San Francisco Bay Delta (Figure 2.1).

The Sacramento River, Napa River, and Alameda Creek all drain into the San Francisco Bay. The Sacramento River is the largest river in California, and the second largest on the West Coast after the Columbia River. The Sacramento River drains over 26,000 square miles in northern California, provides 35% of municipal water for all of California and irrigates over 2 million acres of agricultural land (Lacán and Resh 2016). Samples were collected from the Yolo Bypass and the mainstem of the Lower Sacramento River near Knights Landing. The American and Feather Rivers are tributaries to the Sacramento River. Tissue samples were collected from the American River east of Sacramento near Watt Ave and from the Feather River downstream of Oroville.

The Napa River (North Bay) drains approximately 425 square miles in Napa Valley, most well-known for its vineyards. As the wine industry exploded in Napa in the 1970s, the demand for water from the Napa River also rapidly increased. Napa River discharges through the Napa Sonoma marsh into San Pablo Bay, part of the broader San Francisco Bay-Delta system. Tissue samples were collected north of the city of Napa.

Alameda Creek is the third-largest tributary that flows into the San Francisco Bay via a 12-mile concrete waterway. Geological features indicate that Alameda Creek is a relatively ancient antecedent stream (1 MYA; Sloan 2006). Alameda Creek drains approximately 650 square miles of East Bay interior hills and valleys. Three million San Francisco Bay Area residents get their drinking water from Alameda Creek, and additional water is diverted for grazing and vineyards. Tissue samples were collected near the Sunol Regional Wilderness below Calaveras Reservoir.

The Kings River historically drained into the Tulare Lake basin, but this basin experienced the most extreme water diversions for agricultural production. In the Tulare Lake basin, all water is extracted via canals and aqueducts; there is no longer a natural surface water pathway out of the watershed (Brown & Moyle 2005). The Kings River is the largest river draining the Sierra Nevada mountain range and was historically isolated from the Sacramento-San Joaquin but canal construction in the 1870s piped water from the Kings River to the San Francisco Bay via the San Joaquin River. Kings River tissue samples were collected downstream of Pine Flat Dam.

Sample collection: Tissue samples were collected opportunistically by ongoing salmonid monitoring programs across the state of California in 2018 and 2019. Field sampling methods varied, but most samples came from rotary screw traps. The larval life stage was targeted because this would allow us to infer approximate natal origin. Lampreys do not exhibit strict natal homing behavior, so collections of adults from one stream would likely represent individuals from many different natal origins, complicating the investigation of genetic structure using geography of sampling locations. Though larval individuals were the primary focus, some adult specimens were included in the study to serve as reference individuals for adult morphological species identification. Most samples collected for this study were in the larval phase and not identified to species in the field; many were not identified to the genus level; therefore, our collections included both *Entosphenus* and *Occidentis*.

Genetic Sample Identification: To validate the field identifications provided and to determine which of the collected samples belonged to the *Occidentis* genus, a subset of individuals that were included in the RADseq library were sequenced at the mitochondrial cytochrome b (cyt b) gene, and these sequences were compared to publicly available sequences from voucher specimens confirmed to be members of the *Occidentis* genus (Auringer et al. 2023). Because not all individuals within the RADseq data set were included in the cyt b sequencing, there were still many samples unidentified to the genus level. The individuals identified to species (or at least genus) using the cyt b sequences served as reference individuals to classify all samples in the RADseq dataset by species or at least genus.

Data collection:

Genomic DNA (gDNA) was extracted from all lamprey tissue samples (N=480) collected in 2018-2019 using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol. Samples from each location were divided across plates to reduce plate effects in downstream analyses. Samples were normalized to each contain 50 ng of gDNA and the BestRAD protocol (Ali et al. 2016) was used to prepare samples for sequencing. Genomic DNA was digested with the *SbfI* restriction enzyme, and adaptors with individual barcodes were ligated to the ends of digested gDNA. AMPure XP magnetic beads were used to wash away unligated adaptors and any fragments without adaptors. Individually barcoded samples were then pooled into one library and sonicated to obtain fragments within the target size range (200-500bp). Fragment size was confirmed with gel electrophoresis and size selection was performed using Dynabead M-280 streptavidin magnetic beads to wash away any non-target fragments. The remainder of library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina, following the manufacturer's protocol. The final library fragment size was confirmed with gel electrophoresis and DNA concentration was quantified with a Qubit (Thermo Fisher). Short-read high-throughput sequencing was performed on one lane on an Illumina HiSeq 4000x using paired-end 150 bp sequences. The sequencing was carried out by

the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

Data processing:

Raw reads were demultiplexed and aligned to the Pacific lamprey (*Entosphenus tridentatus*) reference genome (Etrf_v1; somatic genome; Hess et al. 2020) using the mem algorithm in bwa (Li & Durban 2009). Two genome assemblies are available for *E. tridentatus*, but the other version (Etrm_v1; Hess et al. 2020) was assembled from germline cells. In *E. tridentatus*, portions of the germline genome are eliminated from somatic cells in a process called programmed genome rearrangement (Timoshevskiy et al. 2017). Because gDNA for this study was extracted from somatic cells, data was aligned to the somatic genome assembly to avoid mapping errors.

Alignment, mapping quality, and coverage was calculated in SAMtools (Li et al. 2009). BAM files were sorted, indexed, and PCR duplicates were removed using SAMtools. Reads with low mapping quality (MAPQ < 20) were removed and individuals with poor coverage (<100,000 total mapped reads) were dropped from further analysis. Single nucleotide polymorphisms (SNPs) were called using the likelihood ratio test in ANGSD (Korneliussen et al. 2014) if they had a minimum minor allele frequency of at least 0.05, if the SNP was present in at least 50% of individuals, and if the SNP p-value <1.0E-12. Sites with poor raw sequence quality (sequence quality PHRED score <20) and poor mapping quality (map quality PHRED score <20) were excluded from further analyses.

Genotype likelihoods were used when possible to reduce biases related to genotyping errors in low-coverage, reduced-representation genomic sequencing data. Genotype likelihoods were calculated using the GATK model in ANGSD, and posterior probabilities were calculated for each site using allele frequencies as a prior. Genotype likelihoods were carried forward for analyses when posterior probabilities > 0.95.

When genotypes were required for analysis (rather than genotype likelihoods), an individual would only be assigned a genotype at a SNP if that individual's data included at least 8 sequencing reads (8x per-site, per-individual coverage) to reduce genotype error associated with low coverage data.

Data analyses:

We performed Principal Component Analyses (PCA) using the single-read sampling approach in ANGSD (-doIBS). This method randomly samples a single read at each SNP for each individual to estimate the covariance matrix. Principal components (PCs) were calculated from the covariance matrix and PCA plots were created in R (v. 4.5.0; R Core Team 2025).

Eigenvalues were calculated and scree plots were generated in R to determine the optimal number of principal components to retain. We utilized a hierarchical approach to investigate genomic clustering at multiple levels (i.e. genus, species, population). First, we conducted PCA on: a) all individuals (both genera), b) all individuals in the *Occidentis* genus, and c) all individuals within each putative species.

Maximum likelihood phylogenetic analysis was run in RAxML (Stamatakis 2014) using SNP genotype data (coverage >8x). Specifically, the analysis used the GTRGAMMA model of nucleotide substitution with an ascertainment bias correction (ASC_GTRGAMMA) using the Lewis correction method to account for the absence of invariant sites in SNP input data. Node support was assessed using rapid bootstrapping (1000 replicates). All *Occidentis* individuals (N=131) were included in phylogenetic analysis, and one *E. tridentatus* individual was included as an outgroup. After exporting SNP genotypes from ANGSD in PLINK format and converting from PLINK to VCF, SNPs were filtered by missingness (maximum 25% missing data across all individuals) in VCFtools. Then, SNPs in high linkage disequilibrium ($r^2 > 0.25$; calculated in windows containing 1000 SNPs) were removed using BCFtools. The resultant VCF file was re-headered with sample names and converted to PHYLIP format, then input into RAxML. The

resulting maximum likelihood consensus tree was midpoint rooted and visualized using the R package ggtree (Yu et al. 2017).

Population genomic summary statistics (π , F_{ST} , and Edwards genetic distance) were calculated in snpR (Hemstrom and Jones 2023) using SNP genotype data. In these analyses, SNPs were filtered by coverage (excluded if per-site, per-individual coverage $>8x$), missingness (excluded if missing data >0.25), and SNPs were pruned to one SNP per 50 kb. For each of these three population-level metrics (π , F_{ST} , and Edwards genetic distance), weighted averages were calculated to reduce bias introduced by varying levels of missing data between individuals. Individuals (N=10) collected from multiple sites in the Delta were combined into one population due to their close geographic proximity and limited sample sizes at each specific sample location. Few individuals were collected from the Lower American River (N=1) and Feather River (N=2) sampling locations, so we did not estimate nucleotide diversity within these sites. These sites were included in F_{ST} and genetic distance analyses despite their small sample sizes because they provide valuable insight into the connectivity between mainstem and tributary populations.

Results:

Illumina sequencing generated 115,696 Megabases of raw data, and 88% of all bases had a PHRED quality score >30 . After alignment, the average number of mapped reads per individual was 729,355 reads. Within *Occidentis*, this average was 733,847, compared to 725,590 in *Entosphenus*. The average per-individual coverage of RAD loci was 6x. Twenty-four individuals (5% of total samples) were dropped from further analysis due to poor sequencing coverage ($<100,000$ total mapped reads).

Principal Component Analysis (PCA) including all samples (*Entosphenus* and *Occidentis*; N=456) and all SNPs that passed minimum quality filters (177,046 SNPs) showed that the largest axis of variation (PC1; 59.2%) split the samples into two groups that corresponded to the

two genera (*Entosphenus*: N = 325; *Occidentis*: N = 131; Figure 2.S1). Within *Occidentis* (N=131 individuals; 113,025 SNPs), PC1 and PC2 split samples into three main clusters: 1) *O. ayresii*, 2) *O. hubbsi*, and 3) Napa *Occidentis* sp. and Alameda *Occidentis* sp. (Figure 2.2a). Subsequent principal components (PC2 and PC3) split samples similarly, but split samples from the Napa River and Alameda Creek into distinct clusters (Figure 2.S2). The scree plot for PCA within *Occidentis* (Figure 2.S3) supports retaining the first three principal components. Within *O. ayresii* (N = 78 individuals; 82,592 SNPs), no clear clustering pattern is observed with PCA (Figure 2.S4).

Phylogenetic analysis included all *Occidentis* samples (N = 131) and SNPs that passed coverage and missingness filters (N = 975). One *E. tridentatus* individual was included as an outgroup, but all other *Entosphenus* individuals were excluded. Results support patterns observed in hierarchical PCA analyses. Four reciprocally monophyletic lineages are observed: 1) *O. ayresii*, 2) *O. hubbsi*, 3) Alameda *Occidentis* sp., and 4) Napa *Occidentis* sp. (Figure 2.2c). Bootstrap support for each of these nodes is $\geq 95\%$, and bootstrap support was below 75% for any other nodes. Within *O. ayresii*, all samples across sampling locations form one large polytomy.

Population genomic analysis including data across 7,841 SNPs showed that within-lineage nucleotide diversity (π) was similar across all four main lineages, but was highest in *O. ayresii* ($\pi = 0.28$), lower in Alameda *Occidentis* sp. ($\pi = 0.26$), and Napa *Occidentis* sp. ($\pi = 0.25$), and lowest in *O. hubbsi* ($\pi = 0.23$) (Table 2.1). Estimates of pairwise F_{ST} and Edwards genetic distance showed high levels of divergence between all four lineages and very little genomic divergence within *O. ayresii* across sampling sites (Table 2.2).

Discussion:

We found that *Occidentis* is characterized by relatively deep divergences between the migratory *O. ayresii* and all other lineages, corroborating the pattern observed in mtDNA analyses

(Boguski et al. 2012; Auringer et al. 2023; Carim et al. 2023). *Occidentis ayresii* were primarily found in estuarine habitat and lower reaches of freshwater systems. No genomic structure by geographic location was observed within *O. ayresii*. The other highly divergent lineages (*O. hubbsi*, Napa *Occidentis* sp., and Alameda *Occidentis* sp.) were each found at a single location where *O. ayresii* was absent, suggesting speciation was driven by isolation from *O. ayresii* and the resultant lineages are most likely non-migratory species with geographically restricted ranges. Genome-wide analyses support the designation of Napa *Occidentis* sp., and Alameda *Occidentis* sp. as distinct species. The geographic ranges of these two candidate species have not been delineated, but they are likely endemic to the north (Napa) and south (Alameda) regions of the San Francisco Bay Delta. Endemic species are generally at higher risk of extinction than widely distributed species, especially in the face of climate change (Manes et al. 2021). High levels of endemism are observed across native fishes in California, and these endemic lampreys in *Occidentis* further highlight the need for conservation action to protect lampreys and all of California's endemic freshwater fishes.

Evolutionary relationships observed in this study broadly support previous phylogenetic work investigating *Occidentis*. Several of the lineages observed in this study were also reported in past work (Boguski et al. 2012; Auringer et al. 2023; Carim et al. 2024), and this study shows relatively high levels of genomic differentiation between them. Notably, the distinct *Occidentis* lineages originally sampled by Boguski et al. (2012) from Clear Lake, the Russian River, and Paynes Creek are absent in this study due to inability to obtain additional samples from these locations. Future targeted sampling in these areas will allow for a more complete understanding of the species diversity of *Occidentis*.

Due to the opportunistic nature of sample collection, this study was not designed to systematically assess species distribution; however, we can draw some reasonable inferences about how *O. ayresii* utilizes freshwater habitat in the Central Valley. Though this study focuses on *Occidentis*, *Entosphenus tridentatus* is the most widely studied species on the West Coast

and serves as a useful reference point when discussing *O. ayresii*. The described species ranges of *E. tridentatus* and *O. ayresii* overlap extensively, both within California and across the West Coast of N. America. *Entosphenus tridentatus* has a larger body size and is considered a more proficient swimmer than *O. ayresii*, so within-river occupancy likely differs between these two anadromous lampreys. For this study, both *E. tridentatus* and *O. ayresii* were collected in tandem, which allows us to clarify where species ranges overlap at a finer scale.

O. ayresii was observed less frequently overall (N=78; 17% of total samples) than *E. tridentatus* (N=290; 64% of total samples), and *O. ayresii* was found at fewer sampled locations (6 sites; 38% of total sites) than *E. tridentatus* (11 sites; 69% of total sites). Within the Sacramento River, the farthest upstream tributary sampled for this study was the Feather River, where both *O. ayresii* and *E. tridentatus* were present. However, only three *O. ayresii* individuals were collected from Sacramento River tributaries, and the majority (N=75; 96%) of all *O. ayresii* samples collected were concentrated in the Delta and mainstem Sacramento River, suggesting that *O. ayresii* heavily uses estuarine habitat and the lower reaches of the mainstem, but utilizes the upper reaches of the system to a lesser extent. These results support previous findings that *O. ayresii* tends to stay closer to the coastline than *E. tridentatus*, both in freshwater systems (i.e. largely inhabiting estuaries and lower reaches of coastal rivers) and during the marine phase (i.e. most likely to occupy shallow waters close to shore), while *E. tridentatus* was primarily found deeper in the water column and found both at nearshore and offshore marine sampling sites (Weitkamp et al. 2023).

Although both *E. tridentatus* and *O. ayresii* were both present across the Sacramento River basin, a different pattern is observed in the San Joaquin River basin. Here, *E. tridentatus* was found in the Mokelumne, Stanislaus, and Tuolumne Rivers (Auringer et al. 2023), but *O. ayresii* was absent from all of these tributaries to the San Joaquin River. *O. ayresii* was collected from Sherman Lake, a lake at the confluence of the Sacramento and San Joaquin Rivers, and from the very lower reaches of the San Joaquin River within the Delta, but was not found any further

upstream. This suggests that *O. ayresii* might not enter the San Joaquin River system, while *E. tridentatus* utilizes tributaries in the San Joaquin basin relatively extensively (126 individuals collected across three tributaries; Auringer et al. 2023). Considering the patterns of *O. ayresii* habitat use in the Sacramento River (occupancy primarily in the Delta and lower reaches of the mainstem), it is possible that *O. ayresii* is utilizing the mainstem San Joaquin, but samples for this study were only collected in the tributaries and very lower reaches of the San Joaquin within the Delta. Future collections in the San Joaquin basin should target the lower reaches of the mainstem San Joaquin River to determine whether *O. ayresii* is utilizing this habitat.

Within *O. ayresii*, no population structure was observed between locations. However, the geographic scope of this study is limited to the southernmost region of the species range, which extends northward to the Alaskan coast. In this study, *O. ayresii* was found at eight sampling locations that are all geographically proximate (all within the Sacramento-San Joaquin Delta and the lower reaches of the Sacramento River). Previous mtDNA work supported the presence of two distinct phylogenetic clades within *O. ayresii* in California – one spanning the northern coast of California and the other found in the San Francisco Bay Delta and California's Central Valley (Auringer et al. 2023). Notably, when analyzed with mtDNA sequences from range-wide samples, the support for two phylogenetic clades decreases (Carim et al. 2024). *Occidentis ayresii* samples from the northern coast were not available for this study, so our results are unable to refute or support the existence of these two phylogenetic clades with genome-wide data. Further research using genome-wide data with samples across the entire geographic range will clarify patterns of population structure within *O. ayresii*.

While *O. ayresii* appears to be restricted to the Delta and Sacramento River basin, other species in the *Occidentis* genus are present outside this region. *Occidentis hubbsi* is found in the upper San Joaquin River, and was found in this study in the Kings River, the only upper San Joaquin tributary sampled. Additionally, two candidate species occur in the Napa River and Alameda Creek, respectively. The presence of *Occidentis* species in Alameda Creek, the Napa

River and the Kings River suggests that *O. ayresii* previously utilized more upstream sections of California's freshwater systems, but populations in the upper reaches of the system became isolated from the broader *O. ayresii* populations.

Auringer et al. (2023) proposed that these distinct lineages in Alameda Creek and Napa River warrant species descriptions based on mtDNA molecular species delimitation analyses.

Additional molecular species delimitation work including *Occidentis* sequences from a broader geographic range across the West Coast still consistently delimited Alameda Creek and Napa populations as distinct species (Carim et al. 2024). In this study, phylogenetic analysis of genome-wide data shows that these lineages are reciprocally monophyletic, and the F_{ST} and genetic distance estimates (Table 2.2) demonstrate that these two candidate species show similar levels of divergence as those observed between *O. hubbsi* (a known species) and all other lineages. These findings increase support for the designation of Napa *Occidentis* sp. and Alameda *Occidentis* sp. as distinct species. Additional work, specifically morphological descriptions of adult individuals representing these populations, is needed to formally describe these two lineages as species. More collections from neighboring streams in the region will be important to delineate species ranges of Napa *Occidentis* sp. and Alameda *Occidentis* sp.

Evolutionary forces seem to be impacting *O. ayresii* differently in its northern range and southern range, and this is most clearly demonstrated by the genetic relationship between life history variants. At southern latitudes, the freshwater-resident forms tend to become isolated and speciate, while at northern latitudes, gene flow is maintained between life history variants. In the northern portion of the *O. ayresii* species range, the migratory, parasitic form and freshwater-resident, non-parasitic form (previously classified as separate species) were found to be very genetically similar and were recently re-classified as one species (Boguski et al. 2012, Carim et al. 2023, Carim et al. 2024). The freshwater-resident form within *O. ayresii*, now referred to by the common name Western Brook Lamprey, is found frequently at these northern latitudes.

In contrast, at more southern latitudes, Western Brook Lamprey are uncommon. In Oregon, three genetically divergent freshwater-resident species are observed: *O. pacifica*, and two candidate species from Suislaw River and Fourmile Creek (Reid et al. 2011, Boguski et al. 2012, Carim et al. 2023). Further south, in California, this pattern becomes more extreme, with six genetically divergent freshwater-resident species: *O. hubbsi* and five candidate species from Napa River, Alameda Creek, Paynes Creek, Clear Lake, and the Russian River (Boguski et al. 2012, Auringer et al. 2023, Carim et al. 2024). In this study, all freshwater-resident individuals presumed to be Western Brook Lamprey (based on morphology) were found to be part of lineages that are highly genetically divergent from *O. ayresii*.

