bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 **Title**

2 Every cog and wheel: Identifying biocomplexity at the genomic and phenotypic level in a population

3 complex of Chinook salmon

4 **Running Title**

5 Biocomplexity of salmon population complex

6 **Authors**

7 Shannon J. O'Leary[1], Tasha Q. Thompson[2], Mariah H. Meek*[2]

8 **Affiliations**

9 [1]Saint Anselm College, Department of Biology, 100 Saint Anselm Drive, Manchester, NH 03102, United

10 States

11 [2]Department of Integrative Biology, AgBioResearch, and Ecology, Evolution, and Behavior Program,

12 Michigan State University, East Lansing, MI, United States

13 *corresponding author: mhmeek@msu.edu

14

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

15 **Abstract**

16 Genomic diversity is the fundamental building block of biodiversity and the necessary ingredient for
17 adaptation. Our rapidly increasing ability to quantify functional, compositional, and structural genomic
18 diversity of populations forces the question of how to balance conservation goals – should the focus be on
19 important functional diversity and key life history traits or on maximizing genomic diversity as a whole?
20 Specifically, the intra-specific diversity (biocomplexity) comprised of phenotypic and genetic variation
21 can determine the ability of a population to respond to changing environmental conditions. Here, we
22 explore the biocomplexity of California's Central Valley Chinook salmon (*Oncorhynchus tshawytscha*)
23 population complex at the genomic level. Notably, despite apparent gene flow among individuals with the
24 same migration (life history) phenotypes inhabiting different tributaries, each group is characterized by a
25 surprising component of unique genomic diversity. Our results emphasize the importance of formulating
26 conservation goals focused on maintaining biocomplexity at both the phenotypic and genotypic level.
27 Doing so will maintain the species' adaptive potential and increase the probability of persistence of the
28 population complex despite changing environmental pressures.

29 **Keywords**

30 Portfolio effect, intraspecific diversity, conservation genomics

31 **Introduction**

32 *"The last word in ignorance is the man who says of an animal or plant, "What good*
33 *is it?"[...] If the biota, in the course of aeons, has built something we like but do not*
34 *understand, then who but a fool would discard seemingly useless parts? To keep*
35 *every cog and wheel is the first precaution of intelligent tinkering."*
36 ***— Aldo Leopold***

37 Genetic diversity represents the fundamental building block of biodiversity – from genes to species,
38 to communities and ecosystems. The genetic diversity contained within populations forms the cogs and
39 wheels comprising the adaptive potential of a population. Protecting this diversity is a central task of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

40 conservation biology. However, identifying and quantifying biodiversity is a complex undertaking with
41 distinct frameworks for enumerating compositional, structural, and functional diversity (Petchey &
42 Gaston 2002; Duelli & Obrist 2003; Péru & Dolédec 2010). Functional genetic diversity includes both
43 sequence polymorphisms and differences in gene expression, which together shape the phenotypic
44 diversity comprised of differences in morphology, physiology, and life history characteristics present
45 within a population. Under current environmental conditions, it is expected that some components of
46 functional diversity will be selectively neutral, but future shifts in conditions and evolutionary pressures
47 may result in certain phenotypes and genotypes that appeared unimportant under previous conditions
48 becoming critical to the persistence of a species or population (Messer et al. 2016). Therefore, in order to
49 maintain critical evolutionary and ecological processes that sustain biodiversity across scales, sound
50 conservation and management strategies require a fundamental understanding of both past and present
51 patterns of biodiversity to enable the protection of diversity that will form the building blocks for future
52 adaptation (CBD 2011; Hoelzel et al. 2019; Mable 2019).
53 Increasingly, the importance of intra-specific diversity (biocomplexity) has been recognized as a
54 determining factor for the stability and resilience of biological systems (Hilborn et al. 2003; Des Roches
55 et al. 2021). For example, at an ecosystem level, the number of trophic levels and number of species at
56 each level determines the stability of the food web, while at a species level, diversity of life history
57 strategies can be critical for maintaining a temporally stable population through risk partitioning (i.e., the
58 portfolio effect (Hilborn et al. 2003; Schindler et al. 2010)). While biocomplexity is frequently viewed as
59 a horizontal measure within a given level of biodiversity, it cannot always be neatly confined to a single
60 hierarchical level. For example, while intra-specific diversity at a population-level is comprised of
61 phenotypic diversity, there is a genetic component dictating some phenotypic differences. Similarly,
62 species interactions not only shape an ecosystem but simultaneously act as evolutionary pressures
63 determining the fitness of individual phenotypes. Therefore, because selection acts on standing variation,
64 forward-looking conservation strategies must stress the importance of maintaining adequate levels of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

65 genetic diversity within a population to maintain adaptive potential in changing environmental conditions
66 (Mimura et al. 2017; Mable 2019; Hoban et al. 2020; Des Roches et al. 2021).
67 Chinook salmon, *Oncorhynchus tshawytscha*, in the Central Valley are emerging as a case study to
68 understanding the importance of biocomplexity for the persistence of a population complex facing
69 multiple external threats, including habitat fragmentation, overexploitation, and climate change. Here, life
70 history diversity resulting from different migration phenotypes creates a "portfolio of stocks" buffering
71 against spatiotemporally variable environmental conditions and anthropogenic impacts, which in turn
72 increases resilience and promotes interannual stability (Carlson & Satterthwaite 2011; Griffiths et al.
73 2014). Chinook are anadromous with a distinct life history spanning both freshwater and marine
74 ecosystems. Eggs are laid and hatched in the tributaries, where juveniles rear for a period of time before
75 migrating out to the ocean. There, they spend one to several years growing in the ocean before migrating
76 back to their natal river to spawn and die, providing an important source of oceanic nutrients to
77 ecosystems and supporting recreational, commercial, and heritage fisheries (Quinn 2018). The tributaries
78 of the Sacramento-San Joaquin River system contain four distinct run types (migration phenotypes)
79 named for the time of year adults enter freshwater systems to spawn: Winter (endemic to the Central
80 Valley), Spring, Fall, and Late-Fall; the same tributary may support multiple runs (Williams 2006). Early-
81 migrating runs (Winter, Spring) make the trade-off of migrating earlier at a smaller size, leaving behind
82 the nutrient-rich oceanic habitats to access spawning sites higher in the watershed that remain cool over
83 the summer, where they complete maturation in a fasted state while relying on fat stores, spawn, and die
84 (Quinn et al. 2016). By contrast, late-migrating individuals (Fall, Late-Fall) remain in ocean until
85 relatively mature before making their spawning run.
86 This asynchronicity in run timing of adults stabilizes the population complex overall, as temporal
87 and spatial variation in challenges to population components will vary. For example, while environmental
88 conditions within a given year may be poor for early migrating adults, in that same year they may be
89 optimal for their late-migrating counterparts in the same tributary, thus buffering the population complex
90 overall. However, this buffering ability is threatened when one run-type is consistently negatively

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

91 impacted across tributaries. Specifically, dams and other anthropogenic factors have disproportionately
92 affected historical early migrator habitat in much of the Central Valley. As a result both Spring and
93 Winter run are listed under the Endangered Species Act (National Marine Fisheries Service 2005) and the
94 dramatic declines in their abundance and distribution has resulted in a loss of biocomplexity across run
95 types, thus lowering the portfolio effect and making the population complex as a whole more vulnerable
96 (Carlson & Satterthwaite 2011).

97 These demographic changes are also likely to result in an erosion of genetic diversity and
98 consequently adaptive potential due to low effective sizes, which may otherwise have proved important
99 for persistence under future environmental conditions. Measuring the genetic diversity of a population has
100 long been used as a proxy to quantify the "future adaptive potential" of populations (Reed & Frankham
101 2001, 2003). However, despite a plethora of tools to quantify genetic diversity, consensus on which part
102 of the genetic diversity to focus on for conservation remains elusive. By focusing on the structural
103 biodiversity at a genomic level, tools such as diagnostic single nucleotide polymorphism (SNP) panels for
104 Chinook can be used as intrinsic markers to identify (hierarchical sets of) demographic groups, quantify
105 biocomplexity across a species' range, and describe changes in the portfolio of stocks (Meek et al. 2016).
106 Recent technological advances now also allow for the direct interrogation of functional biodiversity at the
107 genomic level. For example, recent studies have identified a single chromosomal region (GREB1L to
108 ROCK1) underlying adult migration timing in Chinook and other salmonids (Prince et al. 2017; Narum et
109 al. 2018; Thompson et al. 2019, Thompson et al. 2020). In addition, the increasing availability of genomic
110 data sets holds promise for identifying genes associated with polygenic traits (Ouborg et al. 2010;
111 Sinclair-Waters et al. 2020). Finally, focusing on the compositional biodiversity at a genomic level will
112 lead to an increased understanding of the differences in genetic diversity and unique components among
113 populations within a complex.

