



Canadian Journal of Fisheries and Aquatic Sciences

Genomic study of longfin smelt (*Spirinchus thaleichthys*) in the northeastern Pacific reveals regional patterns of genetic drift, structure, and gene flow

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2021-0005
Manuscript Type:	Article
Date Submitted by the Author:	05-Jan-2021
Complete List of Authors:	Saglam, Ismail; UC Davis, Animal Science; Koc University, Department of Molecular Biology and Genetics Hobbs, James; University of California - Davis, Baxter, Randall; California Department of Fish and Wildlife Stockton Lewis, Levi; University of California Davis, Department of Wildlife, Fish and Conservation Biology Benjamin, Alyssa ; UC Davis, Animal Science Finger, Amanda; UC Davis, Animal Science
Keyword:	osmeridae, longfin smelt, population genetics, forage fish, CONSERVATION < General
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

1 **Genomic study of longfin smelt (*Spirinchus thaleichthys*) in the**
2 **northeastern Pacific reveals regional patterns of genetic drift,**
3 **structure, and gene flow**

4

5

6 Ismail Saglam^{1,2*}, James Hobbs³, Randall Baxter⁴, Levi S. Lewis³,
7 Alyssa Benjamin², Amanda J. Finger^{2*}

8

9 1. Koç University, Department of Molecular Biology and Genetics,
10 Istanbul, Turkey

11 2. Department of Animal Science, University of California,
12 Davis, CA, 95616 USA

13 3. Department of Wildlife, Fish and Conservation Biology,
14 University of California, Davis, CA 95616 USA

15 4. California Department of Fish and Wildlife, Region 3, 2109
16 Arch-Airport Rd., Suite 100, Stockton CA 95206

17

18 *co-corresponding authors

19

20 Keywords: Forage fish; Local adaptation; Osmeridae; Population genetics

21

22 Abstract

23 The southernmost stock of longfin smelt (*Spirinchus*
24 *thaleichthys*) is approaching extirpation in the San Francisco
25 Estuary (SFE); however, genetic structure and gene flow within
26 the species is poorly understood. Here, we use genome-wide data
27 to evaluate population structure of longfin smelt across a broad
28 latitudinal scale across estuaries ranging from the SFE to
29 Yakutat Bay and Lake Washington, and fine scale within the
30 Fraser River and the SFE. Results indicate high genetic
31 structure between major estuaries, fine-scale structure within
32 the Fraser River, and low levels of structure within the SFE.
33 Genetic structure was more pronounced between northern estuaries
34 whereas southern estuaries showed shared ancestry and ongoing
35 gene flow, most notably unidirectional northward migration out
36 of the SFE. Furthermore, we detected signatures of local
37 adaptation within the Fraser River and the Skeena River
38 estuaries. Taken together, our results identify broad patterns
39 of genetic diversity in longfin smelt shaped by co-ancestry,
40 unidirectional migration and local adaptation. Results also
41 suggest that the SFE population is genetically distinct from
42 northernmost populations and an important source for maintaining
43 southern populations.

44

45

46

47

48

49

Draft

50 **Introduction**

51 Small pelagic forage fishes form an energy conduit from
52 primary producers to top predators and are a critical link in
53 coastal food webs (Lenfest Forage Fish Task Force 2012; Pikitch
54 et al. 2014). While genetic structure and connectivity among
55 populations of commercially important forage fish species such
56 as herring, anchovies, and sardines are relatively well-studied
57 (Hedgecock et al. 1989; Spanakis et al. 1989), other forage fish
58 species are poorly understood despite pressing conservation
59 needs. Various Osmeridae “true smelt” species inhabiting the
60 west coast of North America have shown coast-wide population
61 declines. Eulachon (*Thaleichthys pacificus*) was once abundant in
62 many estuaries from San Francisco Bay to the southeastern Bering
63 Sea, but for the past two decades has declined in the southern
64 portion of its range. The Southern Distinct Population Segment
65 (DPS) of eulachon is listed as threatened under the Federal
66 Endangered Species Act (ESA) (Gustafson et al. 2010; Gustafson
67 et al. 2016; Sutherland et al. 2020). In Puget Sound, surf smelt
68 (*Hypomesus pretiosus*), a common, recreationally-harvested forage
69 fish has precipitously declined (Greene et al. 2015), and in the
70 San Francisco Estuary (SFE), the delta smelt (*H. transpacificus*)
71 and the longfin smelt (*Spirinchus thaleichthys*) are possibly
72 approaching extinction and extirpation, respectively, in the
73 near future (Hobbs et al. 2017).

74 Longfin smelt is a small (fork length < 150 mm) pelagic
75 Osmerid forage fish that inhabits lakes, estuaries, and marine
76 habitats from central California to Prince William Sound, Alaska
77 (Dryfoos 1965, Moyle 2002). The southernmost, and one of the largest,
78 known populations occurs in the SFE (Figure 1). Much of what is
79 known about longfin smelt comes primarily from monitoring
80 surveys in the San Francisco Estuary, California (SFE;
81 Rosenfield and Baxter 2007; Baxter et al. 2008; Maunder, Deriso
82 et al. 2015; Nobriga and Rosenfield 2016; Grimaldo et al. 2017,
83 2020; Lewis et al. 2020) and research studies on the land-locked
84 population in Lake Washington, King County, WA (Moulton 1974;
85 Sibley and Chigbu 1994; Chigbu and Sibley 1998a, 1998b; Chigbu
86 2000). Longfin smelt in the SFE are semi-anadromous, spawning
87 and rearing in fresh to low-salinity waters in winter-spring,
88 migrating to higher-salinity brackish and marine waters in early
89 summer. Little is known about spawning behavior, though it
90 appears that they spawn opportunistically in locations with
91 favorable habitat conditions, with the males arriving first,
92 followed by the females (Moyle 2002). Larvae and juvenile longfin
93 smelt are sometimes observed in smaller estuaries indicating
94 that smaller estuaries are utilized as spawning and rearing
95 habitat (Lewis et al. 2020). In summer, juveniles move into
96 higher salinity coastal waters of the Pacific Ocean, including
97 San Francisco Bay, where they are observed as fisheries by-catch

98 or in monitoring surveys (Rosenfield and Baxter 2007; Garwood
99 2017). Similarly, longfin smelt have been observed in the
100 nearshore coastal environment offshore of the Columbia River
101 plume (Bottom and Jones 1990; Litz et al. 2014), in Bellingham
102 Bay, Washington, near the Nooksack River (Donnelly et al. 1988,
103 Penttila 2007), in Yakutat Bay in the discharge of two glaciers
104 (Arimitsu 2016) and the Cook Inlet, Alaska (Abookire et al.
105 2000; Abookire and Piatt 2005; Jones et al. 2005). Thus, some
106 individuals clearly spend some portion of their life history in
107 the ocean. Documented estuarine rearing (Bottom et al. 1984, Rosenfield and
108 Baxter 2007) and limited detections in nearshore forage fish surveys
109 (Roberge et al. 2002, Penttila 2007) suggest that most fish
110 inhabiting the coast likely remain near their natal estuary and
111 return after a season or two on the coast (e.g., Rosenfield and
112 Baxter 2007). Given their typical 2-year life cycle, it is
113 unclear how long they reside in the ocean, and this may vary
114 among populations, estuaries, or among individuals within an
115 estuary. Little is known about the ocean-phase of their life
116 history, and how ocean rearing and the potential for transport
117 or movement may influence their contemporary population
118 structure, demography, or evolutionary dynamics.

119 Longfin smelt was once one of the most abundant pelagic
120 forage fish in the SFE, even contributing to a commercial bait
121 fishery (Skinner 1962; Moyle 2002). However, based on numerous

122 long-term monitoring surveys, the SFE population has undergone
123 several significant declines in abundance since the 1960s
124 (Thomson et al. 2010; Maunder et al. 2015; Nobriga and
125 Rosenfield 2016; Hobbs et al. 2017). Longfin smelt is one of
126 four species that declined in the early 2000s, an event known as
127 the pelagic organism decline (POD) within the SFE (Sommer et al.
128 2007). The POD was attributed to anthropogenic pressures
129 including habitat alteration, increased water diversions, food-
130 web alteration, and invasive species (Sommer et al. 2007; Mac
131 Nally et al. 2010; Thomson et al. 2010). In response to this
132 decline, in 2009 the SFE population of longfin smelt was listed
133 as threatened under the California Endangered Species Act
134 (California Fish and Game Commission 2009), and a petition to
135 list it as a Distinct Population Segment under the federal ESA
136 was ruled warranted but precluded (USFWS 2012). In Oregon,
137 longfin smelt are managed under Oregon state rule (OAR 635-004-
138 0540 or OAR 635-994-0545) and targeted harvest off the coast is
139 prohibited. North of Oregon there are no other protections or
140 knowledge of long-term population status. Information about
141 genetic diversity and structure is vital for understanding the
142 coast-wide interactions among putative populations and for
143 making decisions about the legal protection and conservation
144 status of the longfin smelt in the SFE.

145 Here, we use genome-wide data to evaluate the population
146 structure of longfin smelt on two scales: a broad latitudinal
147 scale across 6 estuaries and Lake Washington, and fine scale
148 within the Fraser River Estuary and within the SFE. We also
149 examine genetic diversity, adaptive variation, and population
150 connectivity in the major longfin smelt estuaries along the
151 northeastern Pacific Coast. Our results capture patterns of
152 local adaptation and gene flow between longfin smelt populations
153 and highlight the need for further studies aiming to understand
154 the population dynamics of longfin smelt.