The variation in the number of species-level divergences from *O. ayresii* at different latitudes suggests that each non-migratory population sits somewhere along the speciation continuum and the evolutionary trajectory of each is highly dependent on the specific system and how gene flow with the migratory lineage is impacted. Glaciation was much more extensive at northern latitudes during the Last Glacial Maximum (Smith et al. 2010), which may also explain the higher species diversity at more southern latitudes. Additionally, the high species diversity within *Occidentis* in California may be explained by the hydrology and topography of California, which is associated with high levels of endemism in freshwater fish (Moyle 2002, Ball et al. 2013). California has many endemism hotspots, generally characterized by geographical, topographical, or hydrological features that can explain the tendency of divergent lineages to accumulate. However, neither the Napa River nor Alameda Creek are considered highly isolated systems that would accumulate deeply divergent lineages. Both Alameda Creek and the Napa River are in the same hydrologic region (103 regions used by California Department of Water Resources; Figure 2.1) and same zoogeographic subprovince (Moyle 2002). Genetic differentiation between the northern and southern regions of the San Francisco Bay has been observed in other fishes at the population level (rainbow trout; Leitwein et al. 2017) and subspecies level (sculpin; Moyle and Campbell 2022). However, the deep, species-level

divergences observed between Napa *Occidentis* sp. and Alameda *Occidentis* sp. reflect a unique evolutionary process.

Interestingly, *E. tridentatus* was found in both the Napa River and Alameda Creek and was genetically homogenous across these sites and throughout the Central Valley, but *O. ayresii* was not found in the Napa River or Alameda Creek. These results highlight that the evolutionary forces that drove speciation in *Occidentis* in this region did not impact *Entosphenus* in the same way. One possible explanation is that limited migration and dispersal in *Occidentis* (compared to *Entosphenus*) increases the likelihood of isolation and speciation in *Occidentis*. *Occidentis ayresii* has a smaller body size and is a weaker swimmer than *E. tridentatus*, so the migratory form in *Entosphenus* may simply be more capable of accessing habitat and therefore is more likely to maintain some level of gene flow with freshwater-resident forms even amidst periods of isolation. Additionally, the larger population size of *E. tridentatus* compared with *O. ayresii* increases the likelihood of *E. tridentatus* re-connecting with a temporarily isolated population. Further research is needed to understand the differences in diversification and speciation between *Occidentis* and *Entosphenus*. Studies investigating the migratory behavior of *O. ayresii* will help answer these evolutionary questions as well as provide insight into migration timing for *O. ayresii* in California, which is currently unknown. Future genomic work expanding the geographic scope beyond California to the entire West Coast will allow for a more comprehensive investigation of within-species diversity across the entire *O. ayresii* range. Sequencing and assembling the *O. ayresii* genome is another important next step for lamprey research. Additionally, as sequencing costs decrease, whole-genome sequencing will be feasible for future studies, which would eliminate the limitations of reduced-representation sequencing (i.e. non-uniform and incomplete genome coverage) and allow for investigation of structural variation that likely plays an important role in lamprey diversity.

Conclusion:

California holds a high proportion of the known diversity within *Occidentis*. Though little is known about these lampreys, we do know they all rely heavily on the freshwater systems in the San Francisco Bay-Delta and the Central Valley, which means these lampreys are experiencing the same threats as all freshwater fishes in California's ultra-modified waterways. Thus, *Occidentis* populations may be declining and are likely vulnerable to the effects of climate change. Action is needed to protect the diverse lampreys of California, but fundamental knowledge gaps obstruct conservation initiatives focused on lampreys. This study provides phylogenomic resolution for several taxonomically ambiguous lineages and improves our understanding of how *Occidentis* species are utilizing freshwater habitat to facilitate informed conservation management decisions about lampreys in California.

References:

- Al-Samarai, F. R., & Al-Kazaz, A. A. (2015). Molecular markers: an introduction and applications. *European journal of molecular biotechnology*, 9(3), 118-130.
- Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, 202(2), 389-400.
- Auringer, G., Campbell, M. A., Goertler, P. A., & Finger, A. J. (2023). Lampreys in California (Lampetra spp. and Entosphenus spp.): Mitochondrial phylogenetic analysis reveals previously unrecognized lamprey diversity. *North American Journal of Fisheries Management*, 43(6), 1511-1530.
- Ball, J. E., L. A. Bêche, P. K. Mendez, and V. H. Resh. 2013. Biodiversity in Mediterranean-climate streams of California. *Hydrobiologia*, 719: 187-213.
- Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular ecology*, 13(4), 729-744.
- Birnie-Gauvin, K., Lynch, A. J., Franklin, P. A., Reid, A. J., Landsman, S. J., Tickner, D., ... & Cooke, S. J. (2023). The RACE for freshwater biodiversity: essential actions to create the social context for meaningful conservation. *Conservation Science and Practice*, 5(4), e12911.
- Boguski, D. A., S. B. Reid, D. H. Goodman, and M. F. Docker. 2012. Genetic diversity, endemism and phylogeny of lampreys within the genus *Lampetra* sensu stricto (*Petromyzontiformes: Petromyzontidae*) in western North America. *Journal of Fish Biology*, 81(6): 1891-1914.
- Brown, L. R., & Moyle, P. B. (2005). Native fishes of the Sacramento-San Joaquin Drainage, California: a history of decline. In American Fisheries Society Symposium (Vol. 45, pp. 75-98).

- Carim, K. J., D. C. Larson, J. M. Helstab, M. K. Young, and M. F. Docker. 2023. A revised taxonomy and estimate of species diversity for western North American *Lampetra*. *Environmental Biology of Fishes*, 1-20.
- Carim, K. J., Auringer, G., Docker, M. F., Renaud, C. B., Clemens, B. J., Blanchard, M. R., ... & Young, M. K. (2024). Species diversity in the new lamprey genus *Occidentis*, formerly classified as western North American 'Lampetra'. *PLoS One*, 19(12), e0313911.
- Clemens, B. J., and C. J. Wang. 2021. Dispelling misperceptions of native lampreys (*Entosphenus* and *Lampetra* spp.) in the Pacific northwest (USA). *Conservation Science and Practice*, 3(6), e402.
- Close, D. A., M. S. Fitzpatrick, and H. W. Li. 2002. The ecological and cultural importance of a species at risk of extinction, Pacific lamprey. *Fisheries*, 27(7): 19-25.
- Docker, M. F., Silver, G. S., Jolley, J. C., & Spice, E. K. (2016). Simple genetic assay distinguishes lamprey genera *Entosphenus* and *Lampetra*: Comparison with existing genetic and morphological identification methods. *North American Journal of Fisheries Management*, 36(4), 780–787. <https://doi.org/10.1080/02755947.2016.1167146>
- Docker, M. F., & Potter, I. C. (2019). Life history evolution in lampreys: alternative migratory and feeding types. In *Lampreys: biology, conservation and control: Volume 2* (pp. 287-409). Dordrecht: Springer Netherlands.
- Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in ecology & evolution*, 24(6): 332-340.
- Fumagalli, M., Vieira, F. G., Linderroth, T., & Nielsen, R. (2014). ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, 30(10), 1486-1487
- Hemstrom, W., & Jones, M. (2023). snpR: User friendly population genomics for SNP data sets with categorical metadata. *Molecular Ecology Resources*, 23(4), 962-973.

- Hughes, L. C., Bloom, D. D., Piller, K. R., Lang, N., & Mayden, R. L. (2025). Phylogenomic resolution of lampreys reveals the recent evolution of an ancient vertebrate lineage. *Proceedings of the Royal Society B: Biological Sciences*, 292(2038).
- Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, 15(1), 356.
- Laćan, I., & Resh, V. H. (2016). A case study in integrated management: Sacramento–San Joaquin Rivers and Delta of California, USA. *Ecohydrology & Hydrobiology*, 16(4), 215-228.
- Lang, N. J., Roe, K. J., Renaud, C. B., Gill, H. S., Potter, I. C., Freyhof, J., ... & Mayden, R. L. (2009). Novel relationships among lampreys (Petromyzontiformes) revealed by a taxonomically comprehensive molecular data set. In *American Fisheries Society Symposium* (Vol. 72, pp. 41-55). American Fisheries Society.
- Leitwein, M., Garza, J. C., & Pearse, D. E. (2017). Ancestry and adaptive evolution of anadromous, resident, and adfluvial rainbow trout (*Oncorhynchus mykiss*) in the San Francisco bay area: application of adaptive genomic variation to conservation in a highly impacted landscape. *Evolutionary Applications*, 10(1), 56-67.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *bioinformatics*, 25(16), 2078-2079.
- Lucas, M. C., J. B. Hume, P. R. Almeida, K. Aronsuu, E. Habit, S. Silva, C. J. Wang, and B. Zampatti. (2021). Emerging conservation initiatives for lampreys: Research challenges and opportunities. *Journal of Great Lakes Research*, 47: S690-S703.

- Lukic, D., J. Eberle, J. Thormann, C. Holzschuh, and D. Ahrens. 2021. Excluding spatial sampling bias does not eliminate oversplitting in DNA-based species delimitation analyses. *Ecology and Evolution*, 11(15): 10327-10337.
- Manes, S., Costello, M. J., Beckett, H., Debnath, A., Devenish-Nelson, E., Grey, K. A., ... & Vale, M. M. (2021). Endemism increases species' climate change risk in areas of global biodiversity importance. *Biological conservation*, 257, 109070.
- Morin, P. A., Luikart, G., Wayne, R. K., & SNP Workshop Group. (2004). SNPs in ecology, evolution and conservation. *Trends in ecology & evolution*, 19(4), 208-216.
- Moyle, P. B. (2002). Inland fishes of California: revised and expanded. Univ of California Press, Oakland, California.
- Moyle, P. B., & Campbell, M. A. (2022). Cryptic species of freshwater sculpin (Cottidae: Cottus) in California, USA. *Zootaxa*, 5154(5), 501-527.
- Page, L. M., H. Espinosa-Pérez, L. T. Findley, C. R. Gilbert, R. N. Lea, N. E. Mandrak, and R. L. Mayden. 2013. New seventh edition of common and scientific names of fishes: changes include capitalization of common names. *Fisheries*, 38(4): 188-189.
- Parker, C. M. 2022. Spatial and Temporal Patterns of Detection of Native Lampreys in Central Valley California: A Starting Point. Master's thesis. Oregon State University, Corvallis, Oregon.
- Potter, I. C., Gill, H. S., Renaud, C. B., & Haoucher, D. (2015). The taxonomy, phylogeny, and distribution of lampreys. In M. F. Docker (Ed.), *Lampreys: Biology, conservation and control* (Fish & Fisheries 37, pp. 35–73). Springer. https://doi.org/10.1007/978-94-017-9306-3_2
- R Core Team (2025). *_R: A Language and Environment for Statistical Computing_*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

- Reid, S. B., Boguski, D. A., Goodman, D. H., & Docker, M. F. (2011). Validity of *Lampetra pacifica* (Petromyzontiformes: Petromyzontidae), a brook lamprey described from the lower Columbia River Basin. *Zootaxa*, 3091.
- Reid, S. B., & Goodman, D. H. (2016). Pacific lamprey in coastal drainages of California: occupancy patterns and contraction of the southern range. *Transactions of the American Fisheries Society*, 145(4), 703-711.
- Reid, S.B. and D.H. Goodman. 2021. Pacific Lamprey range and distribution in California. U.S. Fish and Wildlife Service.
- Renaud, C. B. (2011). *Lampreys of the world. An annotated and illustrated catalogue of lamprey species known to date.* (Species Catalogue for Fishery Purposes No. 5). Food and Agriculture Organization of the United Nations.
- Rougemont, Q., Dolo, V., Oger, A., Besnard, A. L., Huteau, D., Coutellec, M. A., ... & Evanno, G. (2021). Riverscape genetics in brook lamprey: genetic diversity is less influenced by river fragmentation than by gene flow with the anadromous ecotype. *Heredity*, 126(2), 235-250.
- Rubinoff, D., & Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic biology*, 54(6), 952-961.
- Silver, G. S., Lampman, R. T., Percival, N., Timoshevskaya, N., Smith, J. J., Bentley, K. T., ... & Hess, J. E. (2025). Genetic Identification of Lamprey Genera and Anadromous Ecotypes in Watersheds of the Northeastern Pacific Ocean. *Evolutionary Applications*, 18(5), e70108.
- Sloan, D. (2006). *Geology of the San Francisco Bay Region* (Vol. 79). Univ of California Press.
- Smith, G. R., Badgley, C., Eiting, T. P., & Larson, P. S. (2010). Species diversity gradients in relation to geological history in North American freshwater fishes. *Evolutionary Ecology Research*, 12(6), 693-726.

- Spice, E. K., Whitesel, T. A., Silver, G. S., & Docker, M. F. (2019). Contemporary and historical river connectivity influence population structure in western brook lamprey in the Columbia River Basin. *Conservation Genetics*, 20(2), 299-314.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
- Tickner, D., Opperman, J. J., Abell, R., Acreman, M., Arthington, A. H., Bunn, S. E., ... & Young, L. (2020). Bending the curve of global freshwater biodiversity loss: an emergency recovery plan. *BioScience*, 70(4), 330-342.
- Timoshevskiy, V. A., Lampman, R. T., Hess, J. E., Porter, L. L., & Smith, J. J. (2017). Deep ancestry of programmed genome rearrangement in lampreys. *Developmental biology*, 429(1), 31-34.
- Waters, J. M., D. L. Rowe, C. P. BurrIDGE, and G. P. Wallis. 2010. Gene trees versus species trees: reassessing life-history evolution in a freshwater fish radiation. *Systematic Biology*, 59(5): 504-517.
- Weitkamp, L. A., Tuttle, V., Ward, E. J., Kamikawa, D., Billings, A., Buchanan, J., & Hess, J. E. (2023). Pacific Lamprey and Western River Lamprey marine ecology: Insight from new ocean collections. *North American Journal of Fisheries Management*, 43(6), 1492-1510.
- Wicks-Arshack, A., Dunkle, M., Matsaw, S., & Caudill, C. (2018). An ecological, cultural and legal review of Pacific lamprey in the Columbia River Basin. *Idaho L. Rev.*, 54, 45
- Woodward, G., Perkins, D. M., & Brown, L. E. (2010). Climate change and freshwater ecosystems: impacts across multiple levels of organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1549), 2093-2106.
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in ecology and evolution*, 8(1), 28-36.

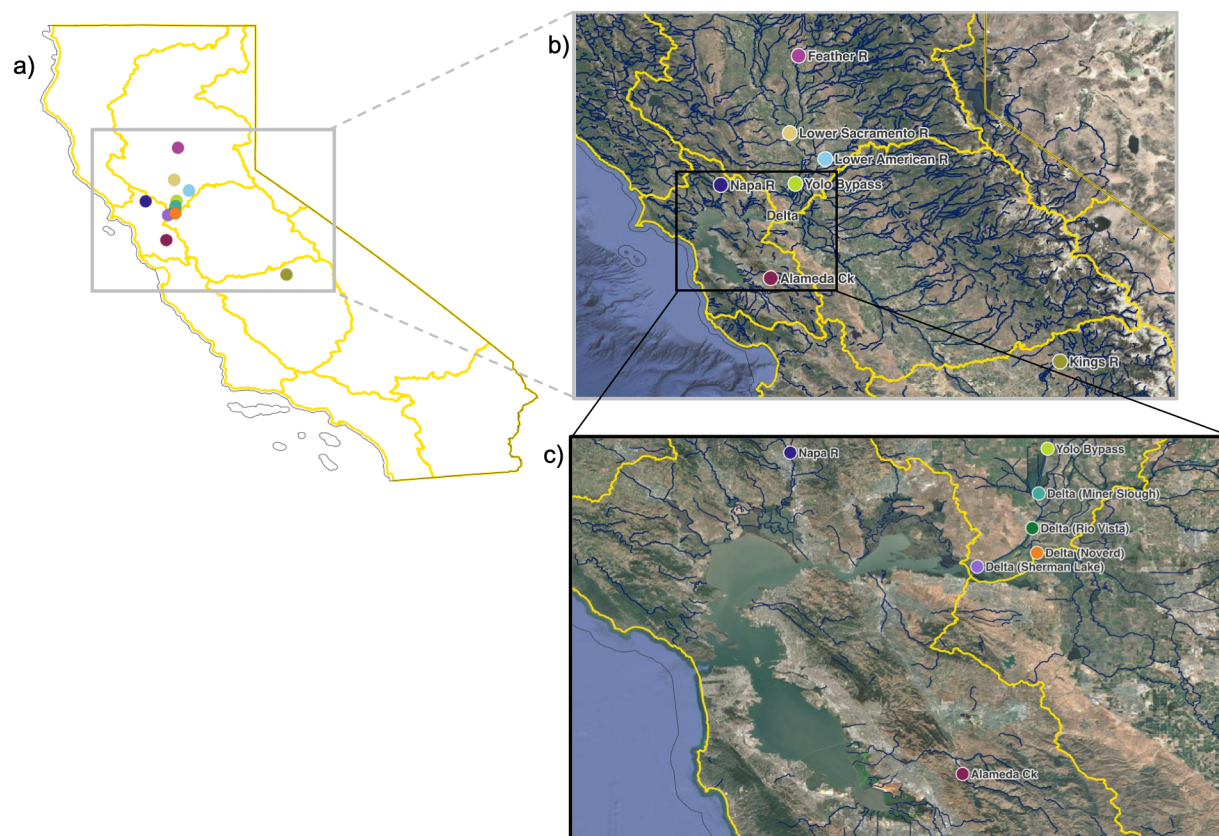


Figure 2.1. Map of sample locations at three scales: a) California state boundary, b) total study area, and c) San Francisco Bay-Delta. Yellow outlines represent the 10 major geographic areas that the California Department of Water Resources uses to manage water resources (i03 Hydrologic Regions; based on river basins and watershed boundaries).

Figure 2.2. Genomic variation and phylogenetic relationships of *Occidentis* individuals (N=131) across study area. Points colored by sampling location. (a) Principal component analysis of all *Occidentis* individuals. (b) Simplified map of sampling locations (c) Maximum likelihood phylogenetic tree using SNP genotypes (genotypes called if individual per-site coverage $\geq 8x$). Nodes with $<50\%$ bootstrap support are collapsed to polytomies, and bootstrap support values $>50\%$ displayed.