114 The increased availability of data quantifying functional, compositional, and structural genomic
115 diversity of populations forces questions about what is important to consider for conservation. There is a
116 need for balancing goals focused on preserving important functional diversity and key life history traits

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

117 with goals aimed at preserving adaptive potential that may be more cryptic, especially when putatively
118 neutral and adaptive markers display divergent signals (Waples & Lindley 2018). This is highlighted by
119 recent studies that identified a major effect locus underlying adult run timing in Chinook salmon (Prince
120 et al. 2017; Narum et al. 2018; Thompson et al. 2019; Ford et al. 2020). The large effect size of this locus
121 (e.g., explaining >85% of the variance in freshwater entry timing; (Thompson et al. 2020) combined with
122 widespread declines and extirpations of spring-run populations and the finding that spring-run alleles are
123 not preserved in the absence of the phenotype (Thompson et al. 2019; Ford et al. 2020) has led to debate
124 on the extent to which conservation policy should address this single locus (Langin 2018; Ford et al.
125 2020). The debate is additionally complicated because it is poorly understood to what extent adaptive
126 variation beyond this locus may be impacted by loss of the spring-run life history. For example, spring-
127 and fall-run fish historically utilized largely disparate spatio-temporal habitats (e.g., spring-run typically
128 spawned earlier and higher in watersheds; Quinn et al. 2016), which could facilitate the development of
129 important local adaptative differences between the runs. However, human activities have homogenized
130 habitat and substantially increased interbreeding between spring- and fall-run Chinook in many locations
131 (Ford et al. 2020), and it is unclear whether local adaptation has been preserved. Despite this, the
132 remaining wild spring-run populations in California's Central Valley still access very distinct habitat from
133 their fall-run counterparts. Furthermore, a great deal of habitat heterogeneity exists *within* the spring-run.
134 Thus, the Central Valley provides an opportunity to examine the extent to which additional unique genetic
135 variation may accumulate between runs and across runs in a population complex where spatio-temporal
136 separation between runs is relatively intact.

137 Here, we aim to understand the genomic biocomplexity, contained within and among the four
138 Central Valley Chinook salmon run types. To achieve this, we leverage a previously published dataset
139 (Meek et al. 2019). The exceeding value of this data set lies in the fact that it encompasses genomic
140 sequences from all the major populations of the critically endangered and threatened Central Valley
141 Chinook salmon, including winter run, which is on the verge of extinction. The analysis of Meek et al.
142 (2019) demonstrated greater population diversity and structure across the Central Valley than had been

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

143 previously described but lacked the genomic resources and analytical methods needed to elucidate the
144 genomic biocomplexity contained within and among populations. This understanding is vital for informed
145 conservation planning that will protect the full portfolio of diversity and better ensure long term
146 persistence and resilience. In this paper, we resolve this problem by using the recently completed Chinook
147 salmon genome (Christensen et al. 2018) and a novel microhaplotype analysis (Willis et al. 2017) to
148 identify > 10,000 multi-allelic loci distributed throughout the coding and non-coding parts of the Chinook
149 genome. Our results not only provide new insight about this important and highly imperiled species, but
150 also demonstrates the power of applying new advances to existing dataset to gain vital biological
151 understanding, without the need for resampling imperiled populations.

152 **Materials and Methods**

153 *Sample collection and sequencing*

154 As reported in Meek et al. 2019, fin clips from all four run types (Fall, Late-Fall, Winter, Spring;
155 Figure 1) were obtained from adult Chinook Salmon from each major tributary in the Central Valley
156 during their spawning migrations. Genomic DNA was extracted and digested using *SbfI* to construct RAD
157 libraries following Miller et al. (Miller et al. 2012). Fifteen libraries consisting of 30 – 47 individuals each
158 were sequenced on single Illumina HiSeq 2000 lanes (100bp, single end).

159 *Genotyping*

160 Raw sequences from Meek et al. 2019 were demultiplexed using *process_radtags* (Catchen et al.
161 2011) and quality trimmed using *fastp* (Chen et al. 2018). Quality trimmed reads were mapped to a
162 Chinook reference genome (Christensen et al. 2018) using *BWA-mem* (Li & Durbin 2009). Reads with a
163 mapping quality < 5 were removed and filtered bam files were concatenated into a single bam file to
164 query coverage per interval using *bedtools* (Quinlan 2014). Mapping intervals > 25bp and < 500bp with
165 coverage > 50 were extracted to create a reduced-representation reference consisting of the recovered
166 RAD-tags, reads were re-mapped to the reduced-representation reference and SNPs were called using
167 *freebayes;* the maximum allowed gap (-E) was set to three and minimum mapping and base quality set to
168 five, otherwise default parameters were used (Garrison & Marth 2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

169 The raw data set was rigorously filtered to remove low quality genotypes, loci, and samples
170 following principles set forth in O'Leary et al. (2018). In short, genotypes with < 5 reads or quality < 20
171 were coded as missing, retained loci had a mean minimum depth of 15, and were called in at least 50% of
172 individuals of a given library, 85% of individuals of a run type, or 90% of individuals overall. Loci were
173 further filtered for allele balance, mapping quality ratio of reference vs alternate allele, depth/quality ratio,
174 and excess heterozygosity to remove paralogs and other technical artifacts. Finally, *rad_haplotyper*
175 (Willis et al. 2017) was used to collapse SNPs on the same RAD-tag into haplotypes. Detailed processing
176 steps and threshold values are contained in Supplementary Material 1. In general, biallelic SNPs contain
177 less information per locus compared to multi-allelic loci such as microsatellites (Morin et al. 2009) and
178 the necessity of thinning SNPs to ensure loci are independent observations (Kaeuffer et al. 2007) further
179 reduces the information content of a data set as the power of a data set resides in the number of
180 independent alleles rather than the number of loci (Kalinowski 2002). Haplotyping a data set rather than
181 thinning preserves the information content of all SNPs in the data set, resolves physical linkage artefacts,
182 and results in more inferential power per locus (Willis et al. 2017; Baetscher et al. 2018). Therefore, this
183 new dataset has much more power to identify genomic diversity and biocomplexity.

184 *Assessment of population structure and differentiation*

185 Population structure was explored using two methods, a clustering analysis based on genetic
186 similarity and an assessment of population differentiation among individuals grouped *a priori* based on
187 run type and tributary. In the first method, individuals were clustered into K = 1 – 10 groups using k-
188 means clustering based on the PCA-transformed genotype matrix (i.e. no assumptions regarding Hardy-
189 Weinberg or linkage disequilibrium) followed by a discriminant analysis of principle components
190 (DAPC) to determine the membership probabilities of each sample to each inferred cluster as
191 implemented in *adegenet* (Jombart et al. 2010). To ensure sufficient variance was retained to discriminate
192 among groups but not overfit the data, the optimum number of principle components to retain was
193 determined using a stratified cross-validation of DAPC.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

194 In the second method, Weir & Cockerham's unbiased estimator of F_{ST} (Weir & Cockerham 1984)
195 was used to calculate population differentiation among individuals grouped *a priori* by run type within
196 tributary. To test for genetic heterogeneity in the data set, global F_{ST} was calculated across all groups,
197 then, pairwise F_{ST} was calculated as a *post hoc* test for pairwise differences among groups. Significance
198 was determined using 95% confidence intervals around each estimate generated by resampling loci 1,000
199 times using *assigner* (Gosselin et al. 2016).

200 *Estimates of effective population size*

201 Estimates of effective population size, *N*e, for each tributary/run group were determined using
202 *LinkNe* (Hollenbeck et al. 2016), an extension of the linkage disequilibrium method (LD) designed for
203 data sets with loci of known linkage relationships. Only SNPs placed on a chromosome, with a minor
204 allele frequency > 0.01, and the first SNP per RAD-tag were used. Genomic distances (bp) were
205 converted to recombination rates (cM) using the size of the female Chinook linkage map (3,118 cM,
206 (Mckinney et al. 2016)) and the length of the genome (2.4 Gbp, (Christensen et al. 2018)) resulting in an
207 estimate of approx. 770kb equivalent to 1cM.

208 *Assessment of Genomic Diversity*

209 All measures of genomic diversity were made for individuals grouped by run type within each
210 tributary, with wild and hatchery individuals treated as separate groups. Four types of parameters were
211 assessed, (1) measures of heterozygosity, (2) measures of allelic diversity, (3) sequence-based parameters,
212 and (4) measures of unique variation.

213 For the first three sets of parameter types, significant heterogeneity was determined using a
214 Friedman's rank sum test followed by a *post hoc* Wilcoxon signed-rang test to test for significant pairwise
215 differences between groups; *p*-values were corrected for multiples comparisons assuming an FDR of 0.05
216 (Benjamini & Hochberg 1995). The observed heterozygosity (Ho) was measured as the proportion of
217 heterozygote genotypes per locus (Nei 1987) and the expected heterozygosity (gene diversity) (Hs) as the
218 proportion of heterozygous genotypes expected under Hardy-Weinberg Equilibrium (Nei 1987). To
219 account for differences in sample size, allelic richness was measured as rarefied allele counts. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

220 evenness of allelic diversity at a given locus was calculated as the ratio of the number of abundant to the
221 number of more rare genotypes using the ration of the Stoddart & Taylor index (diversity index weighted
222 for more abundant alleles) and Shannon-Wiener index (diversity weighted for more rare alleles) as
223 implemented in *poppr* (Kamvar et al. 2014)*;* lower values indicate prevalence of more rare alleles and
224 uneven distributions of allele frequencies. The nucleotide diversity statistic π (Nei 1987) was calculated
225 as the sum of the number of pairwise differences between haplotypes of a given nucleotide over the
226 number of comparisons made; this parameter is biased toward alleles segregating at intermediate rates and
227 will underestimate genetic diversity when many rare alleles are present.

228 Finally, patterns of unique diversity were assessed by comparing (1) the number of fixed loci (i.e. not
229 polymorphic within a group), (2) the number of loci with singletons and the number of singletons per
230 individual by run/tributary group, (3) the number of private polymorphisms, and (4) the number of private
231 alleles. Private polymorphisms are defined as loci where more than one allele is found only in a single
232 group (all other groups are fixed for a single allele). By contrast, private alleles are alleles found in only a
233 single group, though other groups exhibit more than one allele at that locus. To compare whether
234 identified loci are randomly distributed across chromosomes, null distributions were generated by
235 shuffling chromosome designations across loci 1,000 times to determine whether the observed values fall
236 outside the null distribution.

237 **Results**

238 *Genotyping*

239 The final filtered data set consisted of 386 individuals genotyped for 12,983 multi-allelic loci
240 (hereafter loci) with a total of 30,037 alleles (Table 1).