155 **Materials and Methods**

156 ***Sampling, Sequencing and RAD loci discovery***

157 We used DNA from individual longfin smelt collected between
158 2000 and 2015 from 11 locations along the west coast of North
159 America (Table 1 and Figure 1). Four of these collections were
160 from widely dispersed locations within the SFE from long-term or
161 targeted surveys: ALVS (Alviso Slough), CHPI (Chippis Island),
162 PETA (Petaluma River), and SUIB (Suisun Bay). Six collections
163 are from estuary locations further north: HUMB (Humboldt Bay,
164 California) and COLR (Columbia River estuary,
165 Oregon/Washington), PTLC and HRLC (Pitt Lake and Harrison Lake
166 which are tributary lakes in the Fraser River, British
167 Columbia), SKNA (Skeena River estuary), and YBAK (Yakutat Bay,
168 Alaska). The last collection is from the putatively landlocked

169 lake, Lake Washington, WA (LWSH). Samples consisted of an
170 individual fin clip stored in either EtOH or dried in coin
171 envelopes. DNA was extracted using the Qiagen DNEasy kit
172 according to the manufacturer's protocol.

173 To produce a high-quality genomic resource for longfin
174 smelt we sampled the genome using RAD sequencing (Miller et al.
175 2007; Baird et al. 2008). Paired-end 150 bp sequence reads were
176 generated for 190 individuals using the Sfb1 restriction enzyme
177 and the RAD protocol described in Ali et al. (2016). Sequences
178 were sorted into individuals as matching forward and reverse
179 fastq files using unique 8 bp barcodes. RAD-loci discovery and
180 extension was carried out using the custom procedure given in
181 (Sağlam et al. 2016) (for details see supplementary
182 information). This procedure resulted in 23,525 unique RAD-
183 contigs ranging from 300 to 800 bp and served as a de-novo
184 reference for all downstream analyses.

185 ***Alignment and filtering***

186 Sequence reads from individuals were aligned to the *de novo*
187 set of reference RAD contigs using the BWA-MEM algorithm (Li
188 2013). Outputted SAM files were transformed into BAMfiles and
189 indexed after sorting for proper pairs and removing PCR
190 duplicates in samtools (Li et al. 2009). Potential paralogous
191 RAD-loci were tagged using ngsParalogs
192 (<https://github.com/tplinderoth/ngsParalog>) and removed from

193 further analysis. Furthermore, we removed 47 individuals with
194 low read depth that resulted in fewer than 100,000 alignments
195 (Supplementary information, Table S1). This left us with a final
196 sample size of 123 individuals from 11 locations distributed
197 among six estuaries and Lake Washington (Table 1), which formed
198 the basis of all subsequent analyses.

199 ***SNP discovery and genotyping***

200 Due to low sequencing depth, we conducted all analyses
201 using the probabilistic framework implemented in ANGSD
202 (Korneliussen et al. 2014) and related software
203 (http://www.popgen.dk/software/index.php/Main_Page) which do not
204 require calling genotypes and are ideal for working with low
205 depth data (Korneliussen et al. 2013; Fumagalli et al. 2013).
206 Genotype likelihoods, minor allele frequencies (MAF) and
207 genotype probabilities were calculated in ANGSD. We discovered
208 SNPs by screening for sites with an MAF of over 0.05 and the
209 probability of the site not being polymorphic of less than 10^{-12} .
210 Sites were discarded if they did not meet a minimum quality
211 score of 20 and mapping quality of 10. Furthermore, we filtered
212 out any site that had an average per individual read depth below
213 6 and was not present in at least half of the individuals.

214 ***Genetic structure and differentiation***

215 We visualized genetic structure between collection sites
216 using principal component analysis (PCA). We first estimated the

217 genetic covariance matrix between individuals using PCAngsd
218 based on the calculated genotype likelihoods (Meisner and
219 Albrechtsen 2018). Principal component axes summarizing genetic
220 structure were then derived from this matrix using classic
221 eigenvalue decomposition in R version 3.6.1 (R Core Team 2013).

222 To determine shared ancestry between collection sites, we
223 conducted admixture analyses for different values of K. We used
224 K values ranging from 2-11. Individual admixture proportions
225 were estimated in NgsAdmix (Skotte et al. 2013) based on the
226 genotype likelihoods calculated in ANGSD. We ran NgsAdmix 15
227 times per K value, and used the likelihood values from each run
228 to calculate the most likely K with the method described in
229 Evanno et al. (2005).

230 Genetic differentiation between all collection sites was
231 summarized using pairwise F_{ST} statistics. Genome-wide F_{ST} values
232 were estimated in ANGSD by first calculating the joint-SFS
233 between each collection site and then calculating the global
234 weighted F_{ST} using the REALSFS module.

235 **Genetic diversity**

236 Genetic diversity was determined by calculating average
237 pairwise nucleotide differences (θ_{π}) based on the global site
238 frequency spectrum (SFS) derived separately for each collection
239 location (Supplementary information, Figure S1). We used a
240 folded SFS since we did not have a suitable outgroup for

241 determining ancestral states. A SFS for each population was
242 estimated in ANGSD (-doSaf 1) and later used as a prior to
243 calculate θ_π for each RAD-contig (-doThetas). θ_π values for each
244 RAD-contig were transformed into per-site values by dividing
245 each statistic by the length of the contig and genome-wide
246 values were obtained for each population by averaging across
247 RAD-contigs.

248 ***Contemporary migration***

249 We tested ongoing migration between coastal collection
250 sites with samples grouped according to genetic structure
251 analyses (SKNA, PTLC, HRLC, COLR, HUMB, and SFBY) using BayesAss
252 version 3 (Wilson and Rannala 2003), which uses multilocus
253 genotypes to determine immigrant ancestry of individuals within
254 the last three generations (for longfin smelt, generation time
255 is ~2 years). We called genotypes from polymorphic sites using
256 genotype probabilities previously estimated in ANGSD using a
257 posterior cut-off of 80% and outputted as a set of base pairs
258 (AA, AC, CC, etc.) along with the major and minor allele at each
259 site. To minimize the effect of missing genotypes and linkage we
260 filtered out any SNP that had more than 10% missing data (1 or 2
261 non-called genotypes per site) and then chose only a single SNP
262 per RAD contig. This resulted in a final data set of 1,653 SNPs
263 which were used in migration rate estimates. We performed three
264 independent runs with 40 million MCMC iterations each and a

265 sampling frequency of 4,000. After initial trial runs we settled
266 at delta values 0.80, 0.80, 0.80 for parameters 1, 3 and 4
267 (migration rates, allele frequencies and inbreeding
268 coefficients) which resulted in adequate mixing (40% - 50%). The
269 first 4,000,000 iterations of each run were discarded as burn-in
270 and after checking convergence of parameters using TRACER v. 1.6
271 (Rambaut et al. 2014) we combined all independent runs for final
272 parameter estimation.

273 ***Local selection scans***

274 To examine adaptive variation among sample locations, we
275 conducted a genome-wide selection scan as implemented in PCAngsd
276 (-selection 1). Based on the FastPCA model (Galinsky et al.
277 2016), PCAngsd performs a selection scan along all significant
278 PCs and identifies variants whose differentiation along top PCs
279 is significantly greater than the null distribution of genetic
280 drift. The advantage of this method is that it can account for
281 population structure and gene flow between subpopulations and
282 does not require a specific model for detecting signals of
283 genome-wide significance (Galinsky et al., 2016). The resulting
284 selection statistic is X^2 -distributed with 1 degree of freedom
285 and is used to identify variants under selection based on an
286 appropriate probability cutoff value. To minimize any
287 confounding effects due to inbreeding we filtered out SNPs that
288 showed significant deviations from HWE due to inbreeding (-

289 inbreedSites) based on the kinship matrix between all
290 individuals (-inbreed 3) as calculated in PCAngsd. This gave us
291 a final set of 13,296 SNPs which resulted in a genome-wide
292 significance threshold of 9.40×10^{-7} ($\alpha = 0.05$; 13,296 SNPS \times 4
293 principal components).

294 **Results**

295 Alignment of raw reads to the de novo reference resulted in
296 an average mapping success of 65%, and mean individual read
297 depth was around $\sim 5x - 6x$. Raw read counts, number of raw
298 alignments, and number of alignments after filtering for proper
299 pairs and removing PCR duplicates are given in supplementary
300 information, Table S1. Across all collection sites, we
301 discovered 5,864 polymorphic loci and 13,402 high probability
302 ($P < 10^{-12}$) SNPs with a minor allele frequency over 0.05.

303 **Genetic structure and differentiation**

304 Genetic structure for longfin smelt throughout their range
305 was high as sampling locations along the northeastern Pacific
306 coast were clearly separated from one another along the four
307 significant PC axes (Figure 2). PC1 indicated that YBAK is the
308 most differentiated (Figure 2A), followed by LWSH on PC2, and
309 the Fraser River sites, PTLC and HRLC, on PC3. PC3 also
310 separated COLR and HUMB from the four SFE locations. However, we
311 did not capture any separations among the four SFE locations
312 (ALVS, CHPI, PETA and SUIB), between the Fraser locations, nor

313 between COLR and HUMB along any of the significant PC axes
314 (Figure 2). PC 4 separated SKNA from all other locations.

315 Admixture analysis returned support for groupings of $K=3$
316 and $K=5$ based on Evanno's method (Supplementary information,
317 Figure S2). Admixture plots based on $K=3$ gave a similar
318 structure to that defined along PC1 and PC2 separating both YBAK
319 and the landlocked LWSH while indicating some amount of shared
320 ancestry between all other coastal sites (Figure 3). At $K=5$ the
321 analysis was able to separate both the Skeena River (SKNA) and
322 Fraser River locations (PTLC and HRLC) from COLR, HUMB, and the
323 four SFE locations (Figure 3). Similar to PCA we were not able
324 to capture any further structure between sites at higher K
325 values, and could not distinguish between the four SFE
326 locations, the Fraser locations, or COLR and HUMB
327 (Supplementary information, Figure S3).

328 Genetic differentiation based on genome-wide F_{ST} values
329 supported results obtained from structure analysis. Relatively
330 high genetic differentiation ($F_{ST} > 0.1$) was only observed for
331 pairwise comparisons featuring YBAK, SKNA and LWSH, while
332 pairwise F_{ST} values between all other collection sites were
333 generally low ($F_{ST} < 0.1$, Figure 4).