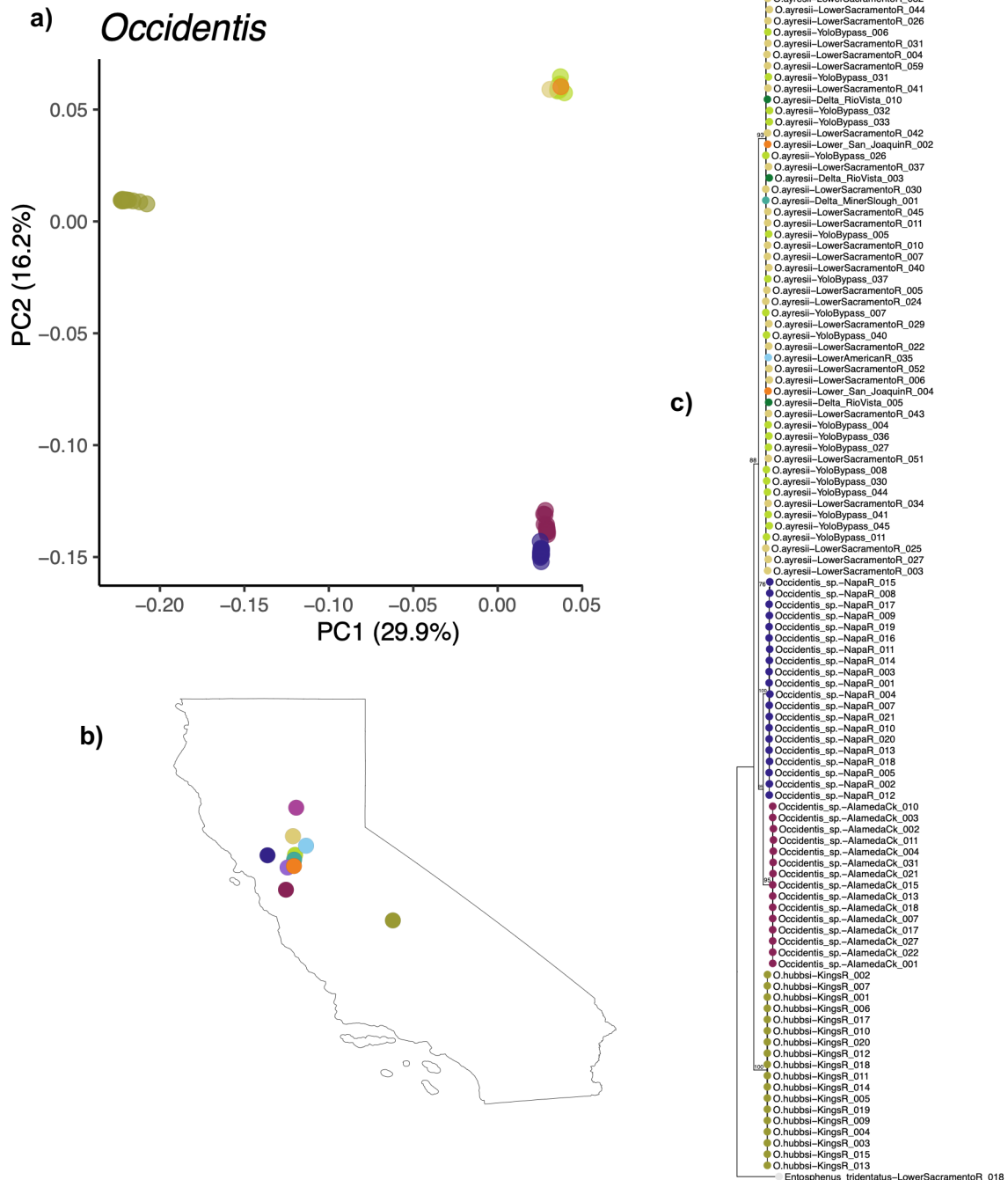


Table 2.1. Sample size and average nucleotide diversity within each sample site (four Delta sites combined here).

Sample Location	Sample Size	Nucleotide Diversity (π)
Delta	9	0.278
Feather River	2	
Lower Sacramento River	38	0.275
Lower American River	1	
Yolo Bypass	28	0.276
Napa River	20	0.251
Alameda Creek	15	0.257
Kings River	18	0.229

Table 2.2. Pairwise average F_{ST} values (above diagonal) and Edwards genetic distance values (below diagonal) comparing all sampling sites (four Delta sites combined here).

	PAIRWISE AVG F_{ST}								
	Delta	Feather River	Lower Sacramento River	Lower American River	Yolo Bypass	Napa River	Alameda Creek	Kings River	
Delta		0.006	2.042E-04	0.006	0.001	0.212	0.186	0.311	Low High
Feather River	0.169		1.571E-04	0.003	2.357E-04	0.224	0.195	0.332	
Lower Sacramento River	0.082	0.160		0.002	2.004E-04	0.203	0.184	0.299	
Lower American River	0.203	0.228	0.200		0.001	0.241	0.204	0.347	
Yolo Bypass	0.086	0.161	0.051	0.200		0.201	0.181	0.297	
Napa River	0.289	0.311	0.310	0.335	0.281		0.165	0.352	
Alameda Creek	0.285	0.310	0.276	0.328	0.277	0.243		0.334	
Kings River	0.369	0.388	0.364	0.400	0.364	0.386	0.386		
	GENETIC DISTANCE								

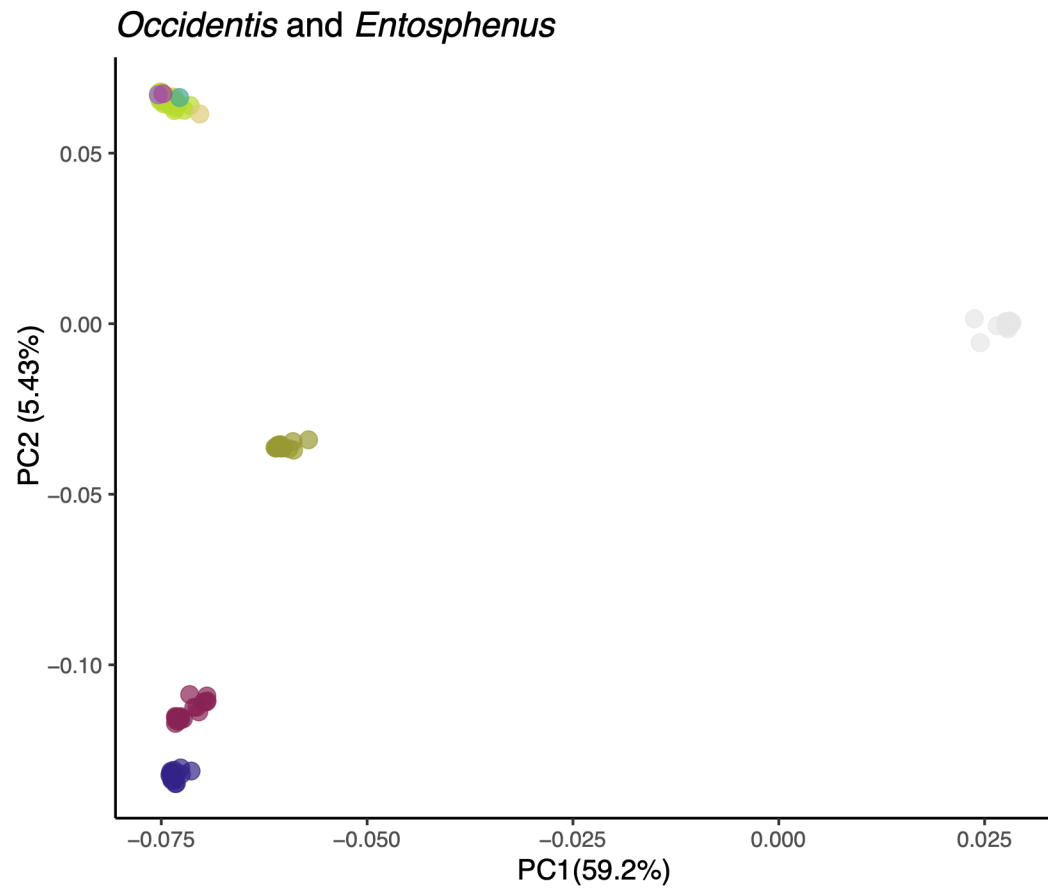


Figure 2.S1. Principal component analysis of all *Occidentis* and *Entosphenus* individuals. Points representing individuals in the *Occidentis* genus are colored by sampling location (refer to Figures 2.1 and 2.2). Grey points represent *Entosphenus* individuals.

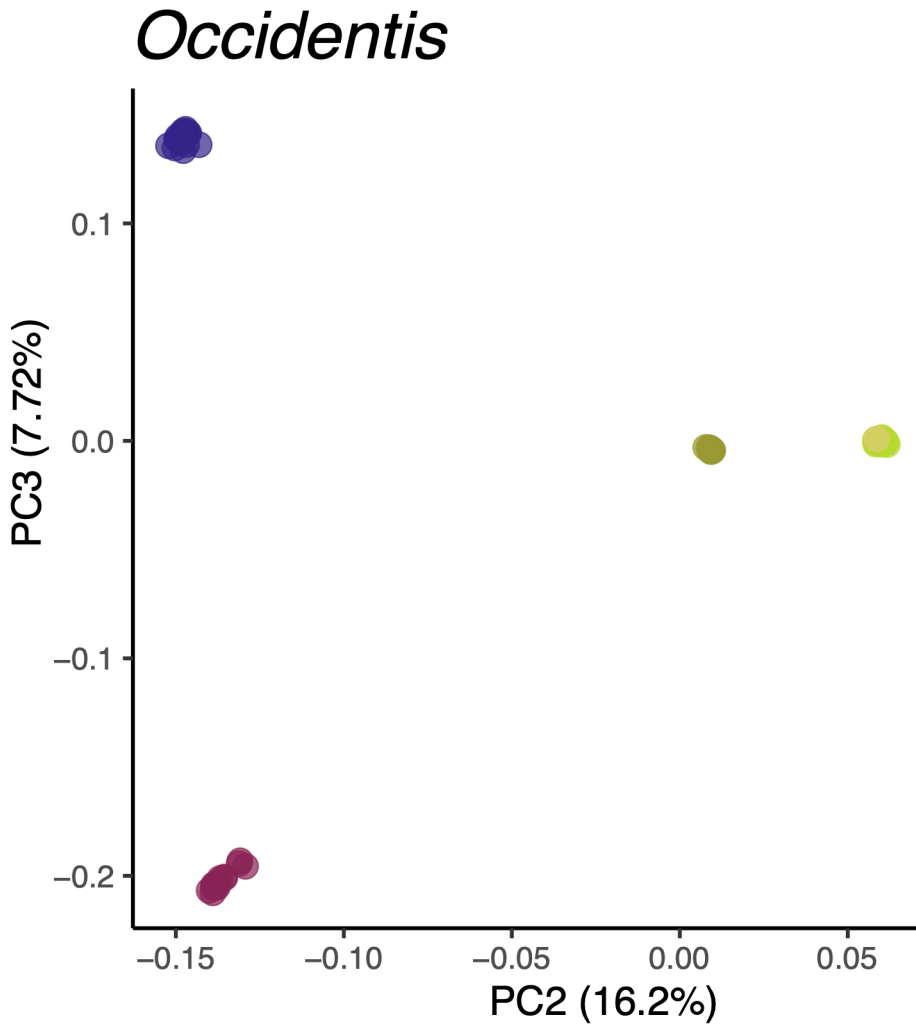


Figure 2.S2. Principal component analysis (PC2 and PC3) of all *Occidentis* individuals. Points are colored by sampling location (refer to Figures 1 and 2).

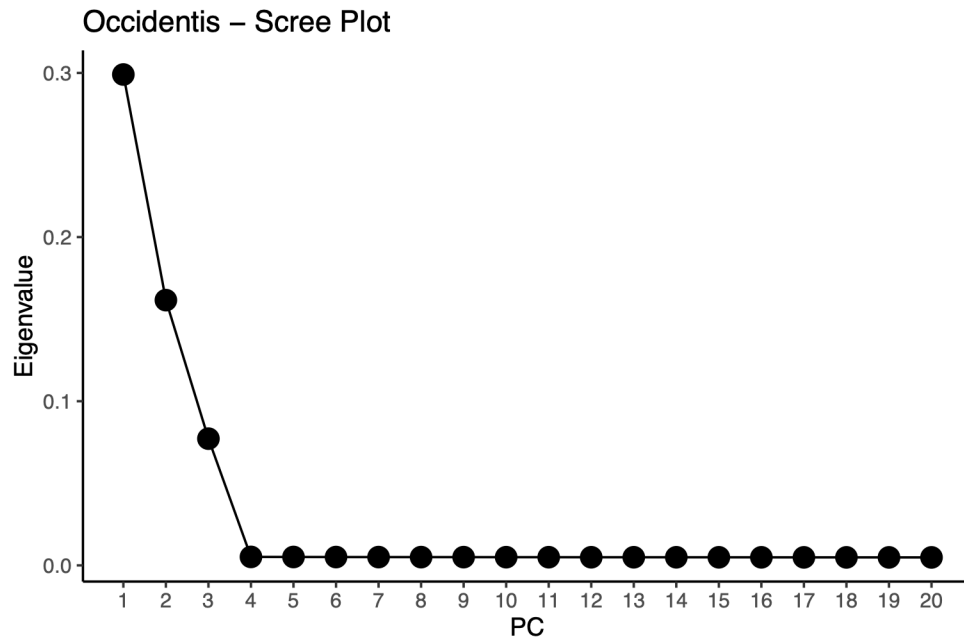


Figure 2.S3. Scree plot for principal component analysis of all *Occidentis* individuals.

Occidentis ayresii

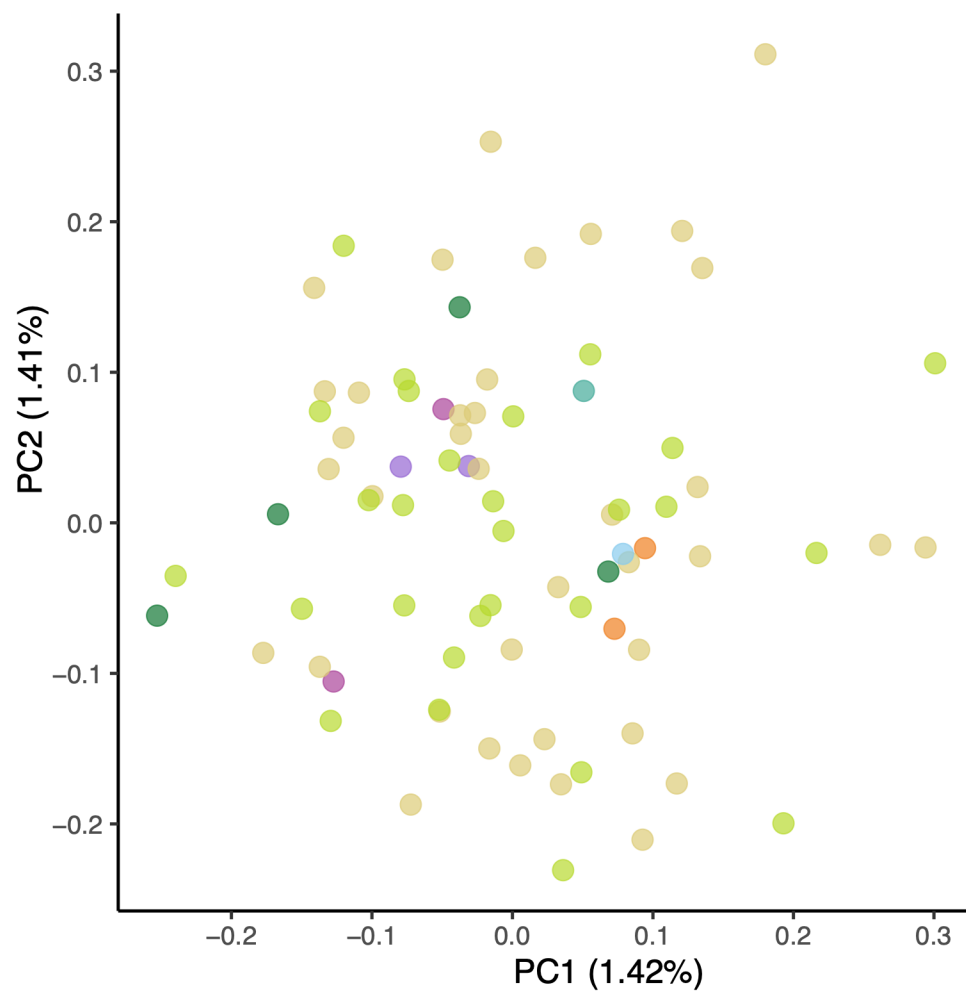


Figure 2.S4. Principal component analysis of all *O. ayresii* individuals. Points are colored by sampling location (refer to Figures 2.1 and 2.2).

Chapter 3: Genomic patterns of differentiation across *Entosphenus* in California

Abstract

Lampreys are an understudied lineage of native fishes in California. Two genera of lampreys (*Entosphenus* and *Occidentis*) are found in the state, but diversity within each genus is poorly understood. In California, three valid species and one unnamed species within *Entosphenus* are recognized based on morphology. *Entosphenus tridentatus* and *E. similis* are parasitic species and *E. lethophagus* is a non-parasitic species. The Goose Lake lamprey is an unnamed parasitic species of conservation concern. Single-locus mitochondrial DNA studies investigating species diversity have not supported these taxonomic classifications, but genome-wide patterns of *Entosphenus* across California remain unknown. This study uses genome-wide SNP data from *Entosphenus* individuals sampled across California to evaluate divergence between and diversity within species. Results support *E. lethophagus* and *E. similis* as distinct species from *E. tridentatus*, but we did not find evidence of genomic differentiation to warrant the designation of Goose Lake lamprey as a distinct species from *E. lethophagus*. Within *E. tridentatus*, no genome-wide differentiation was observed across sampling sites, supporting the findings of previous studies. However, divergence was observed between *E. tridentatus* individuals within a putative inversion region of chromosome 1 that is associated with adaptive variation. These findings clarify taxonomic uncertainty, inform management unit delineation, and highlight the importance of monitoring and conserving adaptive genomic variation across lampreys in California.

Introduction

Conservation of native lampreys on the West Coast of North America is inhibited by a lack of knowledge. The limited evidence we have about lampreys indicates that populations are declining due to habitat loss, land-use change, altered flow regimes, degraded water quality,

and climate change (Wang et al. 2020). These factors threaten many native fishes, but lamprey conservation efforts are at a unique disadvantage because historical misperceptions of lampreys have left gaps in our fundamental understanding of these species (Clemens and Wang 2021).

Native lampreys are an important part of biodiversity and provide multiple ecosystem benefits. However, since European colonization lampreys have typically been considered “pest” fish by non-Indigenous peoples. As the environmental conservation movement gained traction amongst European settlers, salmonids and other native fishes were studied, monitored, and endangered species were legally protected. Meanwhile, lampreys were ignored at best, and at worst, rivers and streams were treated with poison in lamprey eradication efforts, which occurred as recently as the 1970s (Petersen 2006).

Historical misperceptions of lampreys likely contributed directly to population decline but also resulted in a deficit of historical data and a gap in peer-reviewed literature evaluating threats to these fishes. The Endangered Species Act (ESA), the primary U.S. legislative framework for ecological conservation, cannot be leveraged for data-deficient taxa because it requires specific documentation of population decline or scientifically verified evidence of threats. Additionally, taxonomic uncertainty is prevalent throughout the lamprey lineage, and units of biodiversity (i.e. species, subspecies, distinct population segment) must be clearly defined to be eligible for listing under the ESA. In 2003, a petition was submitted to list four West Coast lamprey species as threatened or endangered under the ESA, but this petition was rejected due to the lack of information and undefined distinct population segments (USFWS 2004). This highlighted the urgent need for more research on lampreys.

Several lamprey characteristics present logistical challenges to conducting scientific studies. Survey methods for other fishes may not be suitable for lampreys because they typically spend the first 3-9 years of their life cycle burrowed in sediment at the bottom of the streambed (Docker and Potter 2019). During this larval phase, all species start out very small and remain

morphologically similar until the juvenile phase, so species identification of larvae in the field can be difficult (Maitland et al. 2015). Life history variation and distinguishing morphological features become evident in the adult phase, but lampreys can be elusive during this phase as well, as they generally stay close to the bottom of the streambed and are primarily active at night (Clemens et al. 2017).

In response to the rejected ESA petition, the Pacific Lamprey Conservation Initiative was formed, which led to the development of lamprey-specific research methodology, launched monitoring and management programs to advance lamprey conservation, and substantially increased our scientific understanding of West Coast lampreys (Wang and Schaller 2015).

Despite the increase in lamprey research, many knowledge gaps remain. Most studies have focused on one species, Pacific lamprey (*Entosphenus tridentatus*), a large (adults: 96–800 mm total length), anadromous, parasitic, highly dispersive species that is found in coastal drainages throughout the West Coast from California to Alaska and around the Pacific Rim as far as Japan (Renaud 2011). Several Indigenous communities in the Pacific Northwest rely on *E. tridentatus* for food sovereignty and cultural practices (Shoeships 2014; Wicks-Arshack et al. 2018). Sharp population declines were initially unnoticed, except by these Indigenous Peoples, who can no longer fish for this once-abundant species at many of their traditional harvest sites (Close et al. 2002). These Indigenous stewards were integral to the formation of the PLCI and continue to lead lamprey conservation efforts. Though research on *E. tridentatus* has increased significantly since the PLCI was established, much of this progress has been limited to the watersheds of Washington and Oregon.

Southern California is the southernmost boundary of the *E. tridentatus* range, and though range-edge populations are especially vulnerable to climate change, they may act as reservoirs of critical diversity required for a species to adapt to changing environments (Rehm et al. 2015). Therefore, studying range-edge populations is particularly important; however, most research on *E. tridentatus* takes place in the center of the species range. Furthermore, *E. tridentatus* in

California may be particularly at risk given that California's waterways are heavily impacted by anthropogenic modifications, and dams prevent *E. tridentatus* from accessing approximately 40% of their historical habitat (Reid and Goodman 2020). Improving our understanding of *E. tridentatus* in California is essential for range-wide conservation.