241 *Assessment of population structure and differentiation*

242 Minimum AIC was observed for K = 4 (Figure S1); cross-validation indicated that the optimum number
243 of principle components to retain was 20 – 150 (retaining 11.6 – 75.1% of variance; Table S1). The mean
244 optimum success of assignment declined from 100% (K = 2 – 3) to 96% (K = 7) and then dropped to 89%
245 for K = 8 (Table S1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

246 Figure 1 summarizes membership plots for K = 2 – 8. For K = 2 Winter run individuals form a single
247 cluster set against individuals from all other run types and tributaries. Similarly, for K = 3, Winter run
248 form a single cluster, a second cluster consists of Spring run individuals from Butte, Deer, and Mill
249 Creek, while Spring Feather River Hatchery and Late-Fall individuals form the final cluster along with all
250 Fall groups. Overall, Late-Fall individuals do not emerge as their own distinct cluster until K = 6, while
251 Spring Feather River individuals continue to cluster with Fall individuals from Coleman and Feather
252 River Hatchery for K = 7 – 8. Overall, Spring run individuals form at most three clusters, with Deer and
253 Mill Creek individuals always clustering together. In general, Fall run individuals are assigned to two or
254 three clusters, here, Coleman Hatchery individuals generally form one cluster along with most Feather
255 River Hatchery individuals (both Fall and Spring) and some Deer and Mill Creek individuals; increasing
256 K to 8 results in Coleman Hatchery individuals starting to form a more distinct cluster of their own.
257 For individuals grouped by run type within tributaries, global F_{ST} = 0.0319 [0.0309 – 0.0328]
258 indicates significant heterogeneity among groups. Pairwise comparisons of Winter individuals and all
259 other run/tributary groups exhibit the highest observed pairwise F_{ST}-values (0.14 - 0.161). By contrast, all
260 non-significant comparisons, with CIs including zero, were comparisons among Fall run tributary
261 populations. Late-Fall Upper Sacramento River individuals are significantly different from all other
262 run/tributary groups. For Late-Fall and Winter individuals pairwise F_{ST} is higher (0.016) than for Late-
263 Fall/Fall comparisons (0.001 - 0.01), despite inhabiting the same tributary. Spring Feather River Hatchery
264 individuals stand out as having lower pairwise F_{ST}-values in comparison to Fall (0.003 - 0.013) and Late-
265 Fall (0.016) tributaries than to other Spring tributaries (0.018 - 0.036). Similarly, for all Spring/Fall
266 tributary comparisons the lowest observed value is for Spring/Fall Feather River Hatchery individuals
267 (0.003). Further details are summarized in Table S2/Figure S2.

268 *Estimates of effective population size* Ne

269 Ne estimates for Fall run ranged from 1,125 to 7,432, with the exception of Coleman Hatchery,
270 which was the lowest (Ne = 615, CI = 594 – 637) and Merced River, which was the highest (11,436 CI =
271 6,768; 36,804; Table 1). Estimates for spring groups were lower, ranging from 85 (CI = 84 – 86; Mill

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

272 Creek) to 561 (CI = 529 – 597; Butte Creek), while Late-Fall had an *N*e = 3,114 (CI = 2,398 – 4,435) and
273 Winter run *Ne*=174 (CI = 171 – 177, Table 1). For Fall Deer and Mill Creek and Spring Feather River
274 Hatchery estimates were negative. This occurs when the LD attributed to sampling variation is larger than
275 the LD attributed to drift (i.e. samples sizes too low to accurately estimate *Ne)*. Here, the groups with
276 negative estimates are among the lowest sample sizes present in the data set (Table 1).

277 *Assessment of Genomic Diversity*

278 MEASURES OF HETEROZYGOSITY

279 The mean observed heterozygosity is lowest for Winter run individuals from the Upper Sacramento
280 River (mean = 0.1272), followed by Spring individuals from Butte Creek (mean = 0.1587) and highest for
281 Fall individuals from Coleman and Feather River Hatcheries (0.1713 and 0.1741, respectively (Table S3,
282 Figure 3A). Similarly, mean expected heterozygosity is lowest for Upper Sacramento River winter run
283 individuals (mean = 0.1285), and highest for Fall Feather River Hatchery and Mill Creek individuals
284 (mean = 0.1718 and 0.1714, respectively, Table S4, Figure 3B). For both observed and expected
285 heterozygosity, Spring tributaries exhibit a wider range of distributions compared to Fall tributaries,
286 despite the smaller number of Spring tributaries in the data set (Figure 3A, B). Late-Fall values fall into
287 the range of distributions observed among Fall tributaries (Figure 3A, B).

288 ALLELIC RICHNESS

289 The mean values of rarefied allele counts are comparable across tributary/run groups, ranging from
290 1.51 - 1.52 alleles per locus for all run/tributary groups except Late-Fall Upper Sacramento River (1.36)
291 and Spring Deer Creek individuals (1.48), which exhibit significantly lower mean values (Table S5, Table
292 S6). Despite similar mean values, most pairwise comparisons are significant (Table S5), indicating that
293 even though there is a relatively consistent global number of alleles per locus the patterns of which loci
294 are variable are consistently different across run/tributary groups. For example, Spring Butte Creek
295 individuals exhibit a pattern of rarefied allele counts significantly different from all other locations, and,
296 despite a mean value comparable to most other groups, also exhibit the highest median value (1.71;

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

297 Figure 3C). Overall, Spring tributaries exhibit more variation among tributaries compared to Fall
298 tributaries (median = 1.38 - 1.71; all pairwise comparisons are significant, Table S5, S6). Distributions
299 are more similar to each other across Fall tributaries, with Butte Creek individuals exhibiting the lowest
300 (1.38) and Feather River Hatchery individuals the highest (1.48) median values (Figure 3C). Notably,
301 individuals from different run types from the same tributary may exhibit quite different patterns. For
302 example, Fall individuals from Mill Creek and the Feather River Hatchery exhibit higher allele counts
303 compared to Spring individuals from the same tributary. By contrast, Fall individuals from Deer and
304 Butte Creek have lower allele counts compared to their Spring counter parts. The strongest contrast is
305 Butte Creek, where Spring individuals exhibit the highest allele counts overall while Fall individuals
306 exhibit the lowest allelic richness among all Fall tributaries (Figure 3C).

307 EVENNESS OF ALLELIC DIVERSITY

308 Overall, Winter individuals from Upper Sacramento River exhibit significantly lower evenness of
309 allelic richness across loci, indicating that many loci are characterized by rare alleles (Figure 3D). By
310 contrast, Fall Feather River individuals exhibit the highest median evenness (0.78) compared to all other
311 run/tributary groups (Table S7). Within Fall tributaries, mean levels of evenness are all approx. 0.76
312 though there are some significant differences in their overall distributions (Figure 3D). Again, Spring
313 tributaries exhibit a wider range of mean and median values for evenness compared to Fall (Figure 3D).

314 NUCLEOTIDE DIVERSITY

315 Winter Upper Sacramento River and Spring Butte Creek individuals exhibit lower nucleotide
316 diversity compared to all other locations (Table S8, Figure 3E). By contrast, Fall Feather River Hatchery
317 exhibit the highest mean nucleotide diversity; similarly, among Spring tributaries Feather River Hatchery
318 individuals exhibit the highest nucleotide diversity (Table S8, Figure 3E).

319 FIXED LOCI

320 Late-Fall Upper Sacramento individuals and Spring Butte Creek individuals exhibit the highest
321 number of fixed loci (6,416 and 6,210, respectively); for all other run/tributary groups 3,500 - 5,000 fixed

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

322 loci were identified (Figure 4A). The number of fixed loci in Spring tributaries is generally larger
323 compared to Fall tributaries with the exception of Spring Feather River. By far the two largest intersects
324 are loci fixed exclusively in a single location, Late-Fall Upper Sacramento individuals (460) and Spring
325 Butte Creek individuals (317). All other intersects are < 115 loci. Apart from Spring Feather River
326 hatchery individuals (27), Spring tributaries have more loci fixed in a given tributary (74 - 317) compared
327 to Fall tributaries where 10 - 61 loci are fixed among individuals from a single tributary. Notably, among
328 intersects of loci fixed in two locations, the three largest intersects are all a combination of Late-Fall
329 Upper Sacramento River and a wild Spring population (42 – 115 loci). Overall, about a third of intersects
330 of two run/tributary combinations are loci fixed among Late-Fall Upper Sacramento River individuals and
331 a second location. There is no observed pattern of loci more likely to be fixed among tributaries in
332 geographic proximity (Figure 4A).

333 Loci fixed in Late-Fall Upper Sacramento and Spring Butte Creek individuals also exhibit the
334 highest global allele diversity (mean = 1.46 and 1.41, respectively; Table S9), i.e., loci that are fixed in
335 these groups are more variable when alleles are tabulated across individuals from all runs/tributaries
336 (Figure 4B). By contrast, the global diversity of fixed alleles is lowest in Fall tributary groups, and overall
337 levels are more similar across tributaries (median = 1.13-1.14 with the exception of Fall Deer Creek and
338 Fall Butte Creek at 1.19) compared to spring tributaries where those distributions of global diversity are
339 higher and more variable (median = 1.28 - 1.41; Table S9). The distribution of global diversity among
340 Spring and Fall tributaries varies with some run/tributaries exhibiting much tighter ranges than others
341 (Figure 4B). In general, the proportion of loci that are fixed for a run/tributary group is consistent across
342 chromosomes and there is no distinct, non-random pattern (Tables S10-11).