334 Due to low differentiation and multiple sampling sites in
335 the Fraser River and the SFE, we conducted separate PCA and
336 admixture analysis on just these samples to test for the

337 presences of fine scale population structure. Separate PCA and
338 admixture analysis did not reveal any fine scale structure
339 between SFE locations (Supplementary information, Figure S4)
340 however, we were able to capture significant structure within
341 the Fraser River (Supplementary information, Figure S5). Based
342 on genetic structure and differentiation we conclude that
343 longfin smelt within SFE constitute one single population
344 (hereafter called SFBY).

345 ***Contemporary migration***

346 Migration rates between locations (except LWSH) based on
347 all three BayesAss runs were consistent with each run and
348 converged onto the same parameter space (Supplementary
349 information, Figure S6), so we combined all three runs for final
350 parameter estimation using approximately 30,000 posterior
351 samples. With two exceptions, we observed relatively low
352 migration probabilities between populations (<0.025), with 95%
353 bayesian credible intervals (Highest probability densities, HPD)
354 overlapping with zero, indicating limited migration between most
355 locations (Figure 5; Supplementary information Table S2).
356 However, we did find significant evidence for contemporary
357 migration northward from SFBY to both HUMB and COLR with
358 probabilities of 0.196 (HPD: 0.119 - 0.276) and 0.200 (HPD:
359 0.129 - 0.274) migrants per generation respectively (Table S2).
360 However, migration probabilities in the reverse direction (south

361 from HUMB and COLR to SFBY) were low and not significant.
362 Similarly, we found significant migration from PTLC east and
363 upstream to HRLC (0.191, HPD: 0.109 - 0.273) but none in the
364 reverse direction (Figure 5; Table S2).

365 ***Genetic diversity and local selection scans***

366 Average θ_{π} was similar between all populations and ranged
367 from 0.001 to 0.003 (Figure 6). HRLC had the highest value,
368 followed by PTLC, COLR, and SFBY. To find evidence of adaptive
369 variation within populations we conducted selection scans to
370 find loci that show significant differentiation between
371 populations above that expected by genetic drift and accounted
372 for by structure. Selection scans conducted on the four
373 significant PC axes did not capture any sites that reached
374 genome-wide-significance along PC1 or PC2, but captured 21 and
375 23 sites along PC3 and PC4, respectively (Supplementary
376 information, Table S3 and S4). To understand changes in allele
377 frequencies at these sites we plotted the frequency of the
378 reference allele (taken as the minor allele in the entire
379 sample) at these sites for all populations. Along PC3 (which
380 separates the Fraser River locations PTLC and HRLC), the
381 frequency of the reference allele at sites with positive
382 signatures of selection were higher in the Fraser locations, and
383 some sites were swept to near fixation (Figure 7A). Likewise,
384 along PC4 (which separates SKNA), the frequency of the reference

385 allele at sites showing signs of selection were higher in the
386 SKNA population, to the point where the minor allele becomes the
387 major allele (Figure 7B).

388 **Discussion**

389 This is the first range-wide genomic analysis of the
390 population genetic structure of longfin smelt, a poorly studied
391 forage fish. Our results indicate relatively high genetic
392 structure between the major estuaries sampled, fine-scale
393 structure between Pitt and Harrison Lakes in the Fraser River
394 Estuary, and very low levels of structure within the San
395 Francisco Estuary, by far the largest estuary in the region.
396 Genetic structure was more pronounced between northern estuaries
397 whereas we detected appreciable amounts of shared ancestry and
398 ongoing gene flow in southern estuaries. Migration analyses
399 showed unidirectional migration from Pitt to Harrison Lake in
400 the Fraser River system and strong unidirectional northward
401 migration out of the SFE to nearby estuaries of Humboldt Bay and
402 Columbia River. Furthermore, we detected clear signatures of
403 local adaptation samples taken from the Fraser River system and
404 the Skeena River Estuary. Taken together, our results identify
405 broad patterns of genetic diversity in longfin smelt along the
406 North American west coast shaped by co-ancestry, unidirectional
407 migration and local adaptation and hint at variability in
408 longfin smelt behavior during the ocean phase of their lives.

409 ***Genetic structure, gene flow, and local adaptation in northern***
410 ***estuaries***

411 Overall, the observed patterns of differentiation and low
412 levels of gene flow among northern estuaries indicate low levels
413 of movement between estuaries and/or a lack of regular long-
414 distance migration, suggesting a that genetic drift had a strong
415 influence (e.g., bottlenecks and founder events). The northern-
416 most sampled population in Yakutat Bay was the most
417 differentiated with the lowest genetic diversity, followed by
418 the landlocked Lake Washington, Skeena River, and the Fraser
419 River locations Pitt and Harrison Lakes. Yakutat Bay is a
420 relatively small bay (29km wide) located at the southwest of the
421 Disenchantment Bay waters which captures freshwater runoff from
422 three glaciers, including the Hubbard Glacier. Some
423 geomorphological patterns influencing Yakutat Bay include the
424 advance and retreat cycles of Hubbard Glacier that create
425 Russell Lake via temporary ice dams that then burst sending cold
426 freshwater into Yakutat Bay, as well as high seismic activity
427 (Barclay et al. 2001). This combined with the relatively recent
428 Pleistocene retreat of glaciers opening up the Alaska coast may
429 contribute to the high differentiation of the Yakutat Bay
430 population (Mann and Hamilton 1995).

431 The provenance of longfin smelt in Lake Washington is
432 somewhat mysterious. Lake Washington is a shallow (~74 m) ribbon

433 lake formed by glaciation processes. The original outlet was the
434 Black River, which evidently had a salmon run and was somewhat
435 navigable to Puget sound depending on tides and conditions
436 (Purvis 1934, Chrzastowski 1983). In 1916 a ship canal was built
437 forming the current outlet and causing the lake level to drop 3
438 meters and dry the Black River. The ship canal possesses locks
439 with fish ladders, and the canal is successfully used by salmon
440 to access Lake Washington, but it is unknown whether longfin
441 smelt can use either the ladders or navigate the locks. The Lake
442 Washington population is highly differentiated; it is possible
443 that it is remnant from before the locks were built, or perhaps
444 they were able to colonize Lake Washington later.

445 Though long-distance movement of longfin smelt between
446 estuaries seems to be broadly limited in northern estuaries,
447 migration analyses indicated short-distance, unidirectional gene
448 flow from Pitt to Harrison Lake. The findings from Pitt and
449 Harrison lakes are supported by their hydrology as they are
450 fjord lakes that connect to the Fraser River. Harrison Lake is
451 60 km to the east and upstream of Pitt Lake and has a greater
452 connection to the Fraser River due to higher average flows (446
453 m^3/s) draining a substantially larger watershed ($8,324 \text{ km}^2$)
454 compared to the Pitt watershed ($1,660 \text{ km}^2$ and $186 \text{ m}^3/\text{s}$). In
455 addition, Harrison River was dredged to make it navigable to
456 boats entering Harrison Lake. In contrast, Pitt Lake is one of

457 the largest tidal fjord lakes in the world. Very strong tides
458 create a negative delta that is exposed at low tide, cutting it
459 off from the Pitt River which empties into the Fraser River
460 during part of each day (Ages and Woollard 1976; Fraser River
461 Action Plan 1999). Interestingly, reports on longfin smelt have
462 suggested that Pitt and Harrison Lakes are land-locked (Dryfoos
463 1965, USFWS 2012), though our migration analysis suggests that
464 they are not.

465 There is evidence that local adaptation might have played a
466 role in the Skeena River and Pitt and Harrison lakes.

467 Examination of local adaptation using genome-wide selection
468 scans along significant principal component axes did not reveal
469 any regions reaching genome-wide significance along the two main
470 axes (PC1 and 2) suggesting that differentiation between
471 estuaries is mainly caused by neutral evolution. However,
472 selection scans along PC3 and 4 revealed a large number of sites
473 that reached genome wide significance in Pitt and Harrison Lakes
474 (N=21 sites along PC3) and the Skeena River (N=23 sites along
475 PC4) indicating local adaptation. In all three systems the
476 frequency of the outlier allele had risen to over 60% (going to
477 total fixation in some) indicating strong selective pressures on
478 these genomic regions. The colder and more sterile freshwater
479 habitats of Pitt and Harrison lakes in the Fraser River drainage
480 may provide very different challenges than the mostly anadromous

481 life histories of other locations (Dryfoos 1965) and could be
482 the source of selection in this system. Interestingly, Dryfoos
483 (1965) described Harrison Lake longfin smelt as "stunted" and
484 subsequent descriptions have described populations in both Pitt
485 and Harrison lakes as "pygmy" longfin smelt (e.g., Roberge et
486 al. 2002) although Dryfoos (1965) attributed their small size to
487 the colder low-food environment. Future studies may determine if
488 this phenotype has a genetic basis. In the Skeena River estuary,
489 the limited knowledge of longfin smelt presents no obvious
490 sources of strong selection, though the estuary is one of the
491 more complex, with multiple channels, and several islands that
492 both dissect the estuary and enclose it.

493 Further studies are needed to confirm patterns of local
494 adaptation especially if such patterns exist in other estuaries
495 as well. Since our sample sizes were fairly small and we
496 interrogated a very small portion of the genome (12,695 loci)
497 there is a good chance that we were only able to capture loci
498 under extreme selection pressures linked to traits with high
499 heritability. In addition, this method of selection scanning
500 cannot detect all types of adaptive variation, such as rare
501 variants, and cannot indicate how or when selection occurred,
502 only that it occurred. Therefore, absence of local adaptation in
503 other estuaries might result from limitations of the current
504 study and not necessarily the absence of selection pressures.