Two other valid *Entosphenus* species are found in California: *E. similis* (Klamath River Lamprey), *E. lethophagus* (Pit-Klamath brook lamprey), and one unnamed species, Goose Lake lamprey (*Entosphenus* sp.). *Entosphenus similis* is a small-bodied (136–269 mm total length), parasitic species with a limited range in the Klamath River basin (Renaud 2011). The migratory behavior of *E. similis* is not well-documented; this species is not anadromous but may migrate shorter distances within freshwater in the Klamath River basin. *Entosphenus lethophagus* is a non-parasitic, non-migratory species found in the Pit River, Goose Lake basin, and the upper Klamath River basin. Goose Lake lamprey (*Entosphenus* sp.), found exclusively in the Goose Lake basin, has not been formally described, but was recognized as distinct based on morphology (Hubbs 1925). Goose Lake lamprey is thought to contain at least some parasitic individuals with similar dentition patterns to *E. tridentatus*, but with much smaller body size and different pigmentation (Hubbs 1925, Moyle 2002).

Previous studies have not been able to distinguish between *Entosphenus* species using genetic markers (Docker et al. 2007), and phylogenetic work has also shown very low levels of divergence between species within *Entosphenus* in California (Auringer et al. 2023). However, previous research has been largely limited to one or few mitochondrial markers. A recent phylogenomic study supported at most four distinct clades within *Entosphenus* despite sampling six described *Entosphenus* species (Hughes et al. 2025). This study estimated that all *Entosphenus* species have diverged relatively recently (~1.6 mya; Hughes et al. 2025), which may explain the genetic similarity across *Entosphenus* species, especially in studies relying on highly conserved genes.

Some early genetic studies indicated that *E. tridentatus* was genetically homogenous, (Docker et al. 2007, Goodman et al. 2008), but others found evidence of genetic differentiation between specific regions and an overall pattern of isolation by distance (Lin et al. 2008, Spice et al. 2012). This initial research was limited to one or few genetic markers, and results seemed to vary depending on the loci used and the geographic scope. Despite these variations in genetic results, all studies concurred that *E. tridentatus* does not exhibit natal homing, which is unusual for species with long-distance migration. Hypotheses about drivers of genetic differentiation in *E. tridentatus* include limited dispersal during the marine phase (Spice et al. 2012) and ecotypic variation, possibly related to migration timing (Clemens et al. 2017).

Ecotypic variation in body size (Kostow 2002) and reproductive maturation (Clemens et al. 2013) has been identified in *E. tridentatus* and genetic loci have been associated with these ecotypes (Hess et al. 2014, Parker et al. 2019). Hess et al. (2020) produced the first genome assemblies of *E. tridentatus* and found very high levels of gene flow across the species range, but determined that ecotypic and other adaptive genetic variation is maintained in ‘genomic islands of divergence’. One such genomic region on chromosome 1 (chr1) is of particular interest in this study. Hess et al. (2020) found evidence that this region is likely a polymorphic chromosomal inversion in *E. tridentatus* and SNPs within this region are associated with variation in reproductive maturity. Though this is a putative chromosomal inversion, this region (chr1: position 9908655 – 15463359) will be referred to as “chr1 inversion” in this study.

Despite advances in genomic resources and research on *E. tridentatus*, genome-wide patterns of divergence across *Entosphenus* in California remain unknown. This study utilizes reduced-representation genomic sequence data to further investigate divergence patterns between *E. tridentatus*, *E. similis*, *E. lethophagus*, and Goose Lake lamprey to clarify taxonomic classifications and evaluate genomic diversity within each species, particularly *E. tridentatus*.

Methods

Study Area: This study investigates *Entosphenus* across the San Francisco Bay-Delta, Sacramento and San Joaquin river basins (the Central Valley), the Klamath River basin, and the Goose Lake basin. The Sacramento and San Joaquin rivers drain into the San Francisco Bay. The Sacramento River is the largest river (by discharge) in California, and the second largest on the West Coast of the United States of America after the Columbia River. The Sacramento River drains over 26,000 square miles in northern California, provides 35% of municipal water for all of California and irrigates over 2 million acres of agricultural land (Lacán and Resh 2016). The San Joaquin River is smaller and highly diverted; as much as 80% of total freshwater is extracted (Lacán and Resh 2016). The extreme demand for freshwater in the Central Valley of California and resultant dams and diversions have severely reduced and degraded fish habitat, particularly for anadromous species, such as *E. tridentatus*.

The Klamath River drains over 15,000 square miles in Oregon and northern California, and is considered an important region for biodiversity conservation (DellaSala et al. 2026). Native species richness is very high in the Klamath-Siskiyou region across plant and animal kingdoms (DellaSalla et al. 1999), and the Klamath River basin provides important habitat for several endemic fishes (Moyle 2002), including marbled sculpin (*Cottus klamathensis*) and Klamath speckled dace (*Rhinichthys klamathensis klamathensis*). Though municipal demand is less extreme than in the Central Valley because of less urbanization surrounding the Klamath River, contentious disputes over water rights in the Klamath have been ongoing since the Gold Rush. One of the largest dam removal projects in the world was completed in 2024 with the removal of four large Klamath River dams (Hancock 2026). Samples for this study were collected before dam removal, but will provide a valuable pre-dam removal reference point for future work. Goose Lake is a high-desert, endorheic lake that is typically isolated, though in very wet years, it can overflow into the Pit River watershed (Moyle 2002). In dry years, it can desiccate completely. Though Goose Lake was previously connected to the Klamath River, it has been disconnected for 1-3 million years. The Goose Lake basin is another region with a unique fish

assemblage, which is expected given the long isolation from other basins. In this study, samples from the Goose Lake basin include samples from Lassen Creek, a tributary in California and from several tributaries in Oregon (Figure 3.S1). Due to the small sample size at each tributary in Oregon, the sample location is referred to as 'Goose Lake' for all samples from Oregon tributaries (Figure 3.1). Notably, this study only includes samples collected from tributaries, not from Goose Lake itself, but previous records indicate lampreys are found in the lake as well (Hubbs 1925; Moyle 2002).

Sample collection: Tissue samples were collected opportunistically by ongoing salmonid monitoring programs across the state of California between 2018 and 2024. Field sampling methods varied, but most samples came from rotary screw traps. The larval life stage was targeted because this would allow us to infer approximate natal origin. Because *E. tridentatus* does not exhibit natal homing behavior, collections of adults from one stream would likely represent individuals from many different natal origins, complicating the investigation of genetic structure using geography of sampling locations. Though larval individuals were the primary focus, some adult specimens were included in the study to serve as reference individuals for adult morphological species identification. Because most samples collected for this study were in the larval phase, many were not identified to species or even genus level in the field. Initial collections included both *Entosphenus* and *Occidentis*, but all *Occidentis* individuals were identified and excluded from analyses. Locations for all *Entosphenus* collections are displayed in Figure 3.1.

Data collection: Genomic DNA (gDNA) was extracted from all lamprey tissue samples (N=666) using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol. Samples from each location were divided across plates to reduce plate effects in downstream analyses. Samples were normalized to each contain 50 ng of gDNA and the BestRAD protocol (Ali et al. 2016) was used to prepare samples for sequencing. Genomic DNA was digested with the *SbfI* restriction enzyme, and adaptors with individual barcodes were ligated to the ends of

digested gDNA. AMPure XP magnetic beads were used to wash away unligated adaptors and any fragments without adaptors. Individually barcoded samples were then pooled into one library and sonicated to obtain fragments within the target size range (200–500 bp). Fragment size was confirmed with gel electrophoresis and size selection was performed using Dynabead M-280 streptavidin magnetic beads to wash away any non-target fragments. The remainder of library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina, following the manufacturer's protocol. The final library fragment size was confirmed with gel electrophoresis and DNA concentration was quantified with a Qubit (Thermo Fisher). Short-read high-throughput sequencing was performed on two lanes on an Illumina HiSeq 4000x using paired-end 150 bp sequences. The sequencing was carried out by the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

Genetic Sample Identification: To validate the field identifications provided and to determine which of the collected samples belonged to the *Entosphenus* genus, a subset of individuals that were included in the RADseq library were sequenced at the mitochondrial cytochrome b (cytb) gene. These sequences were compared to publicly available sequences from voucher specimens confirmed to be members of the *Entosphenus* genus (Auringer et al. 2023). Because not all individuals within the RADseq data set were included in the cyt b sequencing, there were still many samples unidentified to the genus level. The individuals with species identifications confirmed using the cyt b sequences served as reference individuals to classify all samples in the RADseq dataset by genus, and species, when possible.

Data Processing: Raw reads were demultiplexed and aligned to the Pacific lamprey (*Entosphenus tridentatus*) reference genome (Etrf_v1; somatic genome; Hess et al. 2020) using the mem algorithm in bwa (Li and Durbin 2009). Two genome assemblies are available for *E. tridentatus*, but the other version (Etrm_v1; Hess et al. 2020) was assembled from germline cells. In *E. tridentatus*, portions of the germline genome are eliminated from somatic cells in a

process called programmed genome rearrangement (Timoshevskiy et al. 2017). Because gDNA for this study was extracted from somatic cells, data was aligned to the somatic genome assembly to avoid mapping errors.

Alignment, mapping quality, and coverage was calculated in SAMtools (Li et al. 2009). BAM files were sorted, indexed, and PCR duplicates were removed using SAMtools. Reads with low mapping quality (MAPQ < 20) were removed and individuals with poor coverage (<100,000 total mapped reads) were dropped from further analysis. Single nucleotide polymorphisms (SNPs) were called using the likelihood ratio test in ANGSD (Korneliussen et al. 2014) if they had a minimum minor allele frequency of at least 0.05, if the SNP was present in at least 50% of individuals, and if the SNP p-value <1.0E-12. Sites with poor raw sequence quality (sequence quality PHRED score <20) and poor mapping quality (map quality PHRED score <20) were excluded from further analyses.

Genotype likelihoods were used when possible to reduce biases related to genotyping errors in low-coverage, reduced-representation genomic sequencing data. Genotype likelihoods were calculated using the GATK model in ANGSD, and posterior probabilities were calculated for each site using allele frequencies as a prior. Genotype likelihoods were carried forward for analyses when posterior probabilities > 0.95.

When genotypes were required for analysis (rather than genotype likelihoods), an individual would only be assigned a genotype at a SNP if that individual's data included at least 8 sequencing reads (8x per-site, per-individual coverage) to reduce genotype error associated with low coverage data.

Data Analyses: We performed Principal Component Analysis (PCA) using the single-read sampling approach in ANGSD (-doIBS). This method randomly samples a single read at each SNP for each individual to estimate the covariance matrix. Principal Components (PCs) were calculated from the covariance matrix and PCA plots were created in R (v. 4.5.0; R Core Team 2025). Eigenvalues were calculated and scree plots were generated in R to determine the

optimal number of principal components to retain. We utilized a hierarchical approach to investigate genomic clustering at multiple levels (i.e. genus, species, population). First, we conducted PCA on: a) all individuals (both genera), b) all individuals in the *Entosphenus* genus, and c) all individuals within *E. tridentatus*, *E. lethophagus* (including Goose Lake lamprey), and *E. similis*, respectively. Additionally, we utilized region-specific PCA to explore the putative inversion regions on chromosomes 1, 2, 4, and 22 identified by Hess et al. (2020).

Maximum likelihood phylogenetic analysis was run in RAxML (Stamatakis 2014) using SNP genotype data (coverage >8x) for all *Entosphenus* individuals (N=307). After exporting SNP genotypes from ANGSD in PLINK format and converting from PLINK to VCF, and SNPs were filtered by missingness (maximum 25% missing data across all individuals) in VCFtools. Then, SNPs in high linkage disequilibrium ($r^2 > 0.25$; calculated in windows containing 1000 bp) were removed using BCFtools. The resultant VCF file was re-headered with sample names and converted to PHYLIP format, then input into RAxML. The RAxML analysis used the GTRGAMMA model of nucleotide substitution with an ascertainment bias correction (ASC_GTRGAMMA) using the Lewis correction method to account for the absence of invariant sites in SNP input data. Node support was assessed using rapid bootstrapping with the number of bootstraps required for stable support values determined automatically using the bootstrap convergence test (autoMRE). The resulting maximum likelihood consensus tree was midpoint rooted and visualized using the R package ggtree (Yu et al. 2017).

Graphical genotyping was conducted on all *Entosphenus* individuals to investigate genotypes at the chr1 inversion. Genotypes were called in ANGSD (doGeno -2) at sites with per-site, per-individual coverage >8x. Individuals and sites with over 75% missing genotype data were removed. Output genotypes (-1, 0, 1, 2) were converted to ABHN genotype format, and genotypes were plotted in R using the ABHgenotypeR package. The non-inverted haplotype was present in the reference genome used, so the reference allele is classified as the non-inverted genotype, and the alternate allele is classified as the inverted genotype.

To determine whether outlier loci are present in other regions of the genome across *E. tridentatus*, genome scans to detect loci under selection were conducted in the individual-level, differentiation-based program, pcadapt (v4.4.0, Luu et al. 2017; Privé et al. 2020). For all *E. tridentatus* individuals (N = 233), genotypes were called in ANGSD (doGeno -4) at sites with per-site, per-individual coverage >8x and output in PLINK format. Loci with >10% missing data were removed, SNPs were thinned to 1 SNP per 1000 bp, unmapped scaffolds were removed, and files were converted to bed/bim/fam format in PLINK. In pcadapt, PCA was conducted and a screeplot was used to determine the number of principal components driving differentiation (optimal K), p-values were computed, and adjusted p-values from a Bonferroni correction were used to identify outlier loci ($p < 0.05$). Outlier loci were visualized via a Manhattan plot, generated using qqman (Turner 2014).

Population genomic summary statistics (π , F_{ST} , and Edwards genetic distance) were calculated in snpR (Hemstrom and Jones 2023) using SNP genotype data. In these analyses, SNPs were filtered by coverage (excluded if per-site, per-individual coverage <8x), missingness (excluded if missing data >0.25), and SNPs were pruned to one SNP per 50 kb. Locations with less than three individuals were excluded from analyses. For each of these three population-level metrics (π , F_{ST} , and Edwards genetic distance), weighted averages were calculated to reduce bias introduced by varying levels of missing data between individuals. Nucleotide diversity (π) was also calculated for SNPs in the chr1 inversion in *E. tridentatus*; in this analysis, individuals were grouped by inversion genotype (homozygous inverted, heterozygous, homozygous non-inverted). Mantel tests for isolation-by-distance in *E. tridentatus* and *E. lethophagus* were also conducted in snpR.

Results:

Illumina sequencing generated 2,158,388 Megabases of raw data, and 92% of all bases had a PHRED quality score >30. After alignment, the average number of mapped reads per individual

was 1,288,115 reads. The average per-individual coverage of RAD loci was 7x. One plate (96 individuals) was dropped from further analysis due to barcode contamination during library preparation. Fifty-eight individuals (9% of total samples) were dropped from further analysis due to poor sequencing coverage (<100,000 total mapped reads). The remaining individuals (N=512) were considered of sufficient quality and were eligible for downstream analyses.

In the PCA of all individuals (N=512), PC1 (53.9%) split samples into the two genera (*Entosphenus* = 307 individuals; *Occidentis* = 88 individuals; Figure 3.S2). In a subsequent PCA excluding all *Occidentis* samples, PC1 and PC2 divided all *Entosphenus* individuals into three clusters: 1) *E. tridentatus*, 2) *E. lethophagus* and Goose Lake lamprey and 3) *E. similis* (Figure 3.2). In a PCA with only *E. lethophagus* and Goose Lake lamprey, PC1 and PC2 separate individuals into three clusters: 1) Goose Lake basin, including Goose Lake lamprey, 2) McCloud River and Pit River, and 3) South Fork Pit River (Figure 3.3). Within *E. similis*, PCA failed to concentrate variance to any particular axis, possibly due to the low sample size (N=12) relative to the number of SNPs (N=79,868) (Figure 3.S3).

In the most highly supported phylogenetic tree, all *E. tridentatus* samples are contained in a large polytomy, and *E. lethophagus* (including Goose Lake lamprey) and *E. similis* are reciprocally monophyletic sister clades, each with 100% bootstrap support (Figure 3.4).

Within *E. tridentatus*, PC1 divides individuals into three clusters, each cluster comprised of samples from various sampling locations (Figure 3.5a). In region-specific PCA of the chr1 inversion (chr 1 pos. 9908655 – pos. 15463359) across all *E. tridentatus* individuals, the same three clusters are observed (Figure 3.5b). When including SNPs from the whole genome, except SNPs from the chr1 inversion, we see no clustering pattern, and all individuals are dispersed randomly on the PCA plot (Figure 3.5c). In the region-specific PCA of the chr1 inversion across all *Entosphenus* individuals, *E. similis* and *E. lethophagus* cluster with the homozygous non-inverted *E. tridentatus* individuals (Figure 3.6). In PCA excluding SNPs in the chr1 inversion (Figure 3.S4), we see the same clustering by species as observed in genome-

wide PCA. Region-specific PCA was run for the three other putative chromosomal inversion regions (on chromosomes 2, 4, and 22) identified by Hess et al. (2020), but PCA results did not indicate that these putative inversions were polymorphic among our samples. Individuals in this study were fixed for the reference allele across these three regions, which indicates our samples all have the same inversion haplotypes as the reference genome (Etrf_v1; Hess et al. 2020). We can infer that all samples in this study are homozygous for the non-inverted haplotype at chr2, homozygous for the inverted haplotype at chr4, and we cannot infer inversion haplotype at chr22 due to the complexity of the putative chromosomal rearrangement.

Graphical genotyping of the entire chr1 across all *Entosphenus* individuals (N=270) shows a large (5.5 Mb) haploblock spanning the chr1 inversion (Figure 3.S5). Within the chr1 inversion, graphical genotyping shows that all *E. lethophagus* (N=54) and *E. similis* (N=12) individuals are homozygous for the reference allele (associated with the non-inverted haplotype) (Figure 3.7).

In *E. tridentatus*, most individuals (N=153, 75%) are homozygous for the alternate allele (associated with the inverted haplotype), many individuals (N=45, 22%) are heterozygous, and few individuals (N=6, 3%) are homozygous for the reference allele at all SNPs in the chr1 inversion. Diversity in genotypes at the chr1 inversion was observed at each geographic sampling location, and heterozygotes were found at every sampling location (Figure 3.8).

The pcadapt genome scan of *E. tridentatus* individuals (N=233) identified 58 outlier loci, with 42 of the 58 loci localized to the chr1 inversion (Figure 3.9). Four outlier loci fell within the 'genomic island of divergence' on chromosome 4 identified by Hess et al. (2020). The other 12 outlier loci were distributed across chromosomes 2, 3, 5, 10, 19, 25, 30, 38, 39, 44, and 60. All outlier loci and their p-values are listed in Table 3.S1.

Summary statistics estimating genomic differentiation (F_{ST} , genetic distance) between species (and geographic locations) showed moderate divergence between *E. tridentatus* and *E. lethophagus* (F_{ST} = 0.216–0.342; genetic distance = 0.301–0.364), and between *E. tridentatus* and *E. similis* (F_{ST} = 0.273–0.335; genetic distance = 0.339–0.353) (Table 3.1). *E. similis* and *E.*

lethophagus were also moderately differentiated from one another ($F_{ST} = 0.305\text{--}0.375$; genetic distance = 0.301–0.323). Samples from Goose Lake basin (including the northern streams in Oregon and Lassen Creek in California) were not substantially differentiated from *E. lethophagus* in the Pit River and McCloud River ($F_{ST} = 0.090\text{--}0.103$; genetic distance = 0.188–0.194). Within *E. tridentatus*, very little divergence was observed ($F_{ST} = 0.000\text{--}0.020$, genetic distance = 0.066–0.153). Mantel tests showed no significant correlation of genetic distance and geographic distance in *E. tridentatus* ($r = -0.054$, $p = 0.50$) or *E. lethophagus* ($r = 0.296$, $p = 0.10$).