343 SINGLETONS

344 Three hundred forty-seven (1.2% of the total) loci exhibit at least one singleton. A comparison of
345 individuals grouped by run/tributary demonstrates that Spring Butte Creek and Deer Creek individuals
346 (mean = 4.0 and 3.0, respectively) and Late-Fall Upper Sacramento River individuals (mean = 2.23),

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

347 exhibit the highest mean number of singletons per individual. In general, Fall tributaries exhibit a lower
348 mean number of singletons per individual compared to other run/tributary groups, indicating that they are
349 comparatively less characterized by rare alleles (Figure 4C; Table S12).

350 PRIVATE POLYMORPHISMS

351 Winter run individuals from Upper Sacramento River have the highest number of loci (N = 153) with
352 private polymorphisms; all other groups have < 100 loci with private polymorphisms (Figure 5). Spring
353 individuals from Butte Creek exhibit the lowest number of loci exclusively polymorphic among
354 individuals of a group (N = 18). Across Late-Fall individuals, 70 loci with private polymorphisms were
355 identified. This number is higher than observed for eight of eleven Fall groups. Notably, individuals from
356 hatcheries fall along the entire range of private polymorphisms; Fall Nimbus Hatchery individuals are on
357 the low end (N = 33) and Spring Feather River Hatchery individuals on the high end (N = 84). Comparing
358 the chromosomal positions of private polymorphisms indicates that Winter Upper Sacramento has the
359 highest mean proportion of loci on a chromosome that are fixed in all other groups (0.10). The second
360 highest mean proportion of loci per chromosome is 0.07, observed in Fall Mill Creek individuals, Spring
361 Deer Creek individuals, and Merced River Hatchery. The mean is lowest for Spring Butte Creek
362 individuals (Figure 5, Table S13). While there are chromosomes with significantly more/less than
363 expected numbers of loci with private polymorphisms no consistent non-random patterns stand out (Table
364 S14). Of the six run/tributary groups with the highest number of chromosomes exhibiting less than the
365 expected number of loci with private polymorphisms, four are Fall run hatcheries (Table S15).

366 PRIVATE ALLELES

367 While most run/tributary groups exhibit 80 - 180 private alleles, Spring Deer Creek individuals
368 exhibit the highest number of private alleles (267) while Fall Butte Creek and Spring Feather River
369 Hatchery individuals exhibit the lowest number of private alleles (34; Figure 5). There is no distinct
370 pattern of hatchery individuals exhibiting more/less private alleles compared to wild individuals, or
371 different run types having consistently higher/lower number of private alleles compared to others; though

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

372 all groups do exhibit private alleles (Figure 5). Most commonly, private alleles are only carried in 1 - 3
373 individuals. Winter Upper Sacramento individuals exhibit the highest mean number of individuals
374 carrying private alleles (N = 4.07) compared to other run/tributary groups, followed by Spring Butte
375 Creek individuals (N = 1.34), for all other groups private alleles are found in a mean of 1.01 - 1.17
376 individuals. Similarly, these two groups have the most "common" private alleles, carried in 25 and ten
377 individuals, respectively (Table S16). Comparing the chromosomal positions of loci with private alleles in
378 a single run/tributary group indicates that Spring Deer Creek has the highest mean proportion of loci with
379 private alleles on a chromosome (Table S17). While there are chromosomes with higher/lower than
380 expected number of private alleles for run/tributary groups no distinct patterns emerge (Table S18-19).

381 **Discussion**

382 The patterns of genetic diversity observed in populations are the product of past evolutionary and
383 ecological processes. Intra-specific diversity determines the standing variation upon which evolutionary
384 forces act, and levels of phenotypic and genetic variation determine the ability of a population to respond
385 to changes in environmental conditions, the resilience and persistence of a species, and the stability of the
386 ecosystems they inhabit (Prieto et al. 2015; Siefert et al. 2015). Therefore, assessing the biocomplexity of
387 a population complex can help predict the range of possible responses to changing conditions.
388 Overlooking the importance of biocomplexity at a genetic and phenotypic level and failing to preserve the
389 adaptive potential of populations could have irreversible consequences for the health and sustainability of
390 populations. Here, we present a first ever fine-scale assessment of the biocomplexity contained at the
391 genomic level in California's Central Valley Chinook salmon population complex. We leverage
392 advancements for genomic analysis to increase our understanding of the hidden diversity contained within
393 a species that is becoming increasingly threatened and at risk of extinction. During this in-depth
394 assessment, we found significant differentiation among and within run types and the tributaries they
395 inhabit, effective population sizes severely below suggested critical thresholds to avoid erosion of genetic
396 diversity and loss in fitness, and corresponding low levels of genetic diversity in populations that have
397 experienced recent demographic declines. Surprisingly, despite apparent gene flow among individuals of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

398 the same run type across tributaries, we found each run/tributary group was indeed characterized by a
399 distinct component of unique genomic diversity. This diversity is very likely important to the overall
400 genetic health of the populations, and population complex as a whole, and vital to consider in
401 conservation efforts. Overall, our results emphasize the importance of not only maintaining life history
402 (phenotypic) diversity within and among groups, but also maintaining the genetic diversity of each run
403 and tributary to enhance the portfolio effect, maintain adaptive potential, and ensure the long-term
404 persistence of Chinook salmon in the Central Valley.

405 Populations become increasingly vulnerable to environmental, demographic, and genetic stochastic
406 effects as their size decreases. Thus, monitoring *N*e of individual components of a population complex is
407 critical for identifying differences in the vulnerability of different groups to genetic issues, evaluating
408 population viability, and predicting evolutionary trajectories to guide conservation managers in their
409 decision-making. Here, individuals from the early migrating run types (Winter, Spring) for each tributary
410 exhibit the lowest *N*e values, well below recommended targets of *N*e = 1,000 to maintain adaptive
411 potential (Frankham et al. 2014). Additionally, we record the lowest estimate of effective population size
412 for the Winter run (Ne = 174) to date. These results further highlight the extreme and alarming risk of
413 extinction facing Winter run.

414 Similarly, we estimate dangerously low effective population sizes within Spring run populations,
415 with Mill and Deer Creek values being similar to those of Winter run (Ne = 85 and 188, respectively) and
416 Butte Creek being just above 500. There is fine-scale structuring within Spring run. Among wild
417 populations Butte Creek is significantly distinct from Mill and Deer Creeks, which are geographically
418 closer and environmentally more similar to each other. Further, all three wild Spring tributaries are
419 significantly distinct from Feather River Hatchery Spring run. Notably, Spring Feather River Hatchery
420 individuals group more closely with Fall individuals from Coleman and Feather River Hatcheries,
421 pointing towards introgression between runs due to hatchery practices (California-HSRG 2012).

422 Contrary to the current practice of managing Late-Fall and Fall runs as a single ESU, both results
423 presented here using multi-allelic haplotyped loci and the initial assessment of the sequence data using

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

424 biallelic SNPs (Meek et al. 2019b) add to the increasing evidence that Late-Fall is genetically distinct
425 from Fall. While our study shows evidence of some distinctions between Fall individuals from more
426 northern tributaries, the differentiation is weak, indicating a loss of biocomplexity with increasing
427 homogenization of Fall run individuals (Williamson & May 2005). In contrast to the early migrating runs
428 (Spring and Winter), *N*e is robust in the late-migrating Fall and Late-Fall groups and at levels sufficient to
429 maintain adaptive potential (Allendorf et al. 2010). The exception to this is Coleman Hatchery Fall
430 individuals, which had an *N*e an order of magnitude lower than other Fall run populations. Also, of note,
431 the *N*e from the Merced River Hatchery Fall run population was also orders of magnitude lower than the
432 *N*e for the putatively wild fish caught in the Merced River. Both of these populations are concerning from
433 a conservation perspective, as maintaining sufficiently large *N*e is a major concern for stocking programs,
434 especially when introgression with wild populations is likely (Ryman & Laikre 1991).

435 The erosion of genetic diversity due to both genetic drift and inbreeding are inversely proportional to
436 *N*e. Both processes lead to a decreased level of heterozygosity as the rate of alleles being lost due to drift
437 increases and individuals become increasingly likely to mate with individuals with similar genotypes.
438 Accordingly, if drift is the primary force shaping the genetic diversity within the declining early migrating
439 populations, we would expect to see low levels of heterozygosity and comparatively higher numbers of
440 fixed alleles. Indeed, Upper Sacramento Winter run individuals exhibit the lowest levels of heterozygosity
441 and other measures of genetic diversity compared to all other groups, and the highest number of fixed loci
442 is found in the Spring run groups. Notably, among Spring run groups we find much wider distributions
443 across all measures of diversity – this underscores the stochasticity of genetic drift. Not only is the effect
444 stronger in smaller populations (resulting in an accelerated loss of diversity), but how each component is
445 affected can differ. Thus, groups may diverge from each other by chance alone. In addition, each tributary
446 experiences a different selection regime, driven by environmental differences, resulting in an increased
447 impact of decoupled demographic and environmental stochastic events affecting each population.

448 The assessment of private polymorphisms (loci only variable in a single group) and private alleles
449 (alleles only found in a single group) reveals that within each tributary, each run exhibits unique

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

450 components of genetic diversity. Notably, despite Winter run individuals from the Upper Sacramento
451 River having the lowest level of diversity when comparing measures related to heterozygosity and allelic
452 diversity, they exhibit the highest level of private polymorphisms, though they exhibit less diversity. The
453 diversity that is present is unique compared to all the other runs. This is different from Butte Creek Spring
454 individuals which exhibit levels of heterozygosity and allelic diversity similar to Winter run individuals,
455 but also harbor a low number of private polymorphisms, indicating that there are differences in the
456 demographic and evolutionary forces that have shaped genetic diversity in these groups. As a result,
457 despite Spring and Winter run types sharing a similar early migration phenotype and evolutionary history
458 at the GREB1L locus, the genome-wide intra-specific diversity is unique within each group, suggesting
459 differences in standing variation for selective pressures to act upon. Additionally, our new analysis
460 reveals clear, important distinctions in the unique diversity harbored by Late-Fall individuals from the
461 Upper Sacramento compared to Fall run groups, despite Fall and Late-Fall individuals being managed as
462 a single ESU and sharing GREB1L genotypes. The Late-Fall population has lower overall allele counts
463 compared to the Fall populations and the number of private polymorphisms is higher for the Late-Fall
464 compared to almost all (8/11) of the Fall run groups, indicating differences in processes shaping these run
465 types and their evolutionary trajectories.