505 Greater sampling coverage may uncover additional adaptive
506 variation and present further avenues for research.

507 ***Reduced genetic structure and northward migration in southern***
508 ***estuaries***

509 In contrast to northern estuaries, the locations sampled at
510 the southern end of our sampling range showed reduced genetic
511 structure and differentiation, and appreciable amounts of shared
512 ancestry, especially between the SFE, Humboldt Bay and Columbia
513 River estuaries. Though this pattern may be due to recent
514 divergence, it is also supported by the detection of relatively
515 high levels of ongoing northward gene flow from the SFE towards
516 Humboldt Bay and the Columbia River estuary.

517 Evidence of contemporary gene flow northward out of the SFE
518 raises a number of questions, the first of which is: what is the
519 process by which genetic influence is moving north? Our findings
520 support contemporary gene flow out of the SFE to estuaries that
521 are a considerable distance north: Humboldt Bay is 418 and the
522 Columbia River is 1,030 sea km north of the SFE. There are
523 several smaller estuaries north of the SFE where longfin smelt
524 have been found, including the Russian, Eel, and Klamath rivers
525 in California (Garwood 2017), and in Coos Bay, Tillamook Bay,
526 and the Yaquina Bay estuaries in Oregon (Bottom and Forsberg
527 1978); it is possible that longfin "estuary hop" in a northern
528 direction rather than undergo long distance migrations. In the

529 Columbia River Estuary, an understanding of the current
530 distribution and abundance trends of longfin smelt is limited.
531 Humboldt Bay is a coastal lagoon, and the second largest estuary
532 on the California coast. Due to low freshwater inputs, Humboldt
533 Bay itself likely does not support (and may have never
534 supported) a large population, but in recent years small
535 collects have occurred regularly in bay tributaries (Garwood
536 2017). Alternatively, a local population in the larger Eel River
537 estuary located just south may support the common occurrence of
538 longfin smelt in Humboldt Bay (Pequegnat and Butler 1982;
539 Garwood 2017). Nevertheless, longfin smelt were considered
540 common in Humboldt Bay in the late 1970s, though the population
541 has likely declined at least since then, if not over the last
542 150 years when habitat alteration began (Garwood 2017; USFWS
543 1994). Given its small size and human alteration, it is not
544 surprising that longfin smelt captured in Humboldt Bay today
545 would exhibit genetic influence from other locations.

546 The second major question regarding gene flow is: why is
547 gene flow only detected in a northern direction from the SFE?
548 Longfin smelt have been found offshore of the SFE during winter
549 months (Garwood 2017), when the near-shore Davidson current is
550 flowing northward, and providing conditions for modest northward
551 movement annually. In extreme years associated with strong El Niños, nearshore
552 organisms can be transported multiple hundreds of kilometers northward as seen in range

553 extensions of demersal species with pelagic egg and larval stages (e.g., California tonguefish;
554 Huyer and Smith 1985, Pearcy et al. 1985). In contrast, the California current
555 which generally flows southward and far offshore during spring
556 through fall, is highly complex, and varies in direction and
557 strength at different depths, distances from shore, and times of
558 the year. Thus, the California current may be less likely to
559 assist southward gene flow. Moreover upwelling areas in the
560 Pacific are notoriously variable and include seasonal, annual,
561 and decadal scale changes in oceanographic patterns (Checkley
562 and Barth 2009). These upwelling changes may cause punctuated
563 episodes of gene flow, and have been linked to abundance and
564 range shift variation along the coast of California in other
565 pelagic forage fish that commonly range well offshore: Pacific sardine
566 (*Sardinops sadax caerulea*) and northern anchovy (*Engraulis*
567 *mordax*) along the coast of California (Lecomte et al. 2004). We
568 cannot determine if there is gene flow due to spawning in
569 smaller estuaries in between the estuaries we sampled, or if
570 longfin smelt actually migrate long distances, or what scale
571 gene flow occurs on.

572 ***Lack of fine-scale structure in the San Francisco Estuary***

573 Due to the large size and complexity of the SFE, it is
574 reasonable to expect fine scale genetic structuring of longfin
575 smelt taken from different sampling locations within the SFE,
576 especially if fish exhibit fidelity to unique spawning and

577 rearing habitats. However, our analysis shows minimal structure,
578 strongly suggesting that the whole estuary is a single genetic
579 unit. Though long-term data sets do not directly target longfin
580 smelt, they have been leveraged by several studies to understand
581 the ecology of longfin smelt in the SFE. For example several
582 studies link abundance to various conditions in the estuary,
583 such as freshwater flows (Stevens and Miller 1983; Jassby et al.
584 1995; Kimmerer 2002b). Rosenfeld and Baxter (2007) synthesized
585 three long term data sets to investigate patterns in longfin
586 smelt abundance and presence throughout their life cycle. Based
587 on declines in detections (and abundance) of the second year of
588 life to nearly zero in late summer and fall followed by
589 increasing detections in winter, Rosenfeld and Baxter (2007)
590 concluded that most longfin smelt must spend the months (June-
591 August) in the coastal ocean. However, given their 2- to 3-year
592 lifespan, they can reside at sea for more than a year
593 (Rosenfeld and Baxter 2007). A couple age classes have been collected in the
594 coastal ocean in the past (CDFG 2009) indicating the potential for coastal rearing beginning their
595 first summer and continuing until returning to the estuary to spawn at the end of their second fall
596 or the following winter (2-year lifespan), or an entire year later (3-year lifespan). They
597 also suggested that anadromy may be an alternative life history
598 strategy, with some fish staying in the estuary as residents.
599 Such residents likely rear in cool marine waters in the lower
600 estuary during the late-summer and fall of their second year,

601 making salinity distinctions between residents and coastal
602 migrants difficult. Otolith microchemistry, using isotopes of
603 strontium and oxygen further support these observations,
604 indicating that individual Longfin Smelt within the SFE can
605 exhibit a variety of spawning, rearing, and adult migratory
606 behaviors (Lewis et al. unpublished).

607 ***Management implications for SFE***

608 Results of this work have several key implications that can
609 help guide conservation efforts for longfin smelt in the SFE.
610 The presence of unidirectional northward migration from the SFE
611 to nearby estuaries further emphasizes the importance of the SFE
612 for Longfin Smelt populations throughout their southern range,
613 and simultaneously raises concern about the vulnerability of the
614 SFE population to long-term extirpation, because the population
615 is apparently not supported by a nearby source population.
616 Together, these results suggest that that protection of the SFE
617 population from extinction is likely warranted.

618 The lack of significant population structure within the SFE
619 was somewhat surprising given its geographic size; however, if
620 longfin smelt do not exhibit unique spawning habitats with high
621 site fidelity, then this result might be expected. Given that
622 the majority of low-salinity spawning and rearing habitat
623 appears to be located in the upper estuary ("Delta"; Moyle
624 2002), it is possible that nearly all longfin smelt spawn within

625 the same geographic region and later disperse throughout the
626 estuary as late-stage larvae and juveniles. Though longfin smelt
627 have been detected spawning and rearing in and near tributaries
628 located throughout the SFE, especially during years of high
629 freshwater outflow (Lewis et al. 2020), the lack of genetic
630 structure may further emphasize the importance of SFE's Delta as
631 critical spawning habitat. The Delta is highly impacted by water
632 diversions, invasive species, pesticides, and many other
633 anthropogenic factors that may be addressed by well-informed
634 policy decisions (Sommer et al. 2007, Baxter et al. 2010).
635 However, before strong conclusions can be made regarding fine-
636 scale population structure and subsequent policies drafted,
637 additional studies are needed examining genomic variation using
638 larger sample sizes collected over longer time scales.

639 **Conclusion**

640 The significant, broad-scale genetic structure and
641 diversity of longfin smelt in the northeastern Pacific Ocean is
642 likely shaped by a multitude of factors including strong genetic
643 drift, local adaptation, shared ancestry, and gene flow. Given
644 the variety of coastal and freshwater habitats used by longfin
645 smelt throughout its range, and the reported diversity of life
646 history types, it is unsurprising to find evidence suggesting
647 that different dynamics govern the genetic structure of separate
648 populations. The most interesting results of this work were the

649 observation of genetic isolation across broad latitudinal scales
650 and the findings of unidirectional, long-distance northward gene
651 flow out of the SFE into other southern estuaries. Though this
652 latter observation challenges previous assumptions of a limited
653 migratory capacity for this species, it is in agreement with the
654 its broad coastal distribution of the species and the complex
655 nearshore coastal oceanography of the region. Further genomic
656 work using larger sample sizes collected over more locations and
657 longer time scales may provide additional insights regarding the
658 genetic structure of this population. We also suggest combining
659 genomics with other tools such as otolith chemistry to shed
660 light on variation in migratory behaviors.

661

662 **Acknowledgements**

663 The authors would like to thank Colin Grant (US Fish and
664 Wildlife Service) who made many of the initial inquiries
665 soliciting longfin smelt samples and Josh Israel (US Bureau of
666 Reclamation) who performed initial genetic analyses. The authors
667 would also like to thank the following individuals and agencies
668 for help collecting samples: Hobbslab and staff - UC Davis; Matt
669 Dekar and staff - US Fish and Wildlife Service; Kathy Hieb,
670 Jennifer Giannetta, Jeremiah Bautista, Aaron Ng, Rebecca
671 Garwood, James Ray, Justin Garwood, and Mike Wallace -
672 California Department of Fish and Wildlife; Laurie Lloyd, Olaf

673 Langness - Washington Department of Fish and Wildlife, Thomas
674 Quinn, University of Washington; Eric Taylor - University of
675 British Columbia, John Kelson - Fish Biologist/Consultant,
676 Smithers, British Columbia; Mayumi Arimitsu - US Geological
677 Survey, Alaska Science Center. Finally, the authors would also
678 like to thank a number of internal reviewers, Genomic Variation
679 Lab Personnel, and other agency biologists who helped this
680 project, including Matthew Campbell, Daphne Gille, Levi Lewis,
681 Andrea Schreier, Josh Israel, and Alisha Goodbla. Funding for
682 this project was generously provided by the California
683 Department of Water Resources, Agreement Number 4600011196.