Genome-wide nucleotide diversity was highest in *E. tridentatus* ($\pi = 0.283\text{--}0.294$), slightly lower in *E. lethophagus* ($\pi = 0.174\text{--}0.231$), and lowest in *E. similis* ($\pi = 0.154$) (Table 3.2). When considering only SNPs in the inversion region on chromosome 1, nucleotide diversity was highest in *E. tridentatus* individuals that were homozygous for the reference haplotype (non-inverted; $\pi = 0.436$), intermediate in heterozygotes ($\pi = 0.398$), and lowest in individuals homozygous for the inversion haplotype ($\pi = 0.286$).

Discussion:

This study finds support for recognizing *E. tridentatus*, *E. similis*, and *E. lethophagus* as distinct species. The Goose Lake lamprey (*Entosphenus* sp.) was consistently nested within *E. lethophagus* and similar to *E. lethophagus* collected from the Goose Lake basin. Within *E. tridentatus*, individuals did not exhibit neutral population genetic structure except at the putative inversion region on chromosome 1, which has previously been identified as a ‘genomic island of divergence’ implicated in local adaptation (Hess et al. 2020). Within *E. similis*, no population structure was observed between the two sampling locations, suggesting that relatively high gene flow occurs in this species across its range in the Klamath River basin. In the non-migratory species, *E. lethophagus*, we observed relatively low levels of genomic differentiation between geographically distant sites, indicating that these populations should all be classified as

one species. Goose Lake lamprey (*Entosphenus* sp.) were not genomically differentiated from *E. lethophagus* collected from the Goose Lake basin, contradicting the hypothesis that Goose Lake lamprey should be classified as a distinct species.

Our findings support the designation of *E. similis* and *E. lethophagus* as distinct species from *E. tridentatus*. Principal component analysis of all *Entosphenus* individuals (Figure 3.2) divides samples into three distinct clusters corresponding to species and phylogenetic analysis resolves *E. similis* and *E. lethophagus* as reciprocally monophyletic clades with maximum bootstrap support (Figure 3.4). Though the magnitude of genomic divergence is less than what has been observed between species in *Occidentis* (Auringer et al. 2026; Chapter 2), pairwise F_{ST} and genetic distance values (Table 3.1) demonstrate clear differentiation consistent with species-level genomic differences. This result contradicts the genetic patterns observed in single gene mtDNA analyses (Auringer et al. 2023), highlighting the importance of genomic data, which provides a more complete representation of genetic patterns than studies relying on a single or few genes. The relatively shallow divergences observed between species within *Entosphenus* in California may be explained by recent gene flow between *E. tridentatus* and the two other species. Alternatively, *E. similis* and *E. lethophagus* may have diverged from *E. tridentatus* relatively recently.

E. similis and *E. lethophagus*

Entosphenus similis individuals collected from the Shasta site and Scott site comprised one phylogenetic clade and clustered together in the PCA of all *Entosphenus* individuals, but PCA within *E. similis* failed to determine a major axis of variation to allow for inferences of population structure (Figure 3.S3). Overall, results suggest relatively high gene flow across the limited range of *E. similis* throughout the Klamath River basin, but population structure may be present across more geographically distant sites.

Further genomic investigations with increased sample sizes and more sampling sites in the Klamath River basin are needed to determine population structure within *E. similis*. Additionally,

Miller Lake lamprey (*E. minimus*), a small, parasitic species endemic to the upper Klamath basin, was not included in this study given that it was not expected in our study area. However, future genomic research should include *E. minimus* to clarify relationships between all *Entosphenus* species in the Klamath River basin. Notably, all samples in this study from the Klamath River were collected before dam removal (completed 2024). *Entosphenus similis* is considered freshwater-resident, but moves between streams and lakes, and the extent of this movement is not well understood. The removal of the Klamath dams may change migratory behavior, dispersal, and patterns of population structure of Klamath River lamprey and all lamprey species in the Klamath River basin.

Genomic differentiation was observed across geographically distant sampling sites within *E. lethophagus* in PCA (Figure 3.3) as well as F_{ST} and genetic distance values (Table 3.1), suggesting genetically distinct populations are found within this species at geographically distant sites. This intraspecific divergence is expected due to the lack of migration and limited dispersal in non-migratory species. Among the sites where *E. lethophagus* was collected, three genetically distinct clusters were distinguished: 1) McCloud River and mainstem Pit River, 2) South Fork Pit River, and 3) Goose Lake basin (northern Oregon sites and Lassen Creek in California).

Goose Lake lamprey

Goose Lake lamprey (*Entosphenus* sp.; formally undescribed) have been taxonomically ambiguous since this group was initially described as anatomically distinct (Hubbs 1925; Moyle 2002). Lampreys within the Goose Lake basin are understudied but records indicate the presence of both parasitic and non-parasitic lampreys. Two major questions have been outstanding regarding Goose Lake lampreys: 1) whether these parasitic and non-parasitic lampreys are life history variants within one species, or two distinct species occupy the Goose Lake basin, and 2) whether the lampreys in Goose Lake should even be classified as a distinct

species, or whether they are a morphologically distinct population of *E. tridentatus* or *E. lethophagus* with sufficient plasticity to express both parasitic and non-parasitic phenotypes. Our analyses included adult individuals morphologically identified as parasitic Goose Lake lamprey (N=4) as well as adults with non-parasitic morphology, identified as *E. lethophagus* (N=3), all collected from the Goose Lake basin. We found no genomic divergence between Goose Lake lamprey and *E. lethophagus* within the Goose Lake basin, refuting the hypothesis that two distinct species inhabit Goose Lake. Hubbs (1925) noted that male Goose Lake lamprey exhibit characteristics more typical of brook lamprey with atrophied intestines and precocious sexual development. Female Goose Lake Lamprey on the other hand, resembled “normal parasitic young” of *E. tridentatus* (Hubbs 1925, p. 589). We do not have associated sex metadata with our samples, but evolutionary trade-offs such as the creation of energy-intensive eggs versus sperm may contribute to different life history strategies for sexes and associated anatomical differentiation (Crook et al. 2010; Rundio et al. 2012). While we are unable to validate that sex may be driving these apparent outward anatomical differences, it is a testable hypothesis. Overall, low genomic divergence between individuals from Goose Lake and *E. lethophagus* from the Pit River, McCloud River, and South Fork Pit River further support the hypothesis that all of the lampreys across these sites are part of a single species with genetically distinct populations.

Under this hypothesis, the single species contains both parasitic and non-parasitic types, and variation in dispersal potential. Within the Goose Lake basin, previous work found lampreys at approximately 40% of all sampled sites between 2007 and 2022 (Dickey 2024), and flow, sinuosity, and elevation are positively associated with lamprey presence in Goose Lake tributaries (Dickey et al. 2025). High gene flow observed across the entire Goose Lake basin suggests that individuals move extensively between the lake and its tributaries. The modest, yet evident genomic differentiation ($F_{ST} = 0.008\text{--}0.194$) between Goose Lake, Pit River, McCloud

River, and South Fork Pit River populations indicate that outside the Goose Lake basin, individuals are relatively immobile and behave more like a typical non-migratory brook lamprey. Because *E. lethophagus* is currently defined as a non-migratory, non-parasitic lamprey, with underdeveloped dentition, reclassifying all Goose Lake lampreys under *E. lethophagus* is insufficient. A possible alternative would be to expand the definition of *E. lethophagus* to include the observed life history diversity. Under this scheme, the common name (Pit-Klamath Brook Lamprey) should be used only to describe the non-migratory, non-parasitic life history variant, since 'brook' in the common name signifies non-migratory lampreys and often implies a lack of parasitism. A separate common name should be established to refer to the parasitic form. Future studies should investigate whether the parasitic form observed in Goose Lake is present across the rest of the *E. lethophagus* range to inform accurate taxonomic classification and description of the life history variation.

Pacific lamprey (*Entosphenus tridentatus*)

Entosphenus tridentatus individuals are highly genetically similar throughout most of the genome, across all sampling locations. Genome-wide genetic distance and F_{ST} values are low ($F_{ST} < 0.02$) between all sampling sites (Table 3.1). Phylogenetic results show that all *E. tridentatus* samples are part of a large polytomy (Figure 3.4), further supporting the lack of genetic differentiation throughout this species in California. These results support the hypothesized lack of natal homing in this anadromous species. Though *E. tridentatus* lacks population structure, nucleotide diversity in this species is the highest of all *Entosphenus* species in California. This is not surprising given the relatively large population size and high levels of gene flow across the entire species range.

Within *E. tridentatus*, genome-wide PCA shows three distinct genetic clusters (Figure 3.5a), but the clusters are not associated with sampling location. This pattern is driven by a single 5.5 Mb region of the genome on chromosome 1 (chr 1 pos. 9908655 – pos. 15463359), which suggests this region may contain large effect loci. In PCA when this region is excluded, individuals are

randomly scattered with no discernable clusters, (Figure 3.5c), indicating that genomic differentiation in *E. tridentatus* is isolated to this genomic region. Previous work has identified this region on chromosome 1 as a putative inversion region (Hess et al. 2020), but this pattern could also be explained by other mechanisms of recombination suppression. Nevertheless, our results from region-specific PCA, pcadapt outlier analysis, graphical genotyping, and region-specific estimates of nucleotide diversity, all substantiate the claim that this haploblock is the result of a chromosomal inversion. Though studies have shown that natural selection can maintain relatively large (10 kb – 500 kb) divergent haploblocks without structural variation (Lamichhaney & Andersson 2019), very large haploblocks, as observed in *E. tridentatus*, are still most likely explained by a chromosomal inversion. Overall, our study provides multiple lines of indirect evidence supporting the hypothesis presented by Hess et al. (2020) that this haploblock on chromosome 1 is most likely caused by a chromosomal inversion, and we will continue to refer to this region as the ‘chr1 inversion’.

Adaptive genetic variation is more easily detected in species without population structure like *E. tridentatus* (Hess et al. 2014). Interestingly, pcadapt genome scan results from this study (Figure 3.9) suggest that adaptive variation is primarily limited to the chr1 inversion. The majority of total outlier loci identified fall within this region (N = 42 SNPs; 72% of all outliers; Table 3.S1) The other potentially adaptive variation falls into the previously identified region on chromosome 4, but few SNPs (N=3) were identified as outliers and *p*-values were significant but substantially higher than all outliers on chromosome 1 (Table 3.S1).

The pcadapt method has important limitations, including the lack of a direct association with environmental variables or phenotypes and the sensitivity to missing data, but is advantageous given that it does not require pre-defined populations and does not rely on population genetics models (Flanagan 2018). Additionally, the key finding from this analysis – that the chr1 inversion contains adaptive variation – is strongly supported by previous findings (Hess et al. 2020) that have conducted association testing with phenotypes and environmental variables.

Each cluster on the region-specific PCA of all *E. tridentatus* individuals at the chr1 inversion (Figure 3.5b) corresponds to the inversion genotypes, verified by graphical genotyping (Figure 3.7). Heterogeneous nucleotide diversity across the putative inversion genotypes (homozygous inverted: $\pi = 0.286$; heterozygous: $\pi = 0.398$; homozygous non-inverted: $\pi = 0.436$) also supports the hypothesis that this haploblock is caused by a chromosomal inversion. The reduced nucleotide diversity across the chr1 inversion in homozygous inverted individuals suggests that the inversion first occurred relatively recently (on evolutionary timescales). Regardless of how this haploblock on chromosome 1 was created and maintained, this region is likely a reservoir of important genetic diversity and should be considered in genetic management of this species. Upon further characterization of this region, it should be considered in delineating distinct population segments of *E. tridentatus*. Adaptive variation, especially large-effect loci like those associated with run-timing in salmon, are generally accepted as an important component of delineating management units (Waples et al. 2022). Across all individuals in the study area, there is a stark imbalance of genotypes, with relatively few heterozygous individuals and very few individuals homozygous for the non-inverted genotype. This excess of homozygous inverted individuals in the study area may be related to the coastal/interior gradient of SNP genotypes observed in this chromosomal region; previous studies have shown that coastal populations have higher proportions of the ocean-maturing genotypes and interior populations have higher proportions of the stream-maturing genotypes (Hess et al. 2020, Clemens et al. 2026). Additionally, SNPs in this region have already been identified as adaptive (associated with reproductive maturation timing), so inversion haplotypes are likely to be out of Hardy-Weinberg Equilibrium (HWE). Because of the extensive gene flow throughout *E. tridentatus*, in order to assess and interpret deviations from HWE, the inversion should be genotyped across the entire species range.

Across the landscape, inversion genotypes are distributed relatively uniformly in the study area. In *E. tridentatus*, genotypes at the chr1 inversion were distributed heterogeneously across the

Sacramento and San Joaquin River basins (Figure 3.8). Though the most abundant genotype in *E. tridentatus* was homozygous inverted (N = 158), none of the sampled locations contained 100% inverted haplotypes. Heterozygotes were less common (N = 45), but were found at all locations. Homozygous non-inverted were the least frequent genotype (N = 6); interestingly, this genotype is disproportionately represented in the Klamath River basin (3 out of 8 individuals). Though previous work has not investigated the distribution of inversion genotypes across the landscape, studies tracking SNP genotypes within this chromosomal region (i.e. Etr_464; Hess et al. 2020, Clemens et al. 2026) have found that there is a coastal vs. interior association with genotypes. The Klamath River is the most coastal site in this study, which may explain the high proportion of homozygous non-inverted genotypes at this site compared to all other sites in the Central Valley. Range-wide monitoring of inversion genotypes will clarify whether the distribution of inversion genotypes is consistent with the SNP genotype patterns.

Previous research has demonstrated an association between SNPs in the chr1 inversion and egg mass, a proxy for reproductive maturation timing (ocean-/river-maturing ecotypes) in *E. tridentatus* (Hess et al. 2012, Parker et al. 2019, Hess et al. 2020). Gene functions within the chr1 inversion were previously investigated with enrichment analysis; while the analysis for fish gene functions had no statistically significant tests, reproductive behavior, response to absence of light, and entrainment of circadian clock by photoperiod were the highest ranked phenotypic terms, and genes related to these phenotypes would likely be good candidates for further exploration of gene functions in this region (Hess et al. 2020). The specific SNPs associated with these phenotypes were not captured in our genomic data, so we cannot associate inversion genotype to phenotype. However, the observed diversity of chr1 inversion genotypes suggests that ecotypic diversity is present in our study area. Also, the large size of this region (~5.5 Mb) and the strong signal of adaptive variation identified in this study suggest that additional important functional genetic diversity is contained in this region. *Entosphenus tridentatus* exhibits high levels of gene flow across its range, and inversions often act as a mechanism to

“bundle” adaptive alleles together in species with high gene flow in order to protect these adaptive ‘supergenes’ from recombination (Zhang et al. 2026).

Based on region-specific PCA of all *Entosphenus* individuals at the chr1 inversion, both *E. similis* and *E. lethophagus* appear to be homozygous for the ancestral (non-inverted) haplotype (Figure 3.6). Given the high levels of genome-wide differentiation between these species even when the chr1 inversion is excluded (Figure 3.S4), the region-specific PCA alone is an insufficient method to determine inversion genotypes for *E. similis* and *E. lethophagus*. When the region-specific PCA results are combined with the graphical genotyping results (Figure 3.7), our data confirm that all *E. similis* and *E. lethophagus* individuals are all homozygous for the ancestral (non-inverted) haplotype.

Four SNPs within the previously identified ‘genomic island of divergence’ on chromosome 4 (Hess et al. 2020) were also identified as putatively adaptive in the pcadapt analysis (Figure 3.9; Table 3.S1). The signal in this region was not as strong as that observed for the chromosome 1 region, but the findings warrant further exploration of this region on chromosome 4 as a putatively adaptive genomic region in California. Though the pcadapt analysis identified 12 additional outlier SNPs, none were clustered together as expected in adaptive regions, nor were they supported by other lines of evidence or previous research. Therefore, these 12 SNPs do not have sufficient evidence to be considered putatively adaptive. The lack of a strong signal of adaptive variation on regions (i.e. chr 2, chr 22) identified as putatively adaptive by Hess et al. (2020) may be explained by the geographic differences between studies. The populations in California sampled in this study are likely fixed for one allele at these loci, while the same loci are polymorphic across a wider geographic scope. Future range-wide studies will help elucidate these patterns.

Conclusion

Genome-wide differentiation is observed between *E. tridentatus*, *E. similis*, and *E. lethophagus*, which largely supports the current taxonomic classifications based on anatomical differentiation. However, the lack of genomic differentiation between Goose Lake lamprey and *E. lethophagus* indicates that Goose Lake lampreys should not be classified as a distinct species or subspecies and should likely be considered part of *E. lethophagus*. The discordance between previous mtDNA results and genome-wide SNP results highlights the limitations of mitochondrial and single-gene studies and emphasizes the utility of genomic data in clarifying evolutionary relationships. Within *E. tridentatus* in California, genomic differentiation is primarily restricted to one specific region of the genome, the chr1 inversion. This region likely contains adaptive variation and should be included in genetic monitoring and may inform translocation decisions. Additionally, this region should be considered when delineating distinct population segments to ensure the maintenance of this adaptive variation.

References

- Auringer, G., Campbell, M. A., Goertler, P. A., & Finger, A. J. (2023). Lampreys in California (Lampetra spp. and Entosphenus spp.): Mitochondrial phylogenetic analysis reveals previously unrecognized lamprey diversity. *North American Journal of Fisheries Management*, 43(6), 1511-1530.
- 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(11), 775-788.
- Clemens, B. J., Beamish, R. J., Coates, K. C., Docker, M. F., Dunham, J. B., Gray, A. E., ... & Wyss, L. A. (2017). Conservation challenges and research needs for Pacific Lamprey in the Columbia River basin. *Fisheries*, 42(5), 268-280.
- Clemens, B. J., & Wang, C. J. (2021). Dispelling misperceptions of native lampreys (*Entosphenus* and *Lampetra* spp.) in the Pacific Northwest (USA). *Conservation Science and Practice*, 3(6), e402.
- Clemens, B. J., Litts, R., Vonderohe, G., Schaefer, J., Medeiros, L. R., Nagler, J. J., & Hess, J. E. (2026). Lamprey genotypes predispose maturation timing phenotypes in coastal and interior river basins. *Canadian journal of fisheries and aquatic sciences*, 83, 1-17.
- 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.
- Crook, D. A., Koster, W. M., Macdonald, J. I., Nicol, S. J., Belcher, C. A., Dawson, D. R., ... & Bannam, L. (2010). Catadromous migrations by female tupong (*Pseudaphritis urvillii*) in coastal streams in Victoria, Australia. *Marine and Freshwater Research*, 61(4), 474-483.
- DellaSala, D. A., Reid, S. B., & Frest, T. J. (1999). A Global Perspective on the Biodiversity of. *Natural Areas Journal*, 19, 4.

- DellaSala, D. A., Baker, B. C., Rogers, M. H., Bond, M., Bury, G., Bury, R. B., & Strittholt, J. R. (2026). An Ecoregional Conservation Assessment for the Klamath-Siskiyou Ecoregion and Proposed Siskiyou Crest Climate Refuge, Southwest Oregon and Northern California, USA. *Diversity*, 18(7), 415.
- Dickey, J. C. (2024). Lamprey of the Goose Lake Basin: A Multi-Scale Habitat & Distribution Analysis. Master's thesis, Oregon State University.
- Dickey, J. C., Clemens, B. J., Dumelle, M., & Davis, M. J. (2025). Modeling lamprey distribution using flow, geomorphology, and elevation in a terminal lake system. *Transactions of the American Fisheries Society*, 154(3), 322-338.
- Docker, M. F., Haas, G. R., Goodman, D. H., Reid, S. B., & Heath, D. D. (2007). PCR-RFLP markers detect 29 mitochondrial haplotypes in Pacific lamprey (*Entosphenus tridentatus*). *Molecular Ecology Notes*, 7(2), 350-353.
- Docker, M. F., & Potter, I. C. (2019). Life history evolution in lampreys: alternative migratory and feeding types. In *Lampreys: biology, conservation and control: Volume 2* (pp. 287-409). Dordrecht: Springer Netherlands.
- USFWS. (2004). Endangered and Threatened Wildlife and Plants; 90-Day Finding on a Petition to List Three Species of Lampreys as Threatened or Endangered, 69 Fed. Reg. 77,158, 77,166 (codified at 50 C.F.R. pt. 17).
- Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary applications*, 11(7), 1035-1052.
- Goodman, D. H., Reid, S. B., Docker, M. F., Haas, 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(2), 400-417.