466 Finally, while all groups exhibit 80 – 180 private alleles, surprisingly it is groups other than Fall run
467 that tend to be characterized by rare alleles. Since rare alleles are expected to be lost first during
468 bottlenecks, the expectation would be that Fall groups (which have relatively large *Ne*) would instead
469 have a larger number of rare alleles than the groups with smaller *Ne* if genetic drift were the main process
470 affecting allele frequencies. Other processes that could produce these patterns include increased
471 homogenization resulting in an exchange of alleles among tributaries, historical hatchery practices, and/or
472 selection. By contrast, spatially and temporally heterogeneous environments promote and maintain
473 polymorphisms and high levels of standing genetic variation (Svardal et al. 2015; Gulisija & Kim 2015;
474 Bertram & Masel 2019). Indeed, Late-Fall and Spring Deer Creek individuals exhibit the highest numbers
475 of singletons, and Winter and Spring Butte Creek individuals exhibit the highest mean number of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

476 individuals carrying private alleles. Even Spring Feather River Hatchery individuals, which have
477 introgressed with Fall individuals in the past, carry private polymorphisms. Overall, these patterns
478 support the conclusion that processes such as balancing selection are occurring to maintain diversity,
479 despite declining population sizes.

480 This presence of unique diversity among and within individual components of the Central Valley
481 population complex underscores the importance of a management strategy that seeks to maintain a robust
482 portfolio at both a phenotypic and genotypic level. While it is important to acknowledge that the (neutral)
483 genetic diversity of a population is not always correlated with functional diversity (Reed & Frankham
484 2001), the variation of genotypes within and among individuals has been demonstrated to be a suitable
485 proxy to predict fitness of individuals and the ability of populations and ecosystems to respond to changes
486 in environmental conditions (Vazquez-Dominguez et al. 1999; Reed & Frankham 2003; Reusch et al.
487 2005; Hoffman et al. 2014). Furthermore, examples from translocation and genetic rescue efforts have
488 demonstrated that heterozygosity and genetic diversity can be more efficient predictors of success than
489 the ability to match (neutral) genotypes as closely as possible to individuals already present in the
490 population (Coleman et al. 2013; Scott et al. 2020). Thus, while the locus that affects run timing is
491 undoubtedly important and of critical conservation concern (Ford et al. 2020), the evolution and
492 accumulation of additional diversity facilitated by differences in run-timing in specific habitats may also
493 be of critical importance for population success (Allendorf et al. 2010). Losing early-run populations
494 therefore runs the risk of losing both early-run alleles (i.e., the ability to recover the early-run phenotype)
495 and the more cryptic yet likely important unique components of genetic diversity harbored among and
496 within migration phenotypes.

497 This loss of diversity and increasing genetic homogenization may be a more important factor driving
498 the loss of the portfolio than demographic synchronization itself (Dedrick & Baskett 2018; Des Roches et
499 al. 2021). Because of their complicated life history, environmental pressures differ widely across salmon
500 life stages such that the genotypes and phenotypes that confer higher survival probability at one life stage
501 do not necessarily translate into the genotypes and phenotypes best matching conditions during a different

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

502 life stage. Additionally, climate change will impact environmental conditions in individual tributaries
503 differently, again necessitating genomic diversity across the Central Valley to allow adaptation to
504 changing conditions (Yates et al. 2008). Important phenotypic traits, including growth, temperature
505 tolerance, and stress responses, are likely polygenic traits, controlled by many loci of small effects, and
506 populations characterized by the presence of a large proportion of polygenic phenotypes are more likely
507 to adapt to new conditions and therefore increase population viability with rapidly changing and
508 fluctuating environmental conditions (Kardos & Luikart 2019).

509 Provided sufficient standing genetic diversity exists and is preserved in the Central Valley, the
510 intraspecific variation present in Central Valley Chinook may allow adaption to changing conditions
511 (Hairston et al. 2005; Richardson et al. 2014; Messer et al. 2016). Even though anthropogenic impacts
512 significantly alter the composition and structure of both neutral and functional diversity at a genetic level,
513 the conservation of intraspecific genetic diversity is frequently overlooked (Laikre et al. 2010; Des
514 Roches et al. 2021) despite serving as the fundamental building block of biodiversity. Indeed, finding
515 unique diversity within each run/tributary group comprising the population complex of Chinook salmon
516 in California's Central Valley underscores the importance of monitoring intraspecific genomic diversity at
517 multiple levels (across and within locations and life history phenotypes) to inform conservation and
518 management policies that counteract genetic homogenization and conserve biocomplexity at a genomic
519 level. Additionally, our results highlight the necessity of managing population complexes as a whole, with
520 a focus on maintaining biocomplexity at multiple scales, as an important factor determining the resilience
521 to changing environmental pressures.

522 **Funding**

523 This work was supported by a grant to MHM from the Delta Stewardship Council #18209.

524 **Acknowledgements**

525 The authors would like to thank the members of the Meek Lab for helpful discussions and Melinda
526 Baerwald for helpful feedback on the manuscript. Additionally, we would like to thank Pascale Goertler
527 for important discussions related to the project.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

528 **Data Availability**

529 Data for this study are available from Dryad (https://doi.org/10.5061/dryad.tht76hdvt).

530 Supplementary Material 1 and 2 contain fully reproducible code supporting the analysis; research

531 compendium available at *https://github.com/sjoleary/ONC_GenDiv*.

532 **References**

533 Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. Nature
534 Reviews Genetics **11**:697–709. Nature Publishing Group. Available from
535 http://www.nature.com/doifinder/10.1038/nrg2844.
536 Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC. 2018. Microhaplotypes provide increased
537 power from short-read DNA sequences for relationship inference. Molecular Ecology Resources
538 **18**:296–305. Blackwell Publishing Ltd. Available from http://doi.wiley.com/10.1111/1755-
539 0998.12737 (accessed April 20, 2021).
540 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach
541 to multiple testing. Journal of the Royal Statistical Society **57**:289–300. WileyRoyal Statistical
542 Society. Available from http://www.jstor.org/stable/2346101%5Cnhttp://about.jstor.org/terms
543 (accessed August 27, 2018).
544 Bertram J, Masel J. 2019. Different mechanisms drive the maintenance of polymorphism at loci subject to
545 strong versus weak fluctuating selection. Evolution **73**:883–896. Society for the Study of Evolution.
546 Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.13719 (accessed December 29,
547 2020).
548 California-HSRG. 2012. California Hatchery Scientific Review Group - California Hatchery Review
549 Report. Prepared for the US Fish and Wildlife Service and Pacific States Marine Fisheries
550 Commission.
551 Carlson SM, Satterthwaite WH. 2011. Weakened portfolio effect in a collapsed salmon population
552 complex. Canadian Journal of Fisheries and Aquatic Sciences **68**:1579–1589. NRC Research Press.
553 Available from https://cdnsciencepub.com/doi/abs/10.1139/F2011-084 (accessed December 21,
554 2020).
555 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks : Building and
556 Genotyping Loci De Novo From Short-Read Sequences. G3: Genes|Genomes|Genetics **1**:171–182.
557 CBD. 2011. Strategic plan for biodiversity 2011–2020, including Aichi Biodiversity Targets.Nagoya,
558 Japan: The Convention on Biological Diversity.
559 Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Pages i884–

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

560 i890 Bioinformatics. Narnia. Available from
561 https://academic.oup.com/bioinformatics/article/34/17/i884/5093234 (accessed August 27, 2019).
562 Christensen KA, Leong JS, Sakhrani D, Biagi CA, Minkley DR, Withler RE, Rondeau EB, Koop BF,
563 Devlin RH. 2018. Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. PLoS
564 ONE **13**:e0195461. Public Library of Science. Available from
565 http://dx.plos.org/10.1371/journal.pone.0195461 (accessed October 24, 2019).
566 Coleman RA, Weeks AR, Hoffmann AA. 2013. Balancing genetic uniqueness and genetic variation in
567 determining conservation and translocation strategies: a comprehensive case study of threatened
568 dwarf galaxias, *Galaxiella pusilla* (Mack) (Pisces: Galaxiidae). Molecular Ecology **22**:1820–1835.
569 John Wiley & Sons, Ltd. Available from http://doi.wiley.com/10.1111/mec.12227 (accessed March
570 25, 2021).
571 Dedrick AG, Baskett ML. 2018. Integrating genetic and demographic effects of connectivity on
572 population stability: The case of hatchery trucking in salmon. American Naturalist **192**:E62–E80.
573 University of Chicago Press. Available from
574 https://www.journals.uchicago.edu/doi/10.1086/697581 (accessed December 29, 2020).
575 Des Roches S, Pendleton LH, Shapiro B, Palkovacs EP. 2021, March 1. Conserving intraspecific
576 variation for nature's contributions to people. Nature Research. Available from
577 https://www.nature.com/articles/s41559-021-01403-5 (accessed March 19, 2021).
578 Duelli P, Obrist MK. 2003. Biodiversity indicators: The choice of values and measures. Agriculture,
579 Ecosystems and Environment **98**:87–98. Elsevier.
580 Ford M et al. 2020. Reviewing and synthesizing the state of the science regarding associations between
581 adult run timing and specific genotypes in Chinook salmon and steelhead:0–47. Available from
582 https://doi.org/10.25923/mv3n-zb79 (accessed January 18, 2021).
583 Frankham R, Bradshaw CJA, Brook BW. 2014, February 1. Genetics in conservation management:
584 Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses.
585 Elsevier.
586 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. Plos One
587 **11**:e0151651.
588 Gosselin T, Anderson EC, Bradbury IR. 2016. assigner: Assignment Analysis with GBS/RAD Data using
589 R. R package.
590 Griffiths JR et al. 2014. Performance of salmon fishery portfolios across western North America. Journal
591 of Applied Ecology **51**:1554–1563. Blackwell Publishing Ltd. Available from
592 https://onlinelibrary.wiley.com/doi/10.1111/1365-2664.12341 (accessed December 21, 2020).
593 Gulisija D, Kim Y. 2015. Emergence of long-term balanced polymorphism under cyclic selection of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