684

685 **References**

- 686 Abookire, A.A., and Piatt, J.F. 2005. Oceanographic conditions
687 structure forage fishes into lipid-rich and lipid-poor
688 communities in lower Cook Inlet, Alaska, USA. *Mar. Ecol.*
689 *Prog. Ser.* 287: 229-240. doi:10.3354/meps287229.
- 690 Abookire, A.A., Piatt, J.F., and Robards, M.D. 2000. Nearshore
691 Fish Distributions in an Alaskan Estuary in Relation to
692 Stratification, Temperature and Salinity. *Estuar. Coast.*
693 *Shelf Sci.* 51(1): 45-59. doi:10.1006/ecss.1999.0615.
- 694 Ages, A., Woollard, A. 1976. *The Tides in the Fraser Estuary.*
695 Institute of Ocean Sciences, Patricia Bay, Victoria, BC,
696 Canada. pp. 7, 11.

- 697 Ali, O.A., O'Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G.,
698 Jeffres, C., and Miller, M.R. 2016. RAD Capture (Rapture):
699 Flexible and Efficient Sequence-Based Genotyping. *Genetics*,
700 202(2): 389-400.
- 701 Arimitsu, M. (2016). The influence of glaciers on coastal marine
702 ecosystems in the Gulf of Alaska. PhD Dissertation,
703 University of Alaska, Fairbanks, Alaska. 154 pages.
- 704 Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver,
705 A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson,
706 E.A. 2008. Rapid SNP discovery and genetic mapping using
707 sequenced RAD markers. *PLoS One*, 3(10): e3376.
- 708 Barclay, D. J., Calkin, P. E., Wiles, G. C. 2001. Holocene
709 history of Hubbard Glacier in Yakutat Bay and Russell
710 Fiord, southern Alaska. *GSA Bulletin*, 113(3): 388-402. DOI:
711 10.1130/0016-7606(2001)113<0388:HHOHGI>2.0.CO;2
- 712 Baxter, R., Breuer, R., Brown, L., Chotkowski, M., Feyrer, F.,
713 Gingras, M., Herbold, B., Mueller-Solger, A., Nobriga, M.,
714 Sommer, T., and Souza, K. 2008. Pelagic Organism Decline
715 Progress Report: 2007 Synthesis of Results. Interagency
716 Ecological Program for the Sacramento-San Joaquin Estuary
717 Technical Report. 78 pages.
718 [https://water.ca.gov/LegacyFiles/iep/docs/pod/synthesis_rep
719 ort_031408.pdf](https://water.ca.gov/LegacyFiles/iep/docs/pod/synthesis_report_031408.pdf)
- 720 Baxter, R., R. Breuer, L. Brown, L. Conrad, F. Feyrer, S. Fong,

- 721 K. Gehrts, L. Grimaldo, B. Herbold, P. Hrodey, A. Mueller-
722 Solger, T. Sommer and K. Souza (2010). Interagency
723 Ecological Program 2010 Pelagic organism decline work plan
724 and synthesis of results through August 2010, Interagency
725 Ecological Program for the San Francisco Estuary: 125
726 pages.
- 727 Bottom, D.L., and Jones, K.K. 1990. Species composition,
728 distribution, and invertebrate prey of fish assemblages in
729 the Columbia River Estuary. Prog. Oceanogr. 25(1-4): 243-
730 270. doi:10.1016/0079-6611(90)90009-q.
- 731 Bottom, D. and B. Forsberg (1978). The Fishes of Tillamook Bay.
732 Federal Aid Progress Reports Fisheries. Portland Oregon,
733 Oregon Dept of Fish and Wildlife: 56 pages.
- 734 California Fish and Game Commission. 2009. California fish and
735 game commission notice of findings: longfin smelt
736 (*Spirinchus thaleichthys*). Meeting: Woodland California,
737 May 29, 2009.
- 738 California Department of Fish and Game (CDFG) (2009). Report to
739 the Fish and Game Commission: A status review of the
740 longfin smelt (*Spirinchus thaleichthys*) in California.
741 Report to the Fish and Game Commission: 46 pages.
- 742 Checkley, D.M. Jr., and Barth, J.A. 2009. Patterns and processes
743 in the California Current system. Progress in Oceanography,
744 83(2009) 49-64. doi:10.1016/j.pocean.2009.07.028

- 745 Chigbu, P. 2000. Population Biology of Longfin Smelt and Aspects
746 of the Ecology of Other Major Planktivorous Fishes in Lake
747 Washington. J. Freshwater Ecol. 15(4): 543-557.
748 doi:10.1080/02705060.2000.9663777.
- 749 Chigbu, P., and Sibley, T.H. 1998a. Feeding ecology of longfin
750 smelt (*Spirinchus thaleichthys* Ayres) in Lake Washington.
751 Fish. Res. 38(2): 109-119. doi:10.1016/s0165-7836(98)00156-
752 8.
- 753 Chigbu, P., and Sibley, T.H. 1998b. Predation by longfin smelt
754 (*Spirinchus thaleichthys*) on the mysid *Neomysis mercedis* in
755 Lake Washington. Freshwater Biol. 40(2): 295-304.
756 doi:10.1046/j.1365-2427.1998.00354.x.
- 757 Chrzastowski, M. 1983. Historical Changes to the Lake Washington
758 Ship Canal, King County, Washington. U.S. Geological
759 Survey, Water Resources Investigation Open-File Report 81-
760 1182. <http://www.jstor.org/stable/40475640>
- 761 Donnelly, R. F., B. S. Miller, J. H. Stadler, L. Christensen, K.
762 Larsen and P. A. Dinnel (1988). Puget Sound dredge disposal
763 analysis (PSDDA) phase II disposal site bottomfish
764 investigations, Final Report. Seattle, WA, Fisheries
765 Research Institute, School of Fisheries, University of
766 Washington: 149 pages.
- 767 Dryfoos, R.L. 1965. The Life History and Ecology of the Longfin
768 Smelt in Lake Washington. Ph.D. thesis, University of

- 769 Washington, Seattle, WA.
- 770 Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the
771 number of clusters of individuals using the software
772 structure: a simulation study. *Mol. Ecol.* 14(8):2611-20.
773 doi:10.1111/j.1365-294x.2005.02553.x.
- 774 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C.,
775 and Foll, M. 2013. Robust demographic inference from
776 genomic and SNP data. *PLoS Genet.* 9(10): e1003905.
- 777 Excoffier, L., and Foll, M. 2011. fastsimcoal: a continuous-time
778 coalescent simulator of genomic diversity under arbitrarily
779 complex evolutionary scenarios. *Bioinformatics*, 27(9):
780 1332-1334. doi: 10.1093/bioinformatics/btr124.
- 781 Federal Register. 2008. Petition to List the San Francisco Bay-
782 Delta Population of the Longfin Smelt (*Spirinchus*
783 *thaleichthys*) as Endangered. U.S. Fish and Wildlife
784 Service, Division of Endangered Species, Sacramento,
785 California. Tuesday, May 6, 2008. Available on the Internet
786 at: <http://edocket.access.gpo.gov/2008/pdf/E8-9835.pdf>.
- 787 Fisheries and Oceans Canada Habitat and Enhancement Branch.
788 1999. Lower Fraser Valley Streams Strategic Review Lower
789 Fraser Valley Stream Review, Vol. 1 Fraser River Action
790 Plan. 360 - 555 W. Hastings St. Vancouver, British Columbia
791 V6B 5G3 1999. 545 pp. Available on the Internet at:
792 <https://waves-vagues.dfo-mpo.gc.ca/Library/240006.pdf>

- 793 Fumagalli, M., Vieira, F.G., Korneliussen, T.S., Linderoth, T.,
794 Huerta-Sánchez, E., Albrechtsen, A., and Nielsen, R. 2013.
795 Quantifying population genetic differentiation from next-
796 generation sequencing data. *Genetics*, 195(3): 979–992.
797 doi:10.1534/genetics.113.154740/-/DC1.
- 798 Galinsky, K.J., Bhatia, G., Loh, P.-R., Georgiev, S., Mukherjee,
799 S., Patterson, N.J., and Price, A.L. 2016. Fast Principal-
800 Component Analysis Reveals Convergent Evolution of ADH1B in
801 Europe and East Asia. *Am. J. Hum. Genet.* 98(3): 456–472.
802 doi: 10.1016/j.ajhg.2015.12.022.
- 803 Garwood, R.S. 2017. Historic and contemporary distribution of
804 Longfin Smelt (*Spirinchus thaleichthys*) along the
805 California coast. *California Fish and Game* 103(3): 96–117;
806 2017.
- 807 Greene, C., Kuehne, L., Rice, C., Fresh, K., and Penttila, D.
808 2015. Forty years of change in forage fish and jellyfish
809 abundance across greater Puget Sound, Washington (USA):
810 anthropogenic and climate associations. *Mar. Ecol. Prog.*
811 *Ser.* 525:153–170. doi:10.3354/meps11251.
- 812 Grimaldo, L., Feyrer, F., Burns, J., and Maniscalco, D. 2017.
813 Sampling Uncharted Waters: Examining Rearing Habitat of
814 Larval Longfin Smelt (*Spirinchus thaleichthys*) in the Upper
815 San Francisco Estuary. *Estuar. and Coast.* 40(6): 1771–1784.
816 <https://doi.org/10.1007/s12237-017-0255-9>.