- Hancock, K. (2026). Comparative study of the Potter Valley Project and Klamath River' respective dam removals. Honor's College thesis, Oregon State University.
- Hess, J. E., Smith, J. J., Timoshevskaya, N., Baker, C., Caudill, C. C., Graves, D., ... & Narum, S. R. (2020). Genomic islands of divergence infer a phenotypic landscape in Pacific lamprey. *Molecular Ecology*, 29(20), 3841-3856.
- Hughes, L. C., Bloom, D. D., Piller, K. R., Lang, N., & Mayden, R. L. (2025). Phylogenomic resolution of lampreys reveals the recent evolution of an ancient vertebrate lineage. *Proceedings of the Royal Society B: Biological Sciences*, 292(2038), 20242101.
- Kostow, K. (2002). *Oregon lampreys: natural history, status, and problem analysis*. Oregon Department of Fish and Wildlife [Fish Division].
- Laćan, I., & Resh, V. H. (2016). A case study in integrated management: Sacramento–San Joaquin Rivers and Delta of California, USA. *Ecohydrology & Hydrobiology*, 16(4), 215-228.
- Lamichhaney, S., & Andersson, L. (2019). A comparison of the association between large haplotype blocks under selection and the presence/absence of inversions. *Ecology and Evolution*, 9(8), 4888-4896.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760.
- Lin, B., Zhang, Z., Wang, Y., Currens, K. P., Spidle, A., Yamazaki, Y., & Close, D. A. (2008). Amplified fragment length polymorphism assessment of genetic diversity in Pacific lampreys. *North American Journal of Fisheries Management*, 28(4), 1182-1193.
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67–77. <https://doi.org/10.1111/1755-0998.12592>

- Maitland, P. S., Renaud, C. B., Quintella, B. R., Close, D. A., & Docker, M. F. (2014). Conservation of native lampreys. In *Lampreys: Biology, Conservation and Control: Volume 1* (pp. 375-428). Dordrecht: Springer Netherlands.
- Mérot, C. (2020). Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Molecular Ecology*, 29(14), 2513-2516.
- Moyle, P. B. (2002). *Inland fishes of California: revised and expanded*. Univ of California Press.
- 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(13), 3171-3185.
- 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. Master's thesis, Oregon State University.
- Privé, F., Luu, K., Vilhjálmsson, B. J., & Blum, M. G. B. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution*, 37(7), 2153–2154. <https://doi.org/10.1093/molbev/msaa053>
- Rehm, E. M., Olivas, P., Stroud, J., & Feeley, K. J. (2015). Losing your edge: climate change and the conservation value of range-edge populations. *Ecology and evolution*, 5(19), 4315-4326.
- Reid, S. B., & Goodman, D. H. (2020). Natural recolonization by Pacific lampreys in a Southern California coastal drainage: Implications for their biology and conservation. *North American Journal of Fisheries Management*, 40(2), 335-341.
- Renaud C. B. (2011) Lampreys of the world: an annotated and illustrated catalogue of lamprey species known to date. *Food and Agriculture Organization of the United Nations, Species Catalogue for Fishery Purposes*, No 5.

- Rundio, D. E., Williams, T. H., Pearse, D. E., & Lindley, S. T. (2012). Male-biased sex ratio of nonanadromous *Oncorhynchus mykiss* in a partially migratory population in California. *Ecology of Freshwater Fish*, 21(2), 293-299.
- Sheoships, G. (2014). Pacific lamprey *Entosphenus tridentatus*: integrating traditional ecological knowledge and contemporary values into conservation planning, and stream substrate associations with larval abundance in the Willamette River Basin, Oregon, USA. Master's thesis, Oregon State University.
- 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.
- Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Journal of Open Source Software*, 3: 731 doi: 10.21105/joss.00731
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in ecology and evolution*, 8(1), 28-36.
- Wang, C., & Schaller, H. (2015). Conserving Pacific Lamprey through collaborative efforts. *Fisheries*, 40(2), 72-79.
- Wang, C. J., Schaller, H. A., Coates, K. C., Hayes, M. C., & Rose, R. K. (2020). Climate change vulnerability assessment for Pacific Lamprey in rivers of the Western United States. *Journal of Freshwater Ecology*, 35(1), 29-55.
- Wicks-Arshack, A., Dunkle, M., Matsaw, S., & Caudill, C. (2018). An ecological, cultural and legal review of Pacific lamprey in the Columbia River Basin. *Idaho L. Rev.*, 54, 45.
- Zhang, X., Hu, Y., Huang, K., Wright, S. I., & Rieseberg, L. H. (2026). Recombination suppression in plant adaptation and speciation. *New Phytologist*, 250(4), 2061-2077.

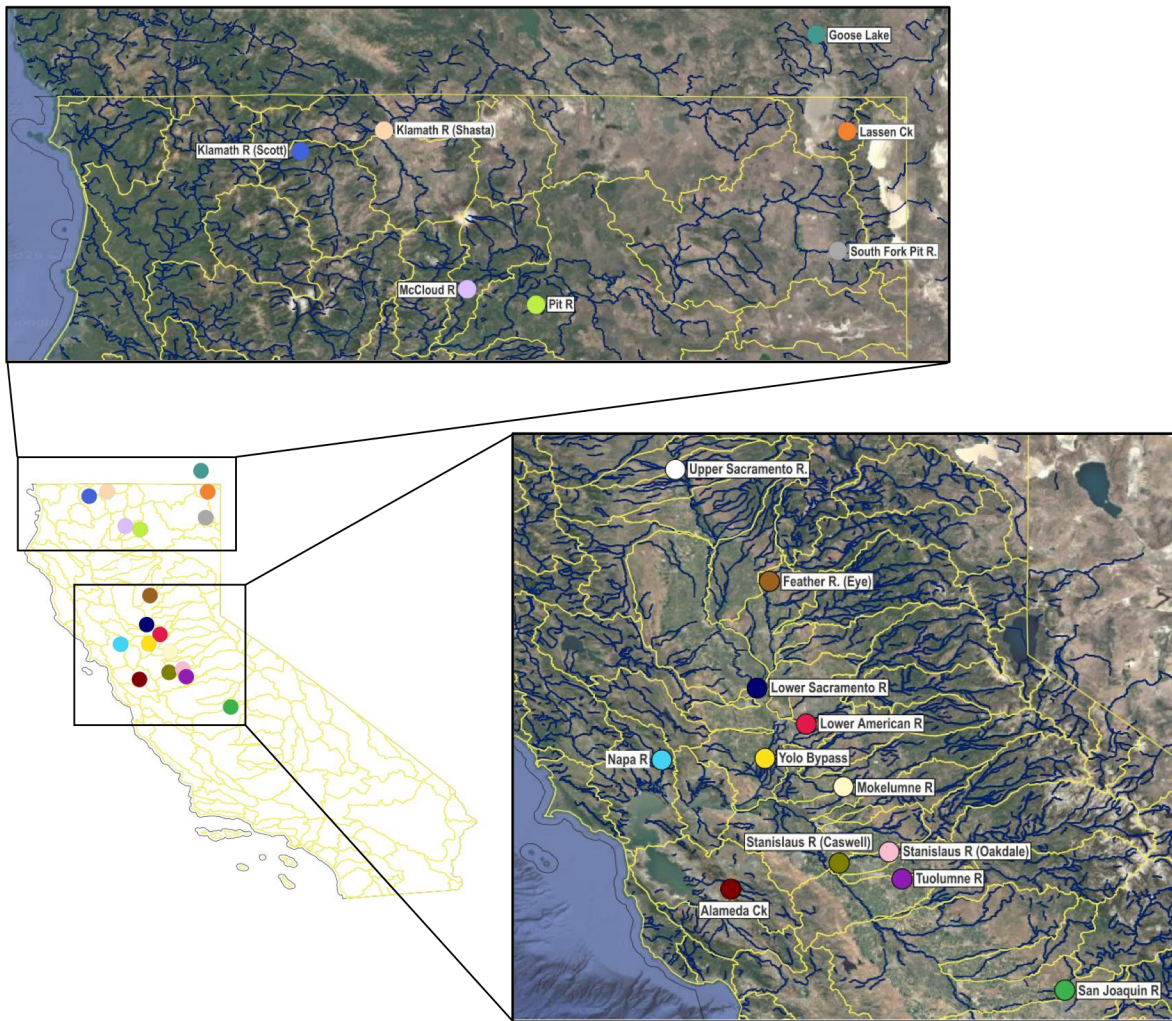


Figure 3.1. Map of all sampled locations for *Entosphenus*. Colors of the circles are specific to that location and utilized in subsequent figures to represent which samples came from each location. In location names, “R” means River and “Ck” means Creek. Yellow lines represent HUC-8 (Hydrologic Unit Code-8) watershed boundaries.

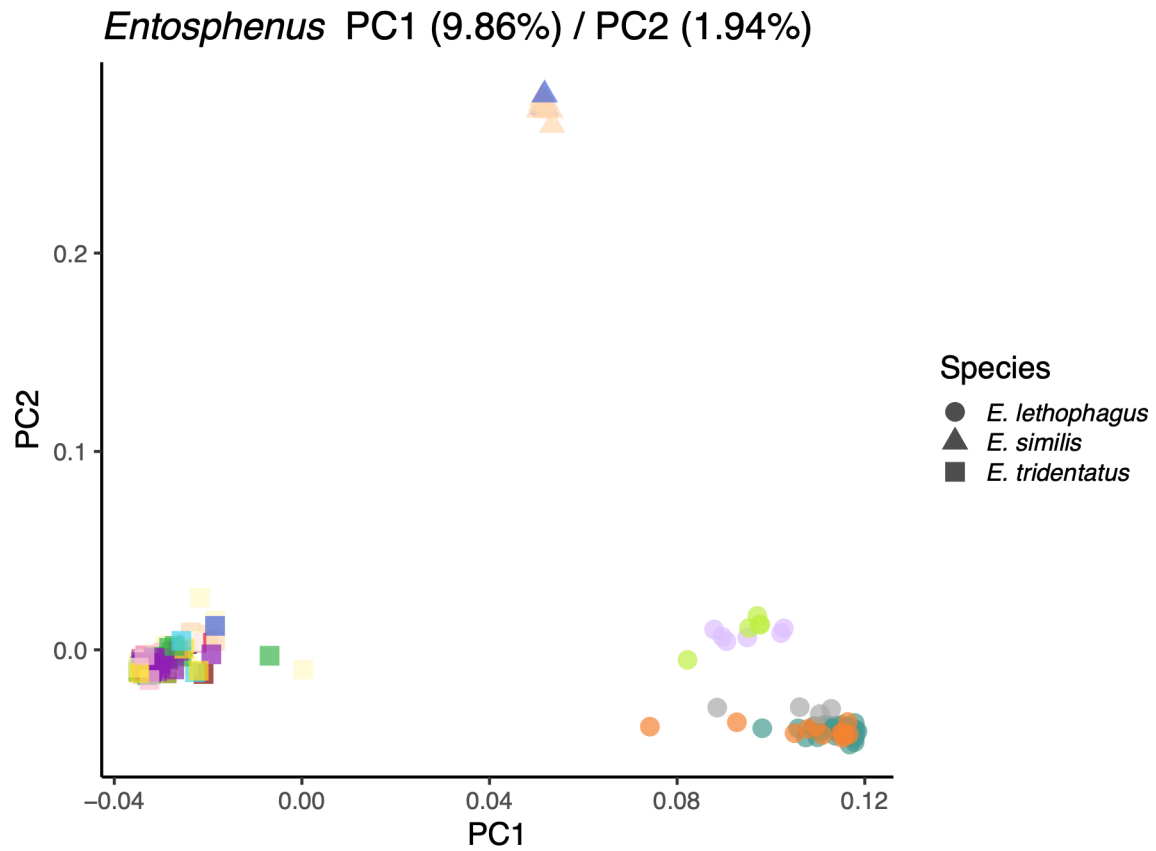
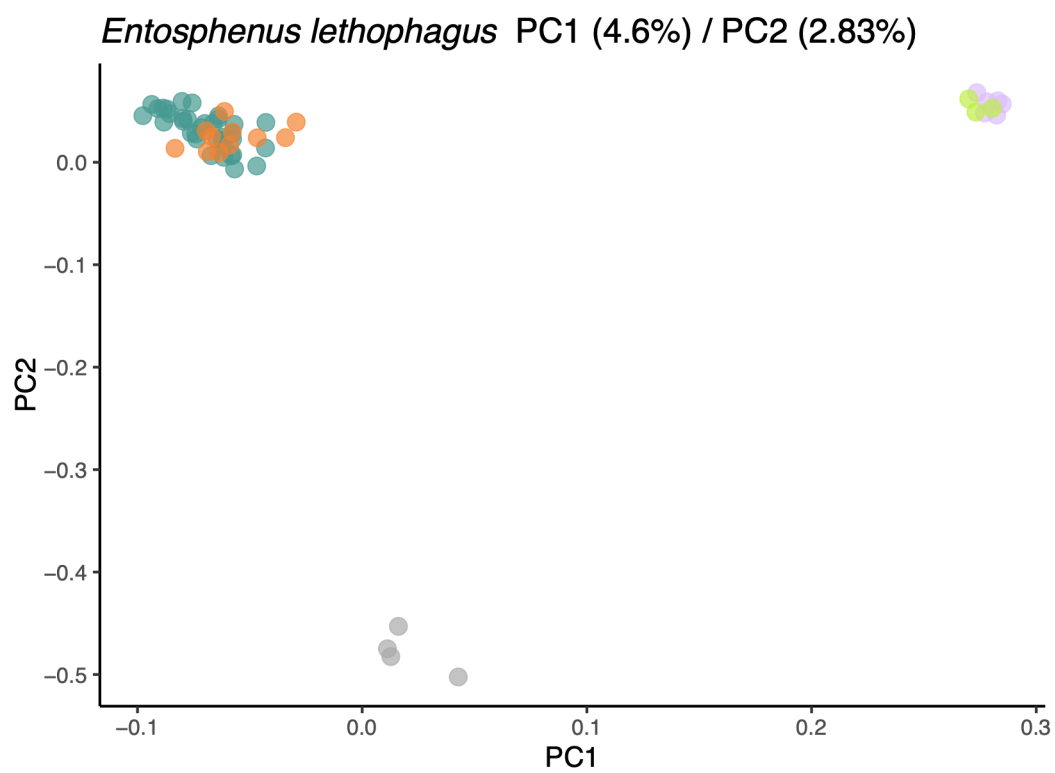
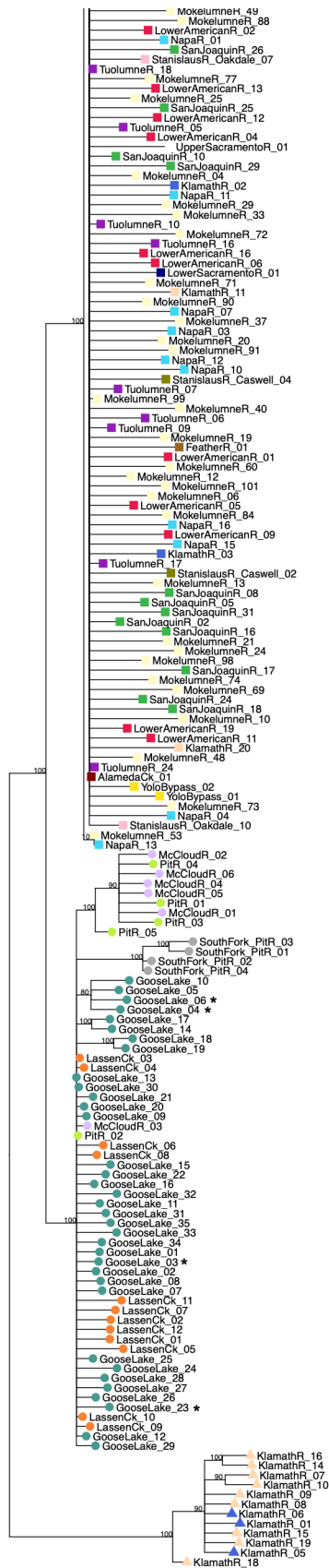


Figure 3.2. PCA of all *Entosphenus* samples. Points are colored by sampling location and point shape is determined by species.





Species
 ● *E. lethopagus*
 ▲ *E. amia*
 ■ *E. tridentatus*

Figure 3.4. (continued) Phylogenetic tree of all *Entosphenus* individuals. Points are colored by sampling location and point shape represents species. Individuals identified morphologically as Goose Lake lamprey are indicated by asterisks. Larger, high resolution version of this figure is attached in supplementary files.

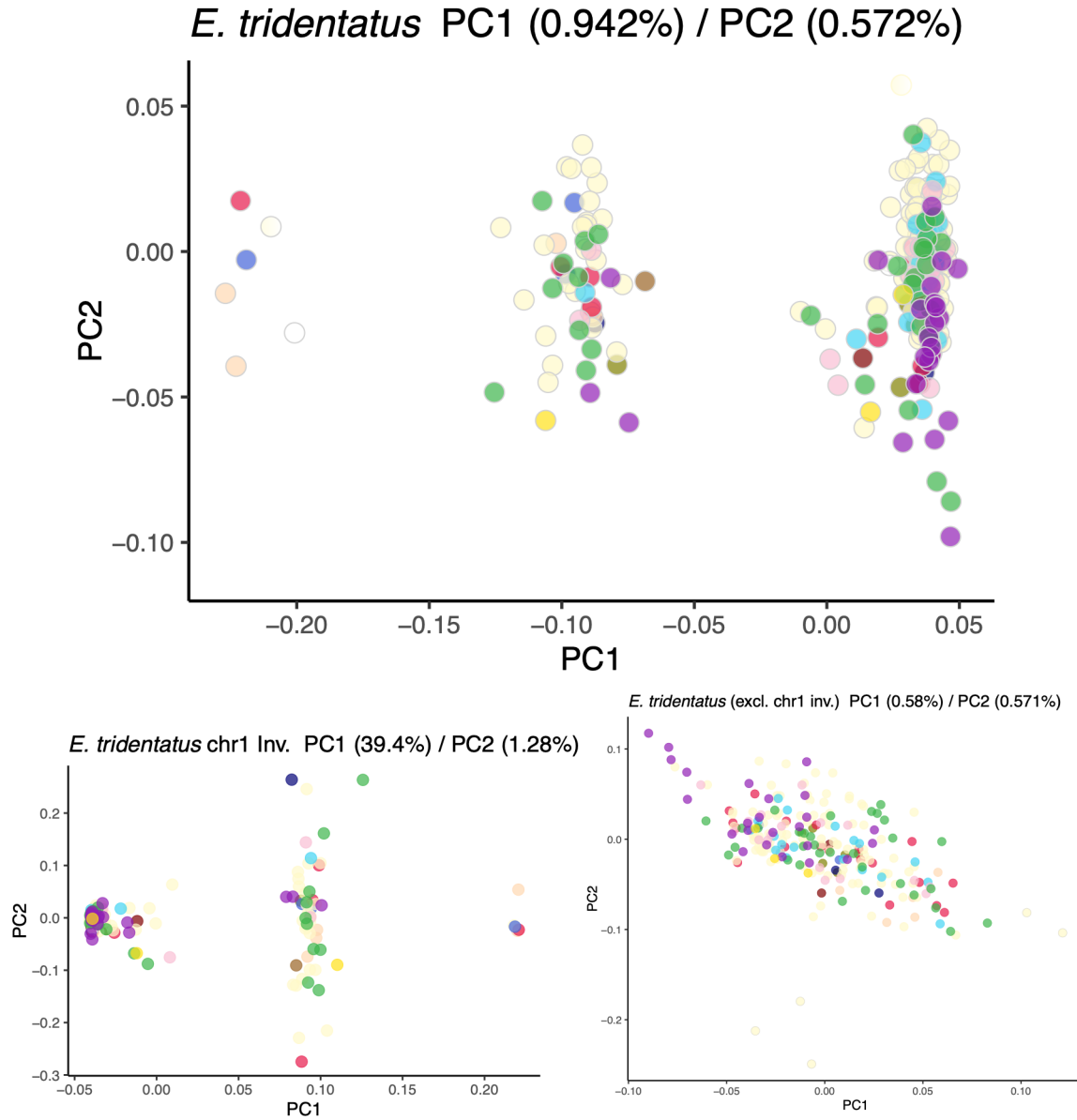


Figure 3.5. a.) PCA of all *E. tridentatus* samples. Points are colored by sampling location. **b.)** Region-specific PCA of chr1 inversion, including all *E. tridentatus* samples. Points are colored by sampling location. **c.)** PCA of all *E. tridentatus* individuals at SNPs throughout the genome, except SNPs in the chromosome 1 putative inversion region. Points are colored by sampling location.