594 spatially variable magnitude. Evolution **69**:979–992. Society for the Study of Evolution. Available
595 from http://doi.wiley.com/10.1111/evo.12630 (accessed December 29, 2020).
596 Hairston NG, Ellner SP, Geber MA, Yoshida T, Fox JA. 2005, October 1. Rapid evolution and the
597 convergence of ecological and evolutionary time. John Wiley & Sons, Ltd. Available from
598 https://onlinelibrary.wiley.com/doi/full/10.1111/j.1461-0248.2005.00812.x (accessed December 29,
599 2020).
600 Hilborn R, Quinn TP, Schindler DE, Rogers DE. 2003. Biocomplexity and fisheries sustainability.
601 Proceedings of the National Academy of Sciences **100**:6564–6568. National Academy of Sciences.
602 Available from http://www.pnas.org/cgi/doi/10.1073/pnas.1037274100.
603 Hoban S et al. 2020. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity
604 Framework must be improved. Biological Conservation **248**:108654. Elsevier Ltd.
605 Hoelzel AR, Bruford MW, Fleischer RC. 2019, February 8. Conservation of adaptive potential and
606 functional diversity. Springer. Available from https://doi.org/10.1007/s10592-019-01151-x
607 (accessed August 7, 2020).
608 Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, Lacy RC, Dasmahapatra KK. 2014.
609 High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of
610 the National Academy of Sciences of the United States of America **111**:3775–3780. National
611 Academy of Sciences. Available from www.pnas.org/cgi/doi/10.1073/pnas.1318945111 (accessed
612 December 25, 2020).
613 Hollenbeck CM, Portnoy DS, Gold JR. 2016. A method for detecting recent changes in contemporary
614 effective population size from linkage disequilibrium at linked and unlinked loci. Heredity **117**:207–
615 216. Nature Publishing Group. Available from
616 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5026753/ (accessed July 20, 2020).
617 Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method
618 for the analysis of genetically structured populations. BMC Genetics **11**:94. BioMed Central Ltd.
619 Available from http://bmcgenet.biomedcentral.com/articles/10.1186/1471-2156-11-94.
620 Kaeuffer R, Réale D, Coltman DW, Pontier D. 2007. Detecting population structure using STRUCTURE
621 software: Effect of background linkage disequilibrium. Heredity **99**:374–380. Nature Publishing
622 Group. Available from www.nature.com/hdy (accessed June 30, 2021).
623 Kalinowski S. 2002. How many alleles per locus should be used to estimate genetic distances? Heredity
624 **88**:62–65. Blackwell Publishing Ltd. Available from www.nature.com/hdy (accessed June 30,
625 2021).
626 Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr : an R package for genetic analysis of populations
627 with clonal, partially clonal, and/or sexual reproduction. PeerJ **2**:e281. PeerJ Inc. Available from

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

628 https://peerj.com/articles/281 (accessed August 27, 2018).
629 Kardos M, Luikart G. 2019. The genetic architecture of fitness drives population viability during rapid
630 environmental change. bioRxiv:660803. Cold Spring Harbor Laboratory. Available from
631 https://www.biorxiv.org/content/10.1101/660803v1.abstract (accessed August 7, 2020).
632 Laikre L, Schwartz MK, Waples RS, Ryman N. 2010, September 1. Compromising genetic diversity in
633 the wild: Unmonitored large-scale release of plants and animals. Elsevier Current Trends.
634 Langin K. 2018. A different animal. Science **360**:590–592.
635 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
636 Bioinformatics **25**:1754–1760.
637 Mable BK. 2019, February 15. Conservation of adaptive potential and functional diversity: integrating old
638 and new approaches. Springer Netherlands. Available from https://doi.org/10.1007/s10592-018-
639 1129-9 (accessed December 22, 2020).
640 Mckinney GJ, Seeb LW, Larson WA, Gomez-Uchida D, Limborg MT, Brieuc MSO, Everett M V., Naish
641 KA, Waples RK, Seeb JE. 2016. An integrated linkage map reveals candidate genes underlying
642 adaptive variation in Chinook salmon (Oncorhynchus tshawytscha). Molecular Ecology Resources
643 **16**:769–783. John Wiley & Sons, Ltd. Available from http://doi.wiley.com/10.1111/1755-
644 0998.12479 (accessed July 20, 2020).
645 Meek MH, Baerwald MR, Stephens MR, Goodbla A, Miller MR, Tomalty KMH, May B. 2016.
646 Sequencing improves our ability to study threatened migratory species: Genetic population
647 assignment in California's Central Valley Chinook salmon. Ecology and Evolution **6**:7706–7716.
648 John Wiley & Sons, Ltd. Available from http://doi.wiley.com/10.1002/ece3.2493 (accessed
649 November 18, 2019).
650 [dataset] Meek MH, Stephens MR, Goodbla A, May BP, Baerwald MR. 2019a. Identifying hidden
651 biocomplexity and genomic diversity in Chinook salmon, an imperiled species with a history of
652 anthropogenic influence. Dryad **doi:10.506**. Available from
653 https://datadryad.org/stash/dataset/doi:10.5061/dryad.tht76hdvt.
654 Meek MH, Stephens MR, Goodbla A, May BP, Baerwald MR. 2019b. Identifying hidden biocomplexity
655 and genomic diversity in Chinook salmon, an imperiled species with a history of anthropogenic
656 influence. Canadian Journal of Fisheries and Aquatic Sciences:cjfas-2019-0171. Available from
657 http://www.nrcresearchpress.com/doi/10.1139/cjfas-2019-0171 (accessed November 18, 2019).
658 Messer PW, Ellner SP, Hairston NG. 2016, July 1. Can Population Genetics Adapt to Rapid Evolution?
659 Elsevier Ltd.
660 Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE, Palti Y, Doe CQ, Thorgaard GH. 2012. A
661 conserved haplotype controls parallel adaptation in geographically distant salmonid populations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

662 Molecular Ecology **21**:237–249. Wiley/Blackwell (10.1111). Available from
663 http://doi.wiley.com/10.1111/j.1365-294X.2011.05305.x (accessed September 15, 2018).
664 Mimura M et al. 2017, February 1. Understanding and monitoring the consequences of human impacts on
665 intraspecific variation. Wiley-Blackwell. Available from
666 https://onlinelibrary.wiley.com/doi/full/10.1111/eva.12436 (accessed December 21, 2020).
667 Morin PA, Martien KK, Taylor BL. 2009. Assessing statistical power of SNPs for population structure
668 and conservation studies. Molecular Ecology Resources **9**:66–73. John Wiley & Sons, Ltd.
669 Available from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1755-0998.2008.02392.x
670 (accessed April 20, 2021).
671 Narum SR, Genova A Di, Micheletti SJ, Maass A. 2018. Genomic variation underlying complex life-
672 history traits revealed by genome sequencing in Chinook salmon. Proceedings of the Royal Society
673 B: Biological Sciences **285**:20180935. The Royal Society. Available from
674 https://royalsocietypublishing.org/doi/10.1098/rspb.2018.0935 (accessed November 20, 2019).
675 National Marine Fisheries Service. 2005. Endangered and threatened species: final listing determinations
676 for 16 ESUs of west coast salmon, and final 4(d) protective regulations for threatened salmonid
677 ESUs. Federal Register **70**:37160–37204. Available from
678 https://www.federalregister.gov/documents/2005/06/28/05-12351/endangered-and-threatened-
679 species-final-listing-determinations-for-16-esus-of-west-coast-salmon-and (accessed December 21,
680 2020).
681 Nei M. 1987. Molecular Evolutionary Genetics. Page Molecular Evolutionary Genetics.
682 O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018. These aren't the loci you're
683 looking for: Principles of effective SNP filtering for molecular ecologists. Molecular Ecology
684 **27**:3193–3206. Wiley/Blackwell (10.1111).
685 Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. 2010, April 1. Conservation genetics in
686 transition to conservation genomics. Elsevier Ltd.
687 Péru N, Dolédec S. 2010. From compositional to functional biodiversity metrics in bioassessment: A case
688 study using stream macroinvertebrate communities. Ecological Indicators **10**:1025–1036. Elsevier.
689 Petchey OL, Gaston KJ. 2002. Functional diversity (FD), species richness and community composition.
690 Ecology Letters **5**:402–411. John Wiley & Sons, Ltd. Available from
691 http://doi.wiley.com/10.1046/j.1461-0248.2002.00339.x (accessed February 19, 2021).
692 Prieto I, Violle C, Barre P, Durand JL, Ghesquiere M, Litrico I. 2015. Complementary effects of species
693 and genetic diversity on productivity and stability of sown grasslands. Nature Plants **1**.
694 Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, Hotaling TJ, Spidle AP, Miller
695 MR. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