- 817 Gustafson, R.G., Ford, M.J., Teel, D., and Drake, J.S. 2010.
818 Status review of eulachon (*Thaleichthys pacificus*) in
819 Washington, Oregon, and California. U.S. Dept. Commerce.
820 NOAA Tech. Memo. NMFS-NWFSC-105, 360 p.
- 821 Gustafson, R.G., Ford, M.J., Adams, P.B., Drake, J.S., Emmett,
822 R.L., Fresh, K.L., Rowse, M., Spangler, E.A.K., Spangler,
823 R.E., Teel, D.J., and Wilson, M.T. 2012. Conservation
824 status of eulachon in the California Current. Fish. Fish.
825 13(2): 121-138. doi:10.1111/j.1467-2979.2011.00418.x.
- 826 Hampe, A., and Petit, R.J. 2005. Conserving biodiversity under
827 climate change: the rear edge matters. Ecol. Lett. 8(5):
828 461-467. doi: 10.1111/j.1461-0248.2005.00739.x
- 829 Hart, J. L., and J. L. McHugh. 1944. The smelts (Osmeridae) of
830 British Columbia. Bull. Fish. Res. Bd. Canada (Ottawa)
831 64:1-27.
- 832 Hay, D. E. and McCarter, P. B. 2000. Status of the eulachon
833 *Thaleichthys pacificus* in Canada. Department of Fisheries
834 and Oceans Canada, Canadian Stock Assessment Secretariat,
835 Research Document 2000-145. Ottawa. 92 p.
- 836 Hobbs, J.A., University of California, Davis, Moyle, P.B.,
837 Fangué, N., Connon, R.E., University of California, Davis,
838 University of California, Davis, University of California,
839 Davis. 2017. Is Extinction Inevitable for Delta Smelt and
840 Longfin Smelt? An Opinion and Recommendations for Recovery.

- 841 SFEWS. 15(2):Article 2. doi:10.15447/sfews.2017v15iss2art2.
- 842 Huyer, A. and R. L. Smith (1985). "The signature of El Nino off
843 Oregon, 1982-1983." Journal of Geophysical Research **90**(C4):
844 7133-7142.
- 845 Jeffries, K. M., Connon, R. E., Davis, B. E., Komoroske, L. M.,
846 Britton, M. T., Sommer, T., Todgham, A. E., and Fangue,
847 N.A. 2016. Effects of high temperatures on threatened
848 estuarine fishes during periods of extreme drought. *J. Exp.*
849 *Biol.* 219:1705-1716. doi:10.1242/jeb.134528.
- 850 Jones, T.M., Bennett, A.L., and Hamon T.R. 2005. Baseline
851 inventory of freshwater fishes of the Southwest Alaska
852 Inventory and monitoring network: Alagnak Wild River,
853 Aniakchak National Monument and Preserve, Katmai National
854 Park and Preserve, Kenai Fjords National Park and Lake
855 Clark National Park and Preserve. National Park Service,
856 Anchorage, AK. 98 pages available at:
857 <https://www.arlis.org/docs/vol1/71090809.pdf>
- 858 Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. 2014.
859 ANGSD: Analysis of Next Generation Sequencing Data. *BMC*
860 *Bioinformatics*, 15: 356. doi:10.1186/s12859-014-0356-4
- 861 Korneliussen, T.S., Moltke, I., Albrechtsen, A., and Nielsen, R.
862 2013. Calculation of Tajima's D and other neutrality test
863 statistics from low depth next-generation sequencing data.
864 *BMC Bioinformatics*, 14: 289. doi:10.1186/1471-2105-14-289

- 865 Lecomte, F., Grant, W.S., Dodson, J.J., Rodriguez-Sanchez, R.,
866 and Bowen, B.W. 2004. Living with uncertainty: genetic
867 imprints of climate shifts in East Pacific anchovy
868 (*Engraulis mordax*) and sardine (*Sardinops sagax*). *Mol.*
869 *Ecol.* 13(8): 2169–2182. doi: 10.1111/j.1365-
870 294X.2004.02229.x
- 871 Lenfest Forage Fish Task Force. 2012. Little Fish Big Impact:
872 Managing a Crucial Link in Ocean Food Webs: a Report from
873 the Lenfest Forage Fish Task Force. Available at
874 [https://www.oceanconservationscience.org/foragefish/files/L](https://www.oceanconservationscience.org/foragefish/files/Little%20Fish,%20Big%20Impact.pdf)
875 [ittle%20Fish,%20Big%20Impact.pdf](https://www.oceanconservationscience.org/foragefish/files/Little%20Fish,%20Big%20Impact.pdf).
- 876 Lewis, L.S., Willmes, M., Barros, A., Crain, P.K., and Hobbs,
877 J.A. 2019. Newly discovered spawning and recruitment of
878 threatened Longfin Smelt in restored and underexplored
879 tidal wetlands. *Ecology*, 00(00): e02868.
880 doi:10.1002/ecy.2868
- 881 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J.,
882 Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000
883 Genome Project Data Processing Subgroup. 2009. The Sequence
884 Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):
885 2078–2079. doi:10.1093/bioinformatics/btp352.
- 886 Litz, M.N.C., Emmett, R.L., Bentley, P.J., Claiborne, A.M., and
887 Barceló, C. 2014. Biotic and abiotic factors influencing
888 forage fish and pelagic nekton community in the Columbia

- 889 River plume (USA) throughout the upwelling season 1999–
890 2009. *ICES J. Mar. Sci.* 71(1): 5–18.
891 doi:10.1093/icesjms/fst082.
- 892 Mac Nally, R., Thomson, J.R., Kimmerer, W.J., Feyrer, F.,
893 Newman, K.B., Sih, A., Bennett, W.A., Brown, L., Fleishman,
894 E., Culberson, S.D., and Castillo, G. 2010. Analysis of
895 pelagic species decline in the upper San Francisco Estuary
896 using multivariate autoregressive modeling (MAR). *Ecol.*
897 *Appl.* 20(5): 1417–1430. doi: 10.1890/09-1724.1.
- 898 Mann, D.H., and Hamilton, T.D., 1995. Late Pleistocene and
899 Holocene paleoenvironments of the North Pacific coast:
900 *Quaternary Science Reviews*, 14:449–471.
- 901 Maunder, M.N., Deriso, R.B., and Hanson, C.H. 2015. Use of
902 state-space population dynamics models in hypothesis
903 testing: advantages over simple log-linear regressions for
904 modeling survival, illustrated with application to longfin
905 smelt (*Spirinchus thaleichthys*). *Fish. Res.* 164: 102–111.
906 doi:10.1016/j.fishres.2014.10.017.
- 907 Meisner, J., and Albrechtsen, A. 2018. Inferring Population
908 Structure and Admixture Proportions in Low-Depth NGS Data.
909 *Genetics*, 210(2): 719–731. doi:10.1534/genetics.118.301336.
- 910 Miller, M.R., Atwood, T.S., Eames, B.F., Eberhart, J.K., Yan,
911 Y.-L., Postlethwait, J.H., and Johnson, E.A. 2007. RAD
912 marker microarrays enable rapid mapping of zebrafish

- 913 mutations. *Genome Biol.* 8(6): R105. doi:10.1186/gb-2007-8-
914 6-r105.
- 915 Morgan, S. G. (2014). "Behaviorally mediated larval transport in
916 upwelling systems." *Advances in Oceanography* 2104(Article
917 ID 364214): 17 pages.
- 918 Morrow, J.E. 1980. *The freshwater fishes of Alaska*. Alaska
919 Northwest Publishing Co., Anchorage, AK. 248 pp.
- 920 Moulton, L.L. 1974. Abundance, Growth, and Spawning of the
921 Longfin Smelt in Lake Washington. *Trans. Am. Fish.*
922 *Soc.*103(1): 46-52. doi:2.0.co;2">10.1577/1548-
923 8659(1974)103<46:agasot>2.0.co;2.
- 924 Moyle, P.B. 2002. *Inland Fishes of California*. University of
925 California Press.
- 926 Nobriga, M.L., and Rosenfield, J.A. 2016. Population Dynamics of
927 an Estuarine Forage Fish: Disaggregating Forces Driving
928 Long-Term Decline of Longfin Smelt in California's San
929 Francisco Estuary. *Trans. Am. Fish. Soc.* 145(1): 44-58.
930 doi:10.1080/00028487.2015.1100136.
- 931 Pearcy, W., R. Fisher, R. Brodeur and S. Johnson (1985). Effects
932 of the 1983 El Nino on coastal nekton off Oregon and
933 Washington. Pages 188-204 in W. S. Wooster and D. L.
934 Fluharty(eds). El Nino North -- Nino effects in the eastern
935 subarctic Pacific Ocean. Washington Sea Grant Program,
936 Seattle, WA.

- 937 Pequegnat, J.E., and Butler, J.H. 1982. The biological
938 oceanography of Humboldt Bay. Pages 39-51 in C. Toole and
939 C. Diebel, eds. Proceedings of the Humboldt Bay Symposium.
940 Center for Community Development, Humboldt State
941 University, Arcata, Calif.
- 942 Penttila, D. (2007). Marine forage fishes in Puget Sound. Puget
943 Sound Nearshore Partnership Report No. 2007-03. Seattle,
944 WA, Seattle District, U.W Army Corps of Engineers: 23
945 pages.
- 946 Pikitch, E.K., Rountos, K.J., Essington, T.E., Santora, C.,
947 Pauly, D., Watson, R., Sumaila, U.R., Dee Boersma, P.,
948 Boyd, I.L., Conover, D.O., Cury, P., Heppell, S.S., Houde,
949 E.D., Mangel, M., Plagányi, É., Sainsbury, K., Steneck,
950 R.S., Geers, T.M., Gownaris, N., and Munch, S.B. 2014. The
951 global contribution of forage fish to marine fisheries and
952 ecosystems. *Fish Fish.* 15(1): 43-64. doi:10.1111/faf.12004.
- 953 R Core Team (2013). R: A language and environment for
954 statistical computing. R Foundation for Statistical
955 Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- 956 Rettinghouse, T. 2010. IEP Newsletter. Interagency Ecological
957 Program for the San Francisco Estuary. 23(2): 5-6.
- 958 Rosenfield, J.A., and Baxter, R.D. 2007. Population Dynamics and
959 Distribution Patterns of Longfin Smelt in the San Francisco
960 Estuary. *Trans. Am. Fish. Soc.* 136(6): 1577-1592.