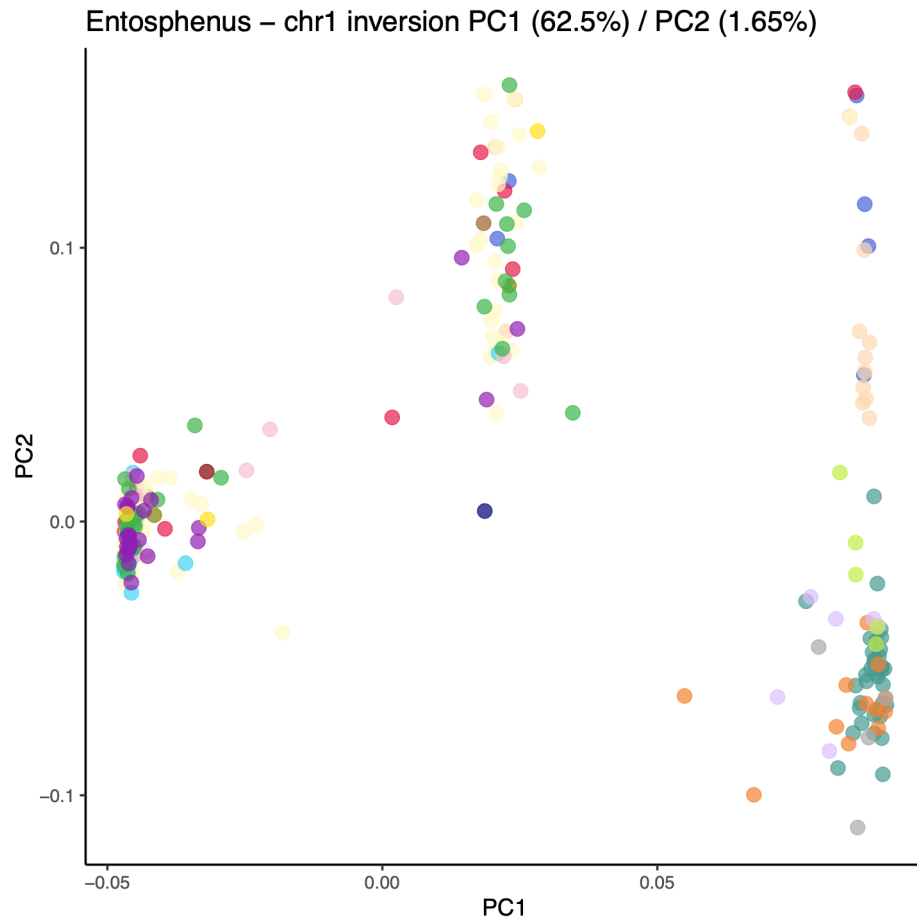


Figure 3.6. Region-specific PCA of chr1 inversion, including all Entosphenus samples. Points are colored by sampling location.

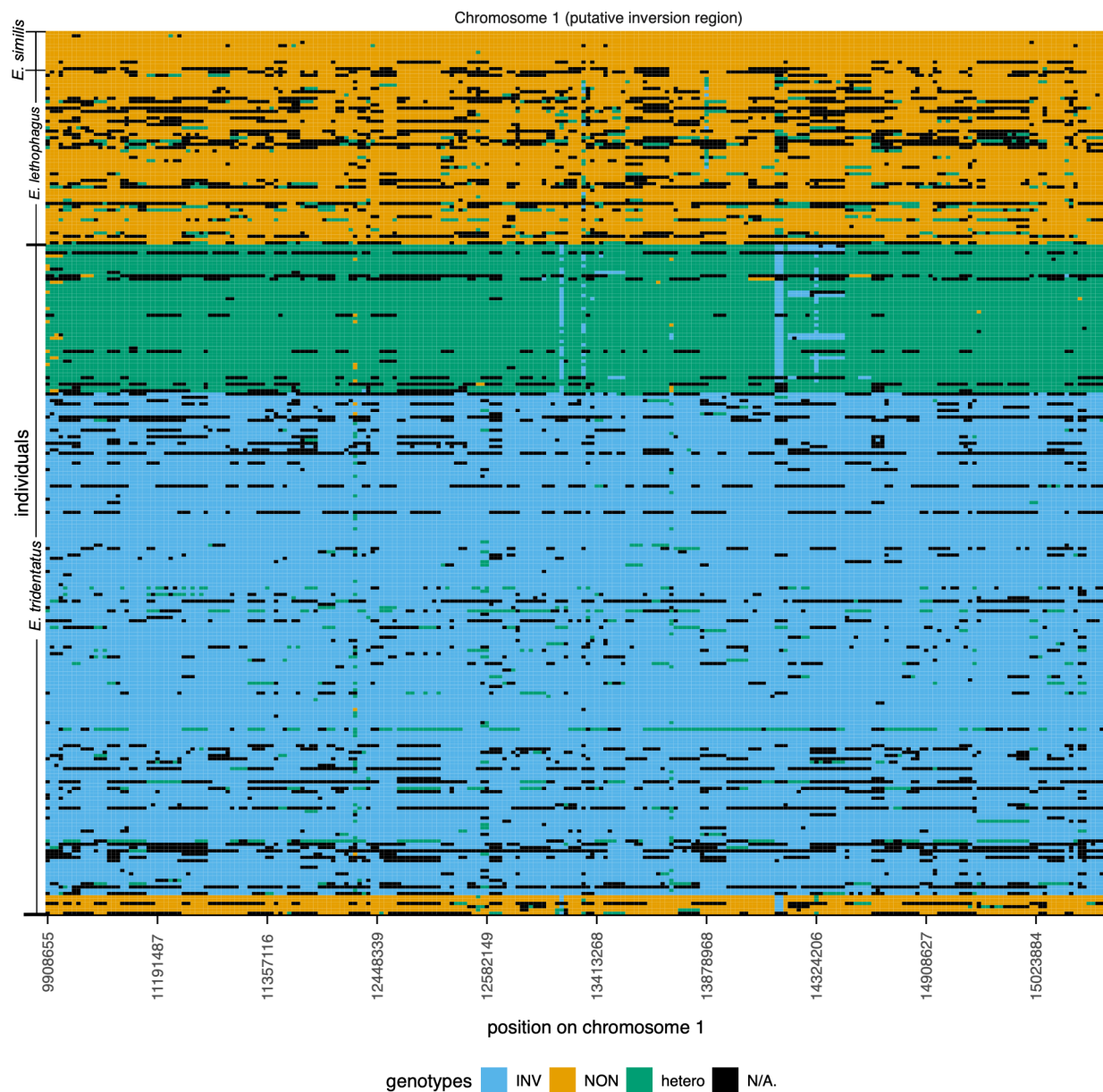


Figure 3.7. Graphical genotyping of chr1 inversion. Each row represents one individual, and each column represents one SNP. Rows are sorted by species, and within *E. tridentatus*, rows are sorted by inversion genotype. Blue indicates homozygous inverted genotypes, green indicates heterozygous genotypes, and orange indicates homozygous non-inverted genotypes.

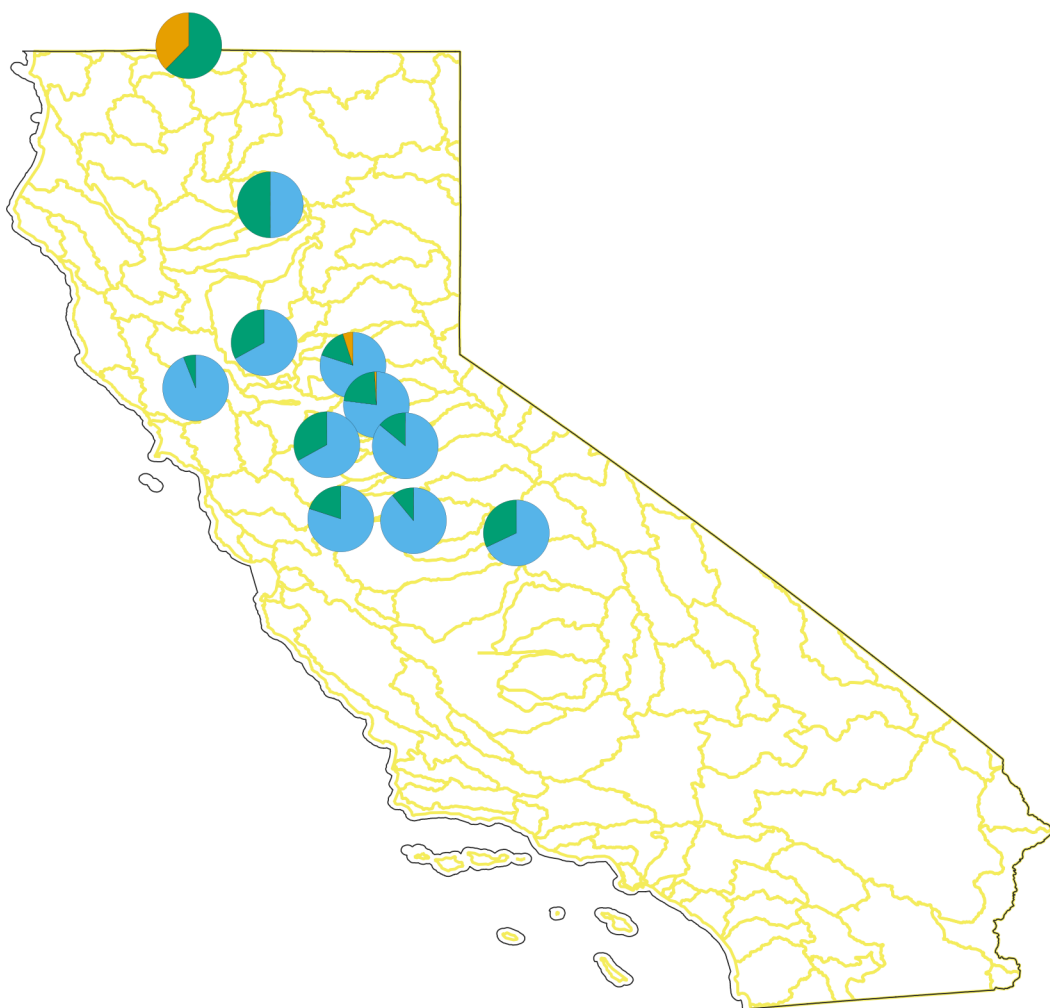


Figure 3.8. Map of the distribution of chr1 inversion genotypes across *E. tridentatus*. Pie charts indicate proportions of genotypes at each sampling location. Blue indicates homozygous inverted genotypes, green indicates heterozygous genotypes, and orange indicates homozygous non-inverted genotypes.

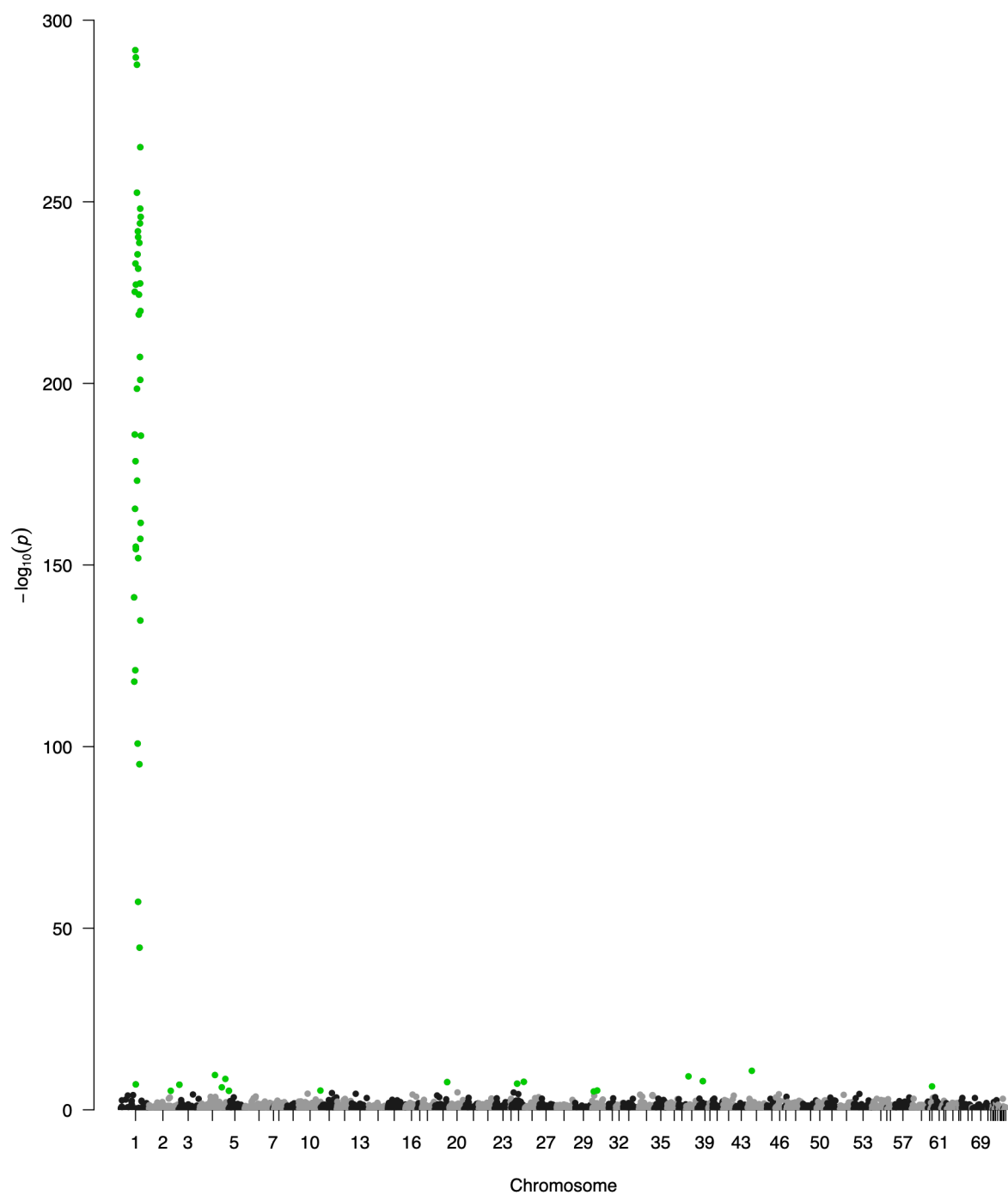


Figure 3.9. Manhattan plot of all loci included in pcadapt genome scan. Outliers are highlighted in green.

	F _{ST}											
	E. lethophagus - Goose Lake	E. lethophagus - Lassen Creek	E. lethophagus - McCloud River	E. lethophagus - Pit River	E. similis - Klamath River	E. tridentatus - Klamath River	E. tridentatus - Lower American River	E. tridentatus - Mokelumne River	E. tridentatus - Napa River	E. tridentatus - Stanislaus River (Oakdale)	E. tridentatus - San Joaquin River	E. tridentatus - Tollumne River
E. lethophagus - Goose Lake		0.011	0.103	0.090	0.375	0.327	0.340	0.289	0.324	0.342	0.315	0.342
E. lethophagus - Lassen Creek	0.124		0.086	0.070	0.353	0.271	0.285	0.254	0.277	0.287	0.269	0.288
E. lethophagus - McCloud River	0.188	0.194		0.008	0.329	0.246	0.263	0.241	0.256	0.265	0.251	0.268
E. lethophagus - Pit River	0.188	0.189	0.164		0.305	0.216	0.235	0.219	0.233	0.237	0.227	0.239
E. similis - Klamath River	0.323	0.320	0.309	0.301		0.321	0.331	0.273	0.315	0.335	0.303	0.334
E. tridentatus - Klamath River	0.351	0.327	0.337	0.319	0.341		0.012	0.015	0.019	0.017	0.013	0.020
E. tridentatus - Lower American River	0.359	0.333	0.344	0.326	0.349	0.145		0.001	0.000	0.001	0.000	0.002
E. tridentatus - Mokelumne River	0.346	0.319	0.331	0.311	0.339	0.122	0.095		0.001	0.001	0.001	0.003
E. tridentatus - Napa River	0.356	0.329	0.340	0.322	0.348	0.138	0.111	0.073		0.001	0.002	0.002
E. tridentatus - Stanislaus River (Oakdale)	0.362	0.336	0.347	0.328	0.353	0.153	0.129	0.101	0.115		0.001	0.004
E. tridentatus - San Joaquin River	0.352	0.326	0.338	0.319	0.344	0.130	0.105	0.066	0.089	0.112		0.002
E. tridentatus - Tollumne River	0.362	0.336	0.347	0.328	0.353	0.150	0.125	0.096	0.109	0.131	0.107	
	GENETIC DISTANCE											



Table 3.1. Pairwise F_{ST} (above diagonal) and genetic distance (below diagonal) values between each sampled location. Cells are colored by value such that red cells have highest values and green cells have lowest values.

Species	Sample Location	Sample Size	Nucleotide Diveristy (π)
<i>E. lethophagus</i>	Goose Lake	14	0.174
<i>E. lethophagus</i>	Lassen Creek	8	0.200
<i>E. lethophagus</i>	McCloud River	5	0.205
<i>E. lethophagus</i>	Pit River	5	0.231
<i>E. lethophagus</i>	South Fork Pit River	2	N/A
<i>E. similis</i>	Klamath River	11	0.154
<i>E. tridentatus</i>	Stanislaus (Caswell)	3	N/A
<i>E. tridentatus</i>	Feather River	1	N/A
<i>E. tridentatus</i>	Klamath River	7	0.293
<i>E. tridentatus</i>	Lower American River	11	0.288
<i>E. tridentatus</i>	Mokelumne	93	0.286
<i>E. tridentatus</i>	Napa	17	0.283
<i>E. tridentatus</i>	Stanislaus (Oakdale)	10	0.294
<i>E. tridentatus</i>	Upper Sacramento River	1	N/A
<i>E. tridentatus</i>	San Joaquin River	20	0.286
<i>E. tridentatus</i>	Tuolumne River	12	0.288
<i>E. tridentatus</i>	Yolo Bypass	1	N/A

Table 3.2. Table including sample sizes for snpR analyses and nucleotide diversity values for each sampling location.

chr_#	chr_name	pos	allele1	allele2	p-value
1	CM025607.2	10261440	G	T	8.19E-142
1	CM025607.2	10438643	A	C	1.29E-118
1	CM025607.2	10808687	G	A	1.21E-186
1	CM025607.2	10833286	G	C	6.04E-226
1	CM025607.2	11017554	G	A	3.39E-166
1	CM025607.2	11191469	C	T	9.34E-122
1	CM025607.2	11199786	A	G	1.81E-292
1	CM025607.2	11356969	A	C	1.02E-233
1	CM025607.2	11391500	C	A	2.64E-179
1	CM025607.2	11540198	T	G	9.37E-156
1	CM025607.2	11553305	C	A	9.90E-08
1	CM025607.2	11561131	C	T	1.79E-290
1	CM025607.2	11616622	C	G	4.02E-155
1	CM025607.2	11653234	A	G	6.34E-228
1	CM025607.2	12448338	G	T	1.83E-288
1	CM025607.2	12450623	G	C	3.06E-253
1	CM025607.2	12463527	A	C	3.00E-199
1	CM025607.2	12582075	G	A	6.00E-174
1	CM025607.2	12908772	C	A	2.87E-236
1	CM025607.2	13003815	G	A	1.45E-101
1	CM025607.2	13167969	T	C	1.36E-242
1	CM025607.2	13318844	T	C	5.42E-58
1	CM025607.2	13378862	G	A	5.26E-241
1	CM025607.2	13413268	T	C	2.60E-232
1	CM025607.2	13417118	A	C	1.31E-152
1	CM025607.2	13878816	C	T	1.11E-219
1	CM025607.2	13980081	T	A	3.67E-225
1	CM025607.2	14285910	C	T	1.96E-239
1	CM025607.2	14324085	A	G	7.27E-96
1	CM025607.2	14426676	A	G	2.23E-45
1	CM025607.2	14733540	G	A	5.31E-208
1	CM025607.2	14758204	T	C	8.67E-245
1	CM025607.2	14874121	G	A	2.95E-228
1	CM025607.2	14908423	G	C	7.67E-249
1	CM025607.2	14937693	G	A	1.07E-201
1	CM025607.2	14971584	G	A	9.40E-266
1	CM025607.2	15017425	A	G	6.33E-158
1	CM025607.2	15023656	G	T	1.91E-135
1	CM025607.2	15122041	G	A	1.20E-220
1	CM025607.2	15251582	A	T	2.49E-162
1	CM025607.2	15271880	T	C	1.39E-246
1	CM025607.2	15463270	T	C	2.46E-186
2	CM025608.2	16656680	A	G	5.85E-06
3	CM025609.2	603343	A	G	1.20E-07
4	CM025610.2	11547120	G	A	2.65E-10
4	CM025610.2	16748031	G	T	6.44E-07
4	CM025610.2	19444160	T	C	3.12E-09
5	CM025611.2	180745	T	C	5.80E-06
10	CM025616.2	16640622	T	C	4.93E-06
19	CM025625.2	9529060	C	A	2.36E-08
25	CM025631.2	4372377	T	C	6.22E-08
25	CM025631.2	9446994	T	C	1.94E-08
30	CM025636.2	1281868	A	T	8.84E-06
30	CM025636.2	3923974	C	T	4.78E-06
38	CM025644.2	2023652	G	A	6.07E-10
39	CM025645.2	4057820	T	C	1.30E-08
44	CM025650.2	3233014	T	C	1.80E-11
60	CM025666.2	1206921	A	G	3.53E-07

Table 3.S1. List of outlier SNPs identified in pcadapt analysis and their respective p-values and metadata.