696 genomics for informing conservation. Science Advances **3**:e1603198. American Association for the
697 Advancement of Science. Available from
698 https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.1603198 (accessed February 5, 2020).
699 Quinlan AR. 2014. BEDTools: The Swiss-Army tool for genome feature analysis. Current Protocols in
700 Bioinformatics **2014**:11.12.1-11.12.34. John Wiley & Sons, Ltd. Available from
701 http://doi.wiley.com/10.1002/0471250953.bi1112s47 (accessed September 3, 2019).
702 Quinn TP. 2018. The Behavior and Ecology of Pacific Salmon and Trout. University of Washington
703 press.
704 Quinn TP, McGinnity P, Reed TE. 2016, July. The paradox of "premature migration" by adult
705 anadromous salmonid fishes: Patterns and hypotheses. NRC Research Press. Available from
706 http://www.nrcresearchpress.com/doi/10.1139/cjfas-2015-0345 (accessed August 7, 2020).
707 Reed DH, Frankham R. 2001. How closely correlated are molecular and quantitative measures of genetic
708 variation? A meta-analysis. Evolution **55**:1095–1103. John Wiley & Sons, Ltd. Available from
709 http://doi.wiley.com/10.1111/j.0014-3820.2001.tb00629.x (accessed August 7, 2020).
710 Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. Conservation Biology
711 **17**:230–237. Blackwell Science, Inc.
712 Reusch TBH, Ehlers A, Hämmerli A, Worm B. 2005. Ecosystem recovery after climatic extremes
713 enhanced by genotypic diversity. Proceedings of the National Academy of Sciences of the United
714 States of America **102**:2826–2831. National Academy of Sciences. Available from
715 www.pnas.orgcgidoi10.1073pnas.0500008102 (accessed December 25, 2020).
716 Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014, March 1. Microgeographic adaptation and the
717 spatial scale of evolution. Elsevier Current Trends. Available from
718 https://www.sciencedirect.com/science/article/pii/S0169534714000159 (accessed September 15,
719 2018).
720 Ryman N, Laikre L. 1991. Effects of Supportive Breeding on the Genetically Effective Population Size.
721 Conservation Biology **5**:325–329. John Wiley & Sons, Ltd (10.1111). Available from
722 http://doi.wiley.com/10.1111/j.1523-1739.1991.tb00144.x (accessed December 4, 2019).
723 Schindler DE, Hilborn R, Chasco B, Boatright CP, Quinn TP, Rogers LA, Webster MS. 2010. Population
724 diversity and the portfolio effect in an exploited species. Nature **465**:609–612.
725 Scott PA, Allison LJ, Field KJ, Averill-Murray RC, Bradley Shaffer H. 2020. Individual heterozygosity
726 predicts translocation success in threatened desert tortoises. Science **370**:1086–1089. American
727 Association for the Advancement of Science. Available from http://science.sciencemag.org/
728 (accessed March 25, 2021).
729 Siefert A et al. 2015, December 1. A global meta-analysis of the relative extent of intraspecific trait

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

730 variation in plant communities. Blackwell Publishing Ltd. Available from
731 https://onlinelibrary.wiley.com/doi/full/10.1111/ele.12508 (accessed December 29, 2020).
732 Sinclair-Waters M, Ødegård J, Korsvoll SA, Moen T, Lien S, Primmer CR, Barson NJ. 2020. Beyond
733 large-effect loci: Large-scale GWAS reveals a mixed large-effect and polygenic architecture for age
734 at maturity of Atlantic salmon. Genetics Selection Evolution **52**:9. BioMed Central. Available from
735 https://gsejournal.biomedcentral.com/articles/10.1186/s12711-020-0529-8 (accessed August 7,
736 2020).
737 Svardal H, Rueffler C, Hermisson J. 2015. A general condition for adaptive genetic polymorphism in
738 temporally and spatially heterogeneous environments. Theoretical Population Biology **99**:76–97.
739 Academic Press Inc.
740 Thompson NF, Anderson EC, Clemento AJ, Campbell MA, Pearse DE, Hearsey JW, Kinziger AP, Garza
741 JC. 2020. A complex phenotype in salmon controlled by a simple change in migratory timing.
742 Science **370**:609–613. American Association for the Advancement of Science. Available from
743 http://science.sciencemag.org/ (accessed January 15, 2021).
744 Thompson TQ et al. 2019. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and
745 restoration potential in wild salmon populations. Proceedings of the National Academy of Sciences
746 of the United States of America **116**:177–186. National Academy of Sciences. Available from
747 http://www.ncbi.nlm.nih.gov/pubmed/30514813 (accessed August 7, 2020).
748 Vazquez-Dominguez E, Pinero D, Ceballos G. 1999. Linking Heterozygosity, Demography, and Fitness
749 of Tropical Populations of Liomys pictus. Journal of Mammalogy **80**:810–822. Allen Press Inc.
750 Available from https://academic.oup.com/jmammal/article-lookup/doi/10.2307/1383250 (accessed
751 December 25, 2020).
752 Waples RS, Lindley ST. 2018. Genomics and conservation units: The genetic basis of adult migration
753 timing in Pacific salmonids. Evolutionary Applications **11**:1518–1526. Wiley-Blackwell. Available
754 from http://doi.wiley.com/10.1111/eva.12687 (accessed December 22, 2020).
755 Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure.
756 Evolution **38**:1358. Society for the Study of Evolution. Available from
757 http://www.jstor.org/stable/2408641?origin=crossref (accessed August 27, 2018).
758 Williams JG. 2006. Central Valley Salmon: A Perspective on Chinook and Steelhead in the Central
759 Valley of California. San Francisco Estuary and Watershed Science **4**. San Francisco Estuary and
760 Watershed Science. Available from https://escholarship.org/uc/item/21v9x1t7 (accessed December
761 21, 2020).
762 Williamson KS, May B. 2005. Homogenization of Fall-Run Chinook Salmon Gene Pools in the Central
763 Valley of California, USA. North American Journal of Fisheries Management **25**:993–1009. Taylor

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

764 & Francis Group. Available from http://doi.wiley.com/10.1577/M04-136.1 (accessed August 7,
765 2020).
766 Willis SC, Hollenbeck CM, Puritz JB, Gold JR, Portnoy DS. 2017. Haplotyping RAD loci: an efficient
767 method to filter paralogs and account for physical linkage. Molecular Ecology Resources.
768 Yates D, Galbraith H, Purkey D, Huber-Lee A, Sieber J, West J, Herrod-Julius S, Joyce B. 2008. Climate
769 warming, water storage, and Chinook salmon in California's Sacramento Valley. Climatic Change
770 **91**:335–350. Springer. Available from https://link.springer.com/article/10.1007/s10584-008-9427-8
771 (accessed December 29, 2020).
772

773

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

774 **Tables**

775 Table 1: Sample sizes and effective population size Ne for all run/tributary groups. Negative point
776 estimates occur when LD attributed to sampling variation is larger than LD attributed to drift. This can
777 either be interpreted as an "infinite" population size (drift is negligible) or due to sample size being too
778 small for an accurate estimate (LD attributed to sampling variation is a function of sample size).

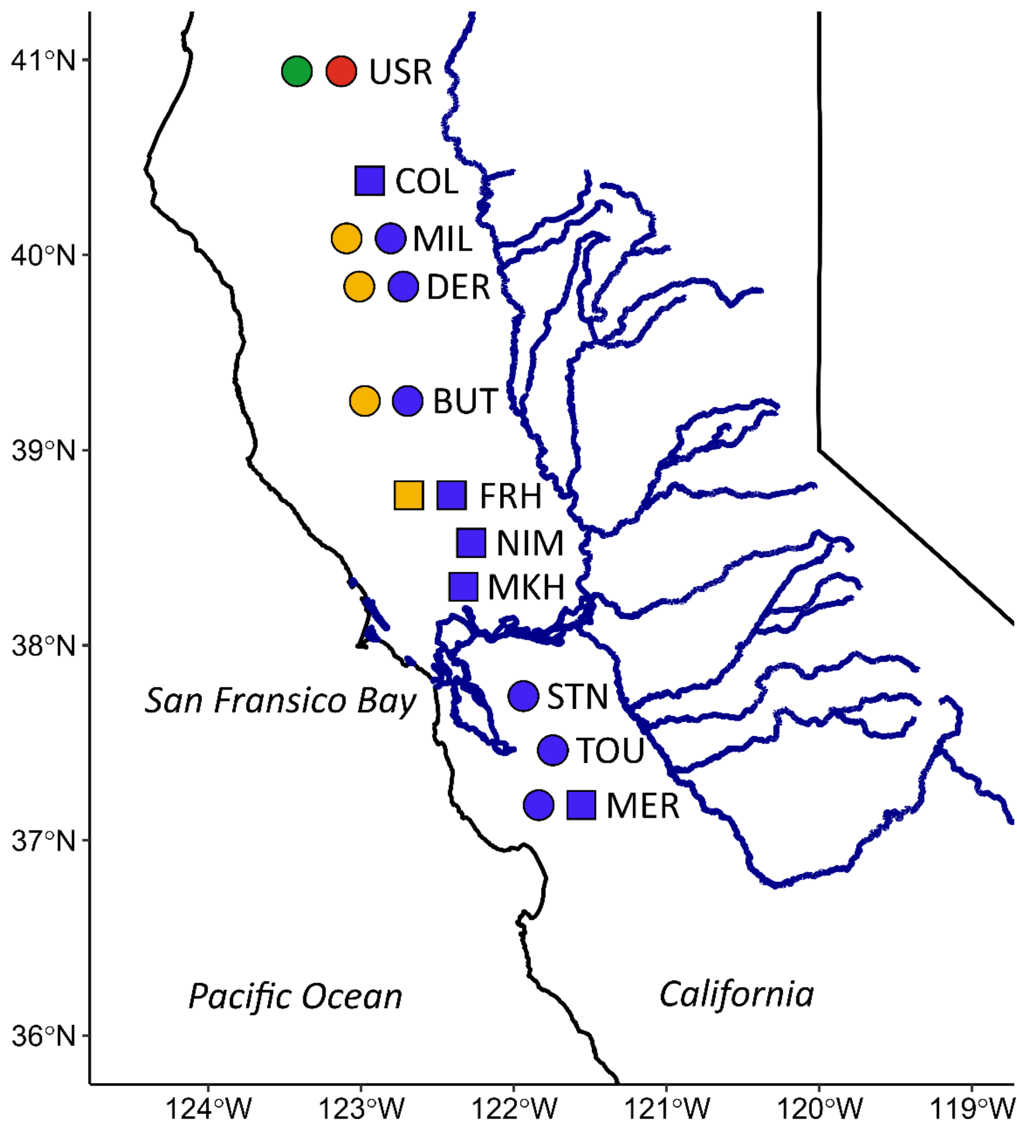
Run	Tributary/Hatchery	Sample Size	*Ne* (95% CI)
Fall	Coleman Hatchery	30	615 (594; 637)
	Mill Creek	20	*-6,801 (-56,952; -3,618)*
	Deer Creek	15	*-3,939 (-9,323; -2,497)*
	Butte Creek	21	3,802 (2,708; 6,378)
	Feather River Hatchery	27	5,172 (3,823; 7,988)
	Nimbus Hatchery	30	7,432 (5,147 – 13,351)
	Mokelumne River Hatchery	28	4,106 (3,207; 5,704)
	Stanislaus River	23	2,567 (2,040; 3,460)
	Tuolumne River	30	6,353 (4,383;11,538)
	Merced River Hatchery	15	1,125 (940; 1,399)
	Merced River	31	11,436 (6,768; 36,804)
Late-Fall	Upper Sacramento River	21	3,114 (2,398; 4,435)
Winter	Upper Sacramento River	26	174 (171; 177)
Spring	Mill Creek	16	85 (84; 86)
	Deer Creek	27	188 (185; 191)
	Butte Creek	19	561 (529; 597)
	Feather River Hatchery	7	*-94 (-98; -91)*