- 961 doi:10.1577/t06-148.1.
- 962 Rosenfield, J. A. 2010. Life history conceptual model and sub-
963 models for longfin smelt, San Francisco Estuary population.
964 Report for Delta Regional Ecosystem Restoration
965 Implementation Plan. California Department of Fish and
966 Wildlife, Sacramento, CA. Available from:
967 <https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=28421>
- 968 Sağlam, İ.K., Baumsteiger, J., Smith, M.J., Linares-Casenave,
969 J., Nichols, A.L., O'Rourke, S.M., and Miller, M.R. 2016.
970 Phylogenetics support an ancient common origin of two
971 scientific icons: Devils Hole and Devils Hole pupfish. *Mol.*
972 *Ecol.* 25(16): 3962-3973. doi: 10.1111/mec.13732.
- 973 Sibley, T.H., and Chigbu, P. 1994. Feeding Behavior of Longfin
974 Smelt (*Spirinchus thaleichthys*) May Affect Water Quality
975 and Salmon Production in Lake Washington. *Lake Reserv.*
976 *Manage.* 9(1): 145-148. doi:10.1080/07438149409354747.
- 977 Skinner, J. E. 1962. An historical review of the fish and
978 wildlife resources of the San Francisco Bay Area. Department
979 of Fish and Game Water Projects Branch Report No. 1.
980 http://stripersurf.com/california_history.html
- 981 Skotte, L., Korneliussen, T.S., and Albrechtsen, A. 2013.
982 Estimating individual admixture proportions from next
983 generation sequencing data. *Genetics*, 195(3): 693-702.
984 doi:10.1534/genetics.113.154138.

- 985 Sommer, T., Armor, C., Baxter, R., Breuer, R., Brown, L.,
986 Chotkowski, M., Culberson, S., Feyrer, F., Gingras, M.,
987 Herbold, B., Kimmerer, W., Mueller-Solger, A., Nobriga, M.,
988 and Souza, K. 2007. The Collapse of Pelagic Fishes in the
989 Upper San Francisco Estuary. *Fisheries*, 32(6): 270-277.
990 doi:10.1577/1548-8446(2007)32[270:tcopfi]2.0.co;2.
- 991 Thomson, J.R., Kimmerer, W.J., Brown, L.R., Newman, K.B., Mac
992 Nally, R., Bennett, W.A., Feyrer, F., and Fleishman, E.
993 2010. Bayesian change point analysis of abundance trends
994 for pelagic fishes in the upper San Francisco Estuary.
995 *Ecol. Appl.* 20(5): 1431-1448. doi:10.1890/09-0998.1.
- 996 US Fish and Wildlife Service (USFWS). 2012. Notice of 12-Month
997 Petition Finding (Longfin Smelt). Docket No. FWS-R8-Es-
998 2008-0045.
- 999 Wilson, G.A., and Rannala, B. 2003. Bayesian inference of recent
1000 migration rates using multilocus genotypes. *Genetics*,
1001 163(3): 1177-119.

1002 **Table Legend:**

1003 **Table 1.** List of samples used, location code, number collected,
1004 year collected, and approximate GPS coordinates.

1005

1006 **Figure Legends:**

1007 **Figure 1.** Map of the northeastern Pacific coast, with sample
1008 locations of longfin smelt collections for this study.

1009 **Figure 2.** Genetic structure of longfin smelt along the
1010 northeastern Pacific coast as summarized by PCA for comparisons
1011 between the four significant principal component axes. PC1-PC2
1012 (A), PC1-PC3 (B) and PC1-PC4 (C). Proportion of variance
1013 explained by each PC axis is given in parentheses.

1014 **Figure 3.** Shared ancestry between longfin smelt along the
1015 northeastern Pacific coast as revealed by admixture analysis for
1016 $K=3$ and $K=5$.

1017 **Figure 4.** Genetic differentiation between longfin smelt
1018 collections along the northeastern Pacific coast as measured by
1019 F_{st} statistics

1020 **Figure 5.** Contemporary migration rates between collections of
1021 longfin smelt, as probability of migrants received within the
1022 last three generations in each population. Each box represents a
1023 different sink population, and the code locations below indicate
1024 sources.

1025 **Figure 6.** Genome wide values of per site $\theta\pi$ (average pairwise
1026 nucleotide differences) for longfin populations along the
1027 northeastern Pacific coast.

1028 **Figure 7.** Reference allele frequencies of sites under selection
1029 along PC3 (A) and PC4 (B)

1030

1031

1032 **Supplemental Tables**

1033 **Table S1.** Alignment statistics for all individuals used in the
1034 study. Individuals in bold were removed from the study due to
1035 having less than 100,000 alignments after removal of duplicate
1036 reads (PCR clones).

1037 **Table S2.** Mean, median, and 95% Bayesian credible intervals
1038 (highest probability densities; HPD) for migration probability
1039 estimates between longfin smelt populations. Each box represents
1040 a different sink population, the below locations represent
1041 sources.

1042 **Table S3.** Result of selection scans along PC3 for sites showing
1043 allele frequency differentiation significantly greater than
1044 expected by genetic drift as determined by a log likelihood
1045 ratio with a genome wide significance threshold of 9.40×10^{-7}
1046 (LRT > 24). Frequency of the reference allele (minor allele with
1047 reference to all collection sites) in each population is also
1048 given.

1049 **Table S4.** Result of selection scans along PC4 for sites showing
1050 allele frequency differentiation significantly greater than
1051 expected by genetic drift as determined by a log likelihood
1052 ratio with a genome wide significance threshold of 9.40×10^{-7}
1053 (LRT > 24). Frequency of the reference allele (minor allele with
1054 reference to all collection sites) in each population is also
1055 given.

1056 **Supplemental Figures**

1057 **Figure S1.** Folded site frequency spectra of longfin populations

1058 **Figure S2.** Max Delta K graph showing the most likely Ks using
1059 the Evanno method (Evanno et al. 2005)

1060 **Figure S3.** Shared ancestry between longfin smelt collection
1061 locations for admixture analysis from K=2 to K=11

1062 **Figure S4.** Genetic structure within the SFE as determined by
1063 admixture analysis for K=2 to K=4 and PCA

1064 **Figure S5.** Genetic structure between Pitt and Harrison lakes
1065 within the Fraser River system as determined by PCA and

1066 admixture analysis for K=3 and K=4

1067 **Figure S6.** Log probability trace files and density curves for
1068 the three independent runs of BayesAss for estimation of
1069 migration rates between populations. Trace files in all three
1070 runs show adequate mixing and unbiased exploration of the
1071 parameter space while density curves of all three runs stack on

1072 top of each other indicating convergence of all three runs to
1073 the same parameter space.

1074

Draft

Table 1. List of locations where longfin smelt samples were collected, including location code, number (N) collected, year(s) collected, and approximate GPS coordinates.

Estuary	Location	N	Year(s) collected	GPS coordinates
San Francisco Estuary (SFBY)	Alviso Slough (ALVS), California	12	2013	37.4435508, -122.0055163
	Chippis Island (CHPI), California	19	2007, 2008, 2009	38.0551975, -121.911904
	Petaluma River (PETA), California	6	2015	38.1104751, -122.4874776
	Suisun Bay (SUIB), California	20	2008, 2010	38.07, -122.07
Humboldt Bay (HUMB)	HFAC Weir, Freshwater Creek, California	8	2009, 2012, 2015	40.759830294, -124.217332464
Columbia River Estuary (COLR), Oregon	Trestle Bay, North Channel, Oregon	22	2011	46.213389, -123.978778
Lake Washington (LWSH)	Lake Washington, Washington	20	2008	47.6215, -122.2558
Fraser River (FRAS)	Pitt Lake (PTLC), British Columbia	20	2000	49.4399889, -122.5400974
	Harrison Lake (HRLC), British Columbia	13	2006	49.500000, -121.833336
Skeena (SKNA)	Skeena River, British Columbia	10	2011	54.1499994, -129.9499962

Yakutat Bay, Alaska
(YBAK)