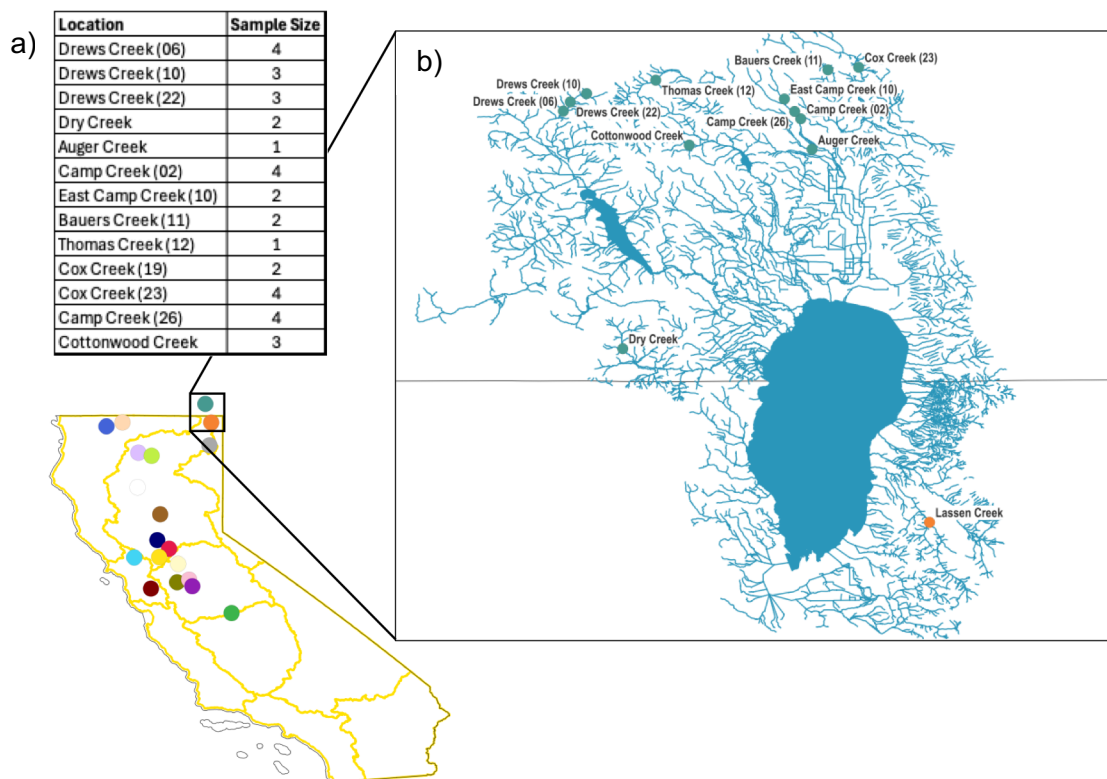


Figure 3.S1. a) Sample sizes of each sampled Oregon tributary of Goose Lake. b) Map of all 'Goose Lake' sampling sites, including Lassen Creek and all sampled tributaries in Oregon.

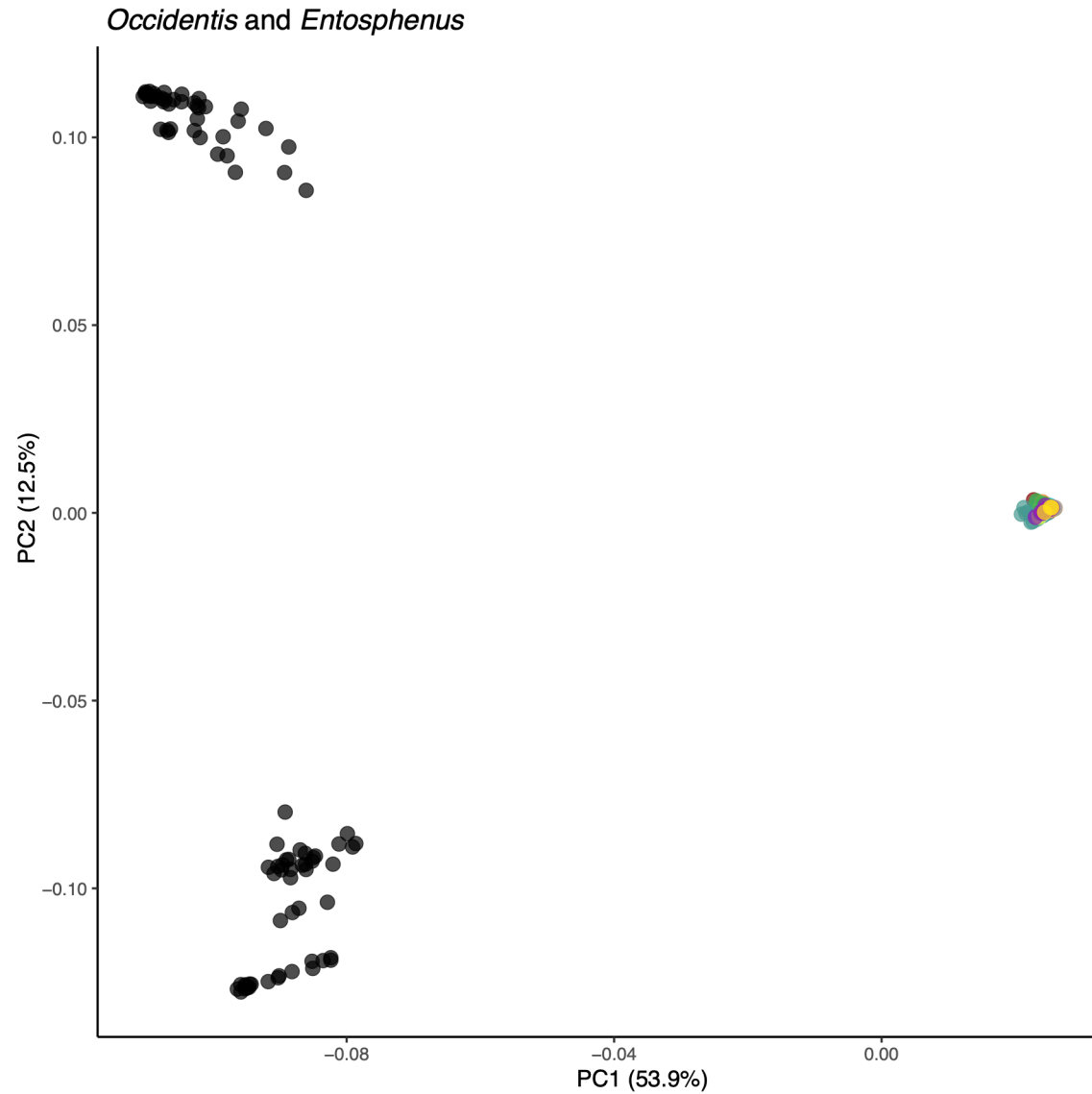


Figure 3.S2. PCA plot of all individuals (*Entosphenus* and *Occidentis*) that passed minimum quality filters. *Entosphenus* points are colored by sampling location (reference Figure 3.1 in main text), and *Occidentis* points are black.

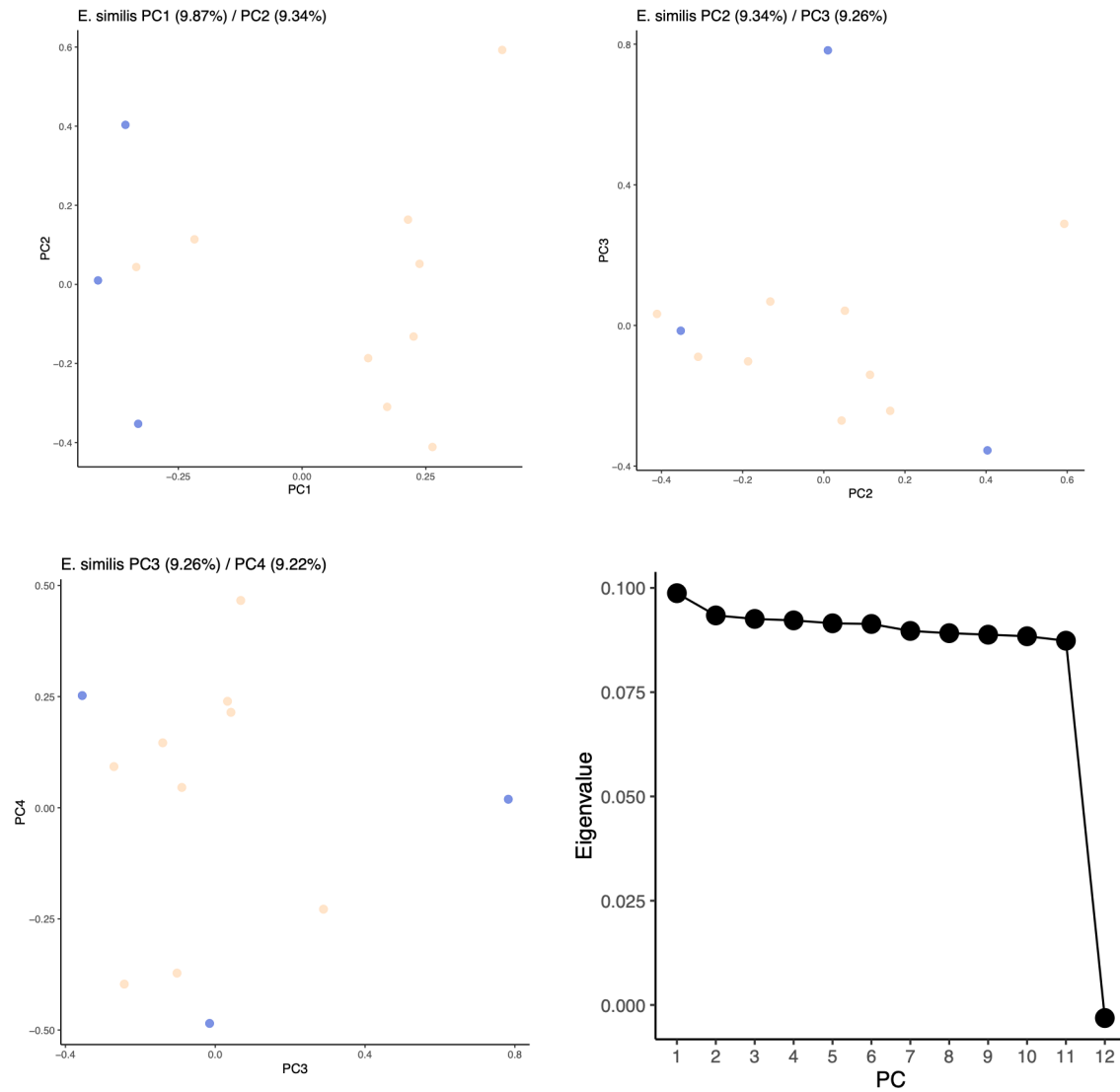


Figure 3.S3. a-c) PCA plots of all *E. similis*. Plots for first 4 PCs displayed.
d) Scree plot for PC 1-11.

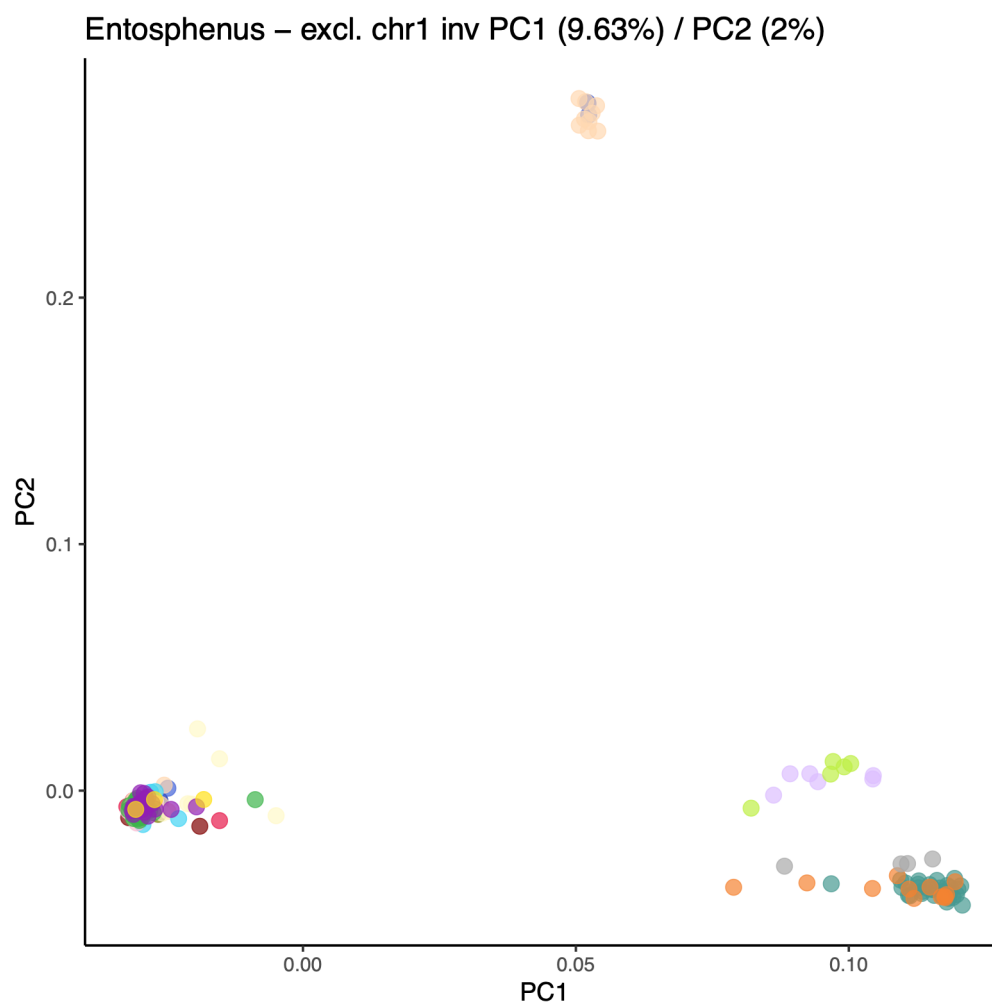


Figure 3.S4. PCA plot of genome-wide SNPs excluding chr1 inversion across all *Entosphenus* individuals. Points are colored by sampling location.

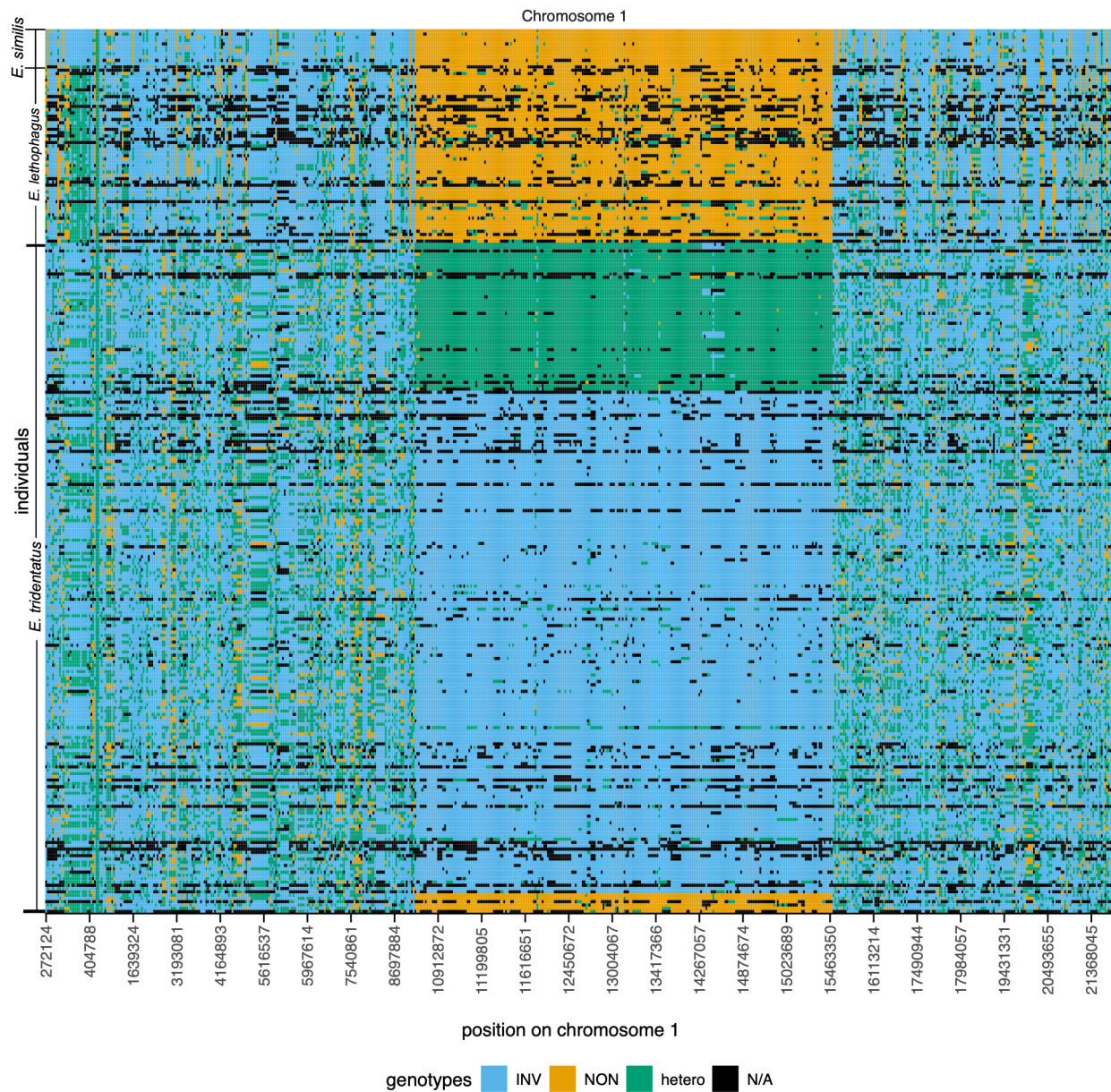


Figure 3.S5. Graphical genotyping of entire chromosome 1. Each row represents one individual, and each column represents one SNP. Rows are sorted by species, and within *E. tridentatus*, rows are sorted by inversion genotype. Blue indicates homozygous inverted (reference) genotypes, green indicates heterozygous genotypes, and orange indicates homozygous non-inverted (alternate) genotypes.