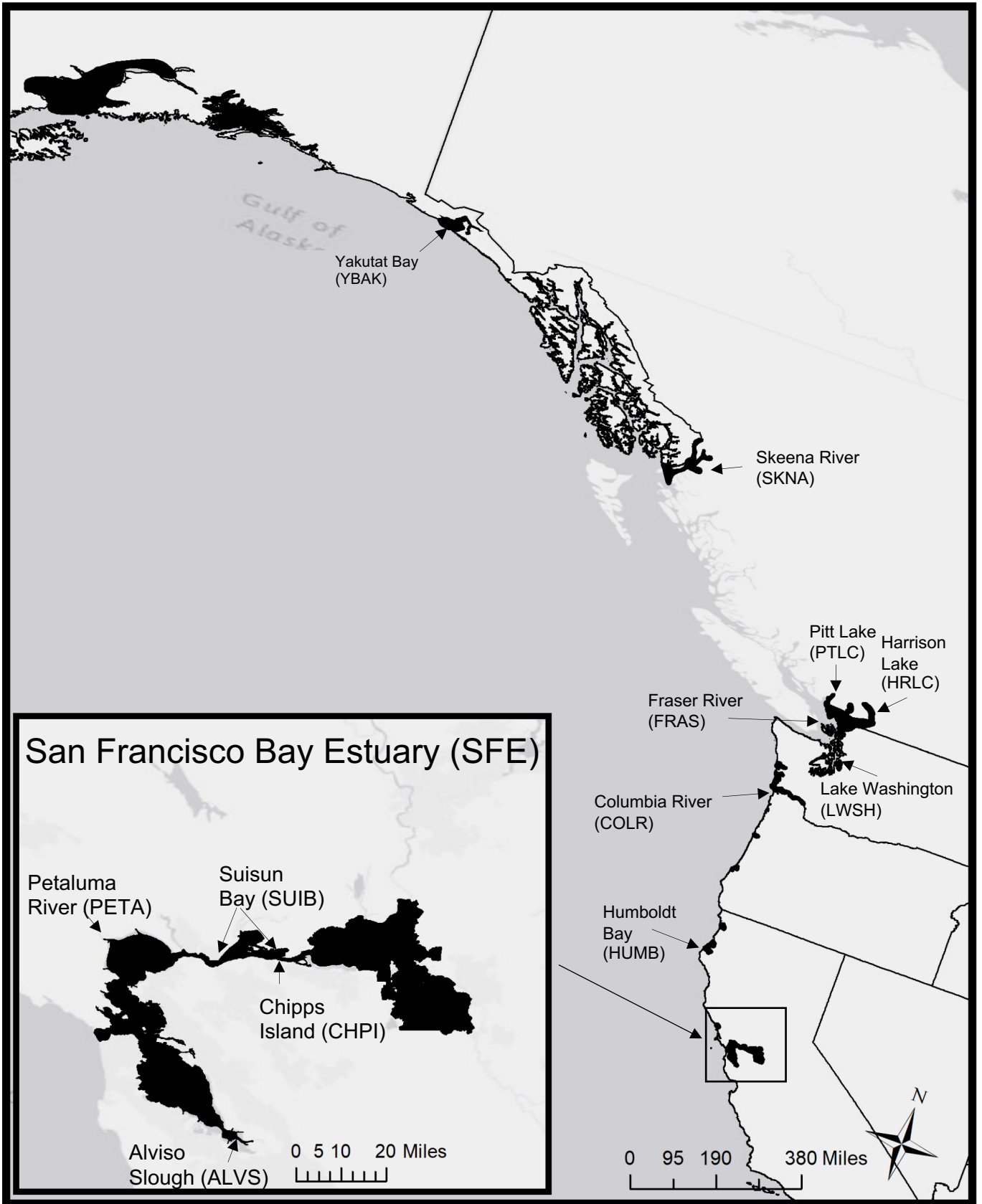
Yakutat Bay, Alaska

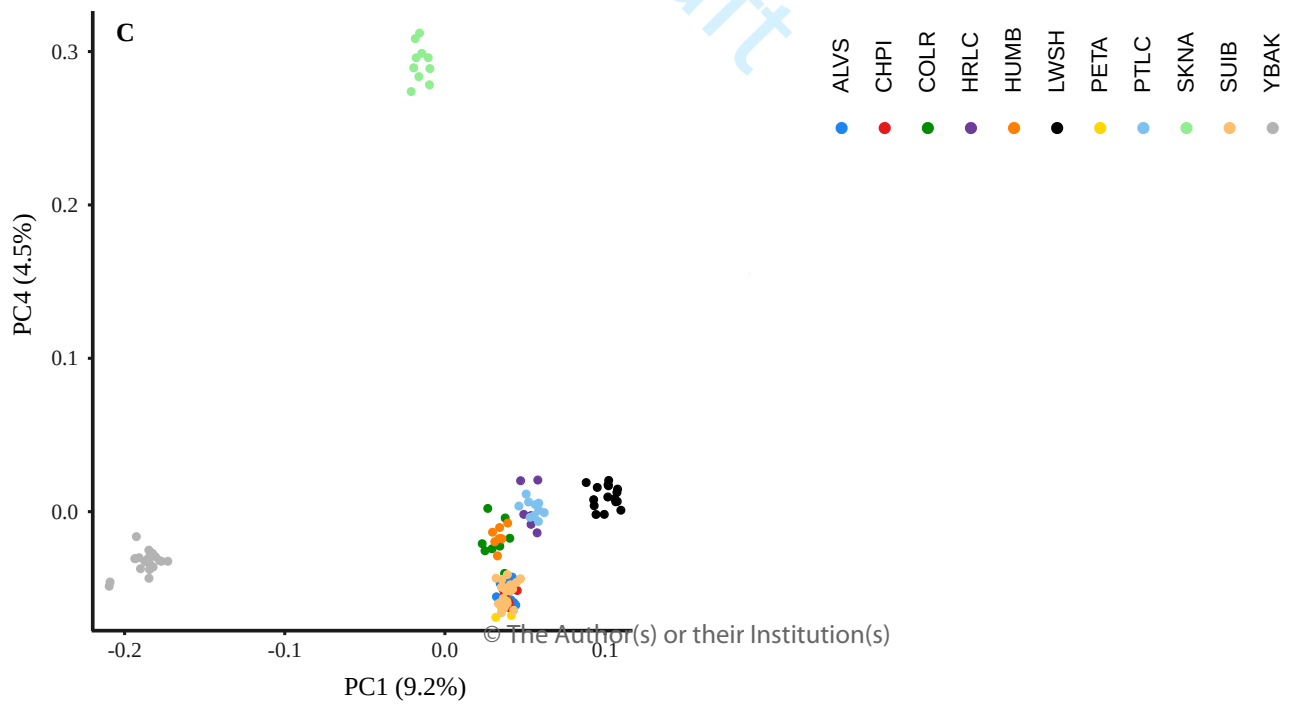
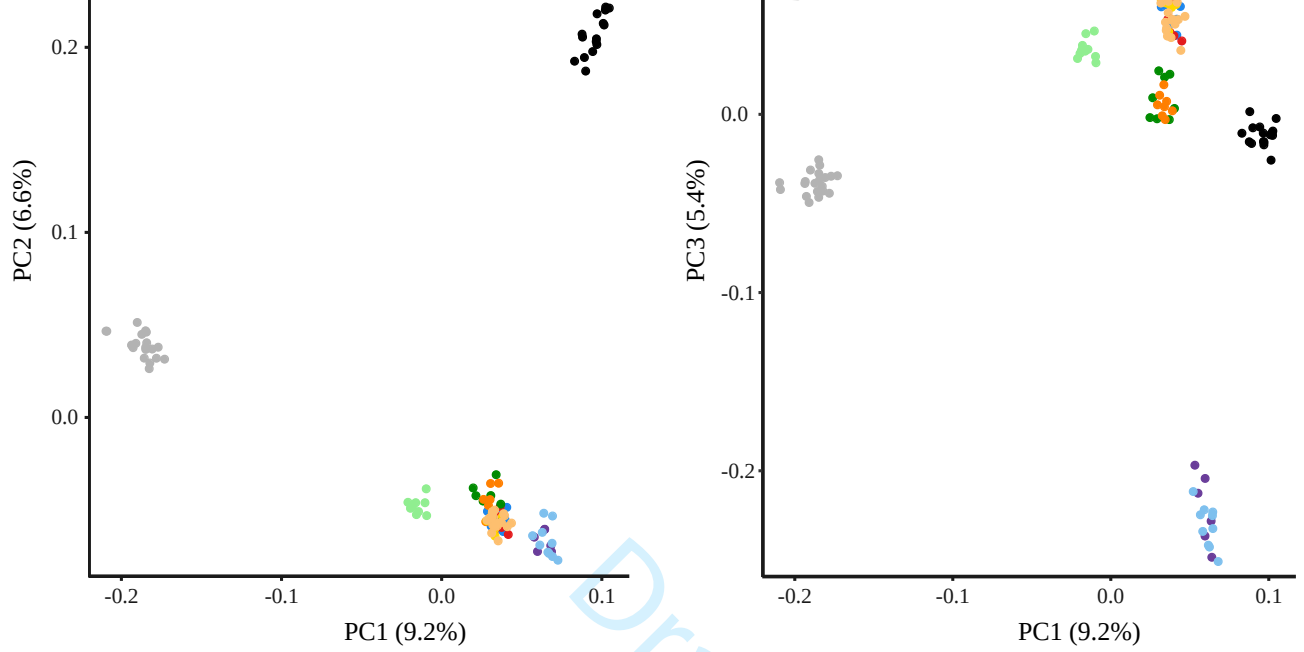
20

2013

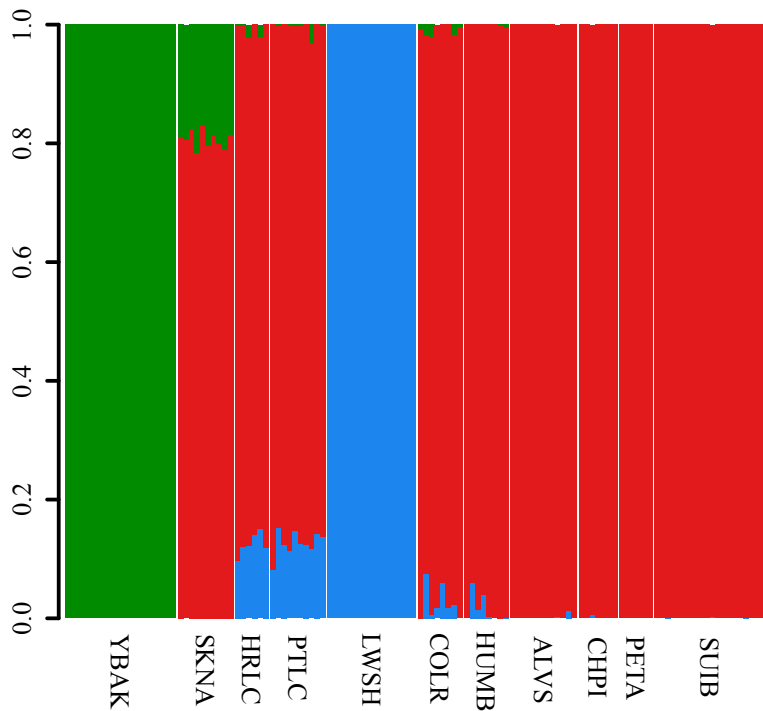
59.725663764,
-139.836496654

Draft





Admixture, K=3



Admixture, K=5