779

780

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

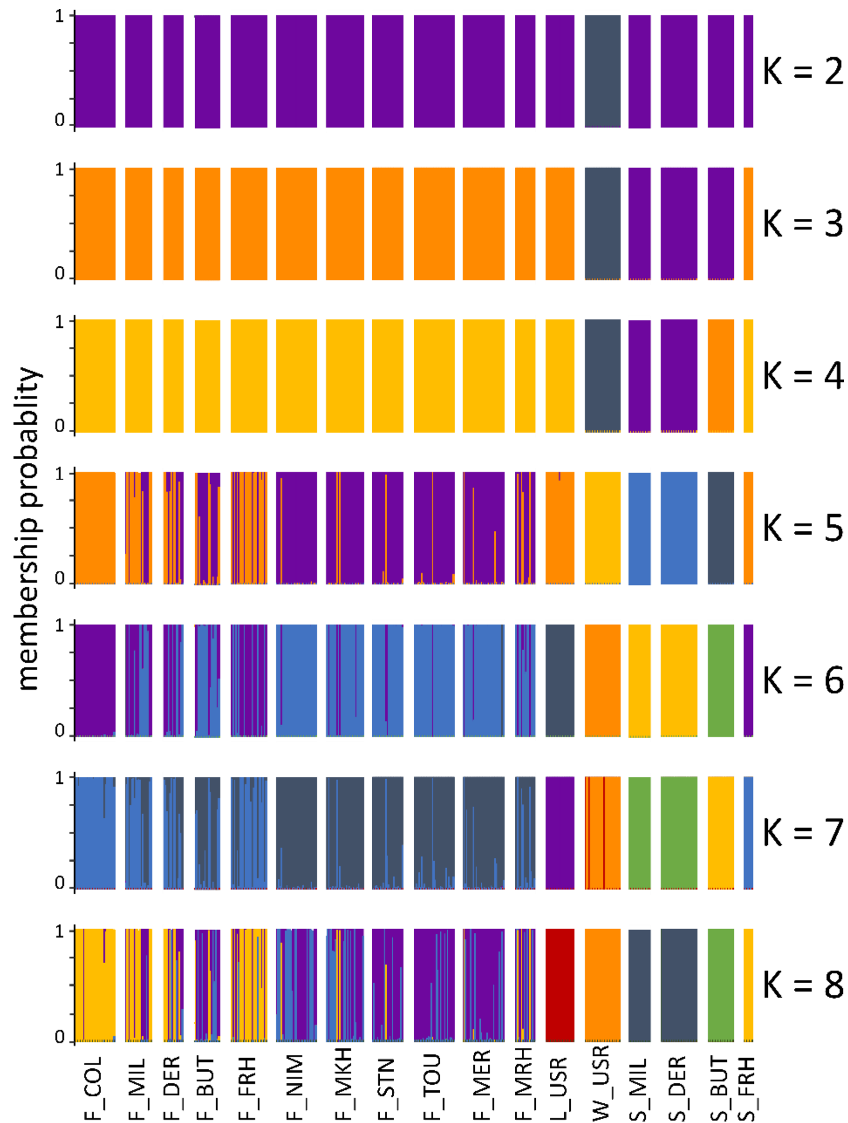
781 **Figures**



782

783 Figure 1: Tributaries of the California's Central Valley sampled for this study. Hatchery individuals are
784 represented by squares and wild populations by circles. Colors represent the sampled run type at each
785 location (Spring = green, Fall = blue, Late-Fall = red, Winter = yellow). Abbreviations for tributaries used
786 throughout: USR (Upper Sacramento River), COL (Coleman Hatchery/Battle Creek), MIL (Mill Creek),
787 DER (Deer Creek), BUT (Butte Creek), FRH (Feather River Hatchery), NIM (Nimbus
788 Hatchery/American River), MKH (Mokelumne River Hatchery), STN (Stanislaus River), TOU
789 (Tuolumne River), MER (Merced River).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

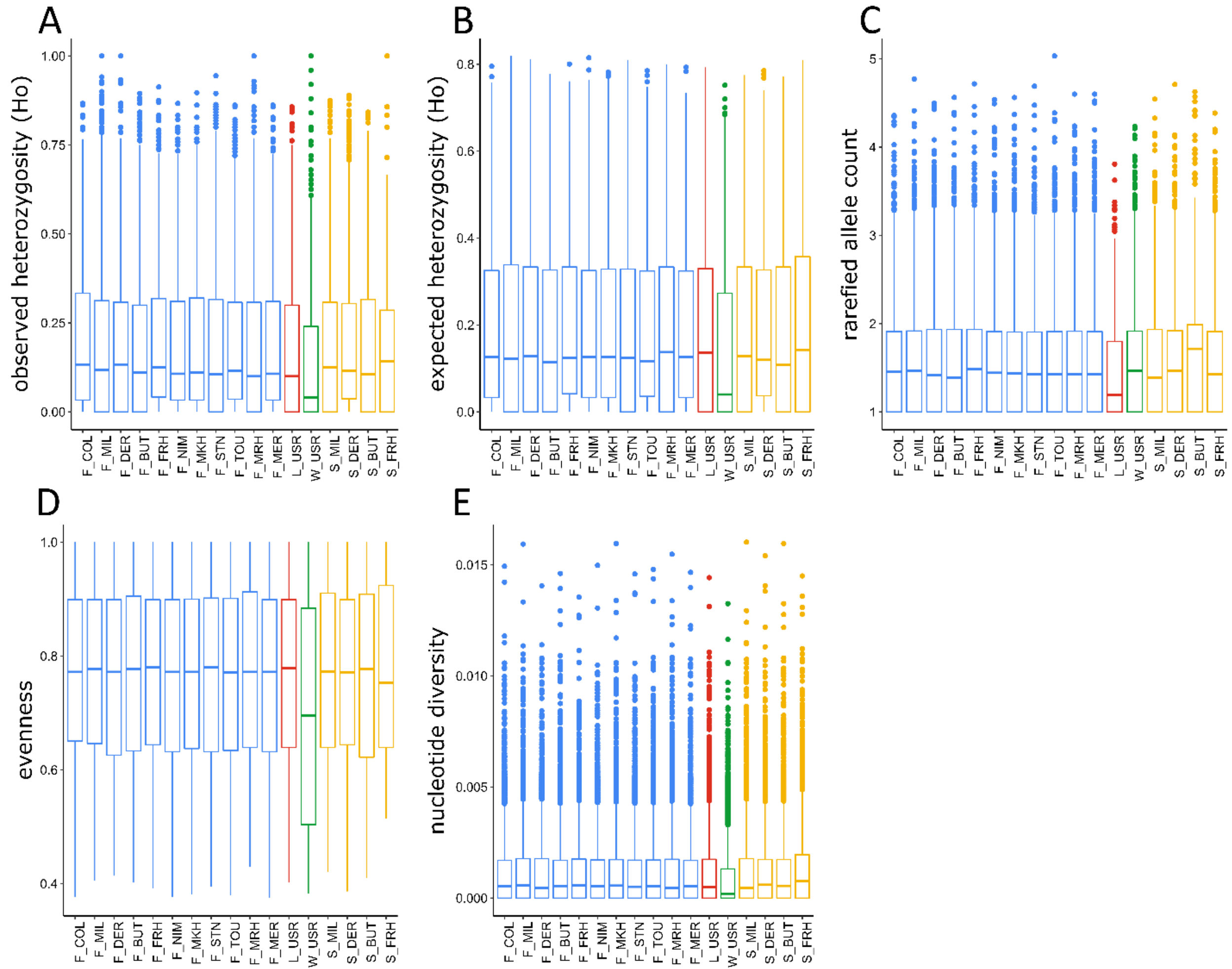


790

791 Figure 2: Membership probability of each individual to clusters identified using k-means hierarchical

792 clustering for K = 2 – 8.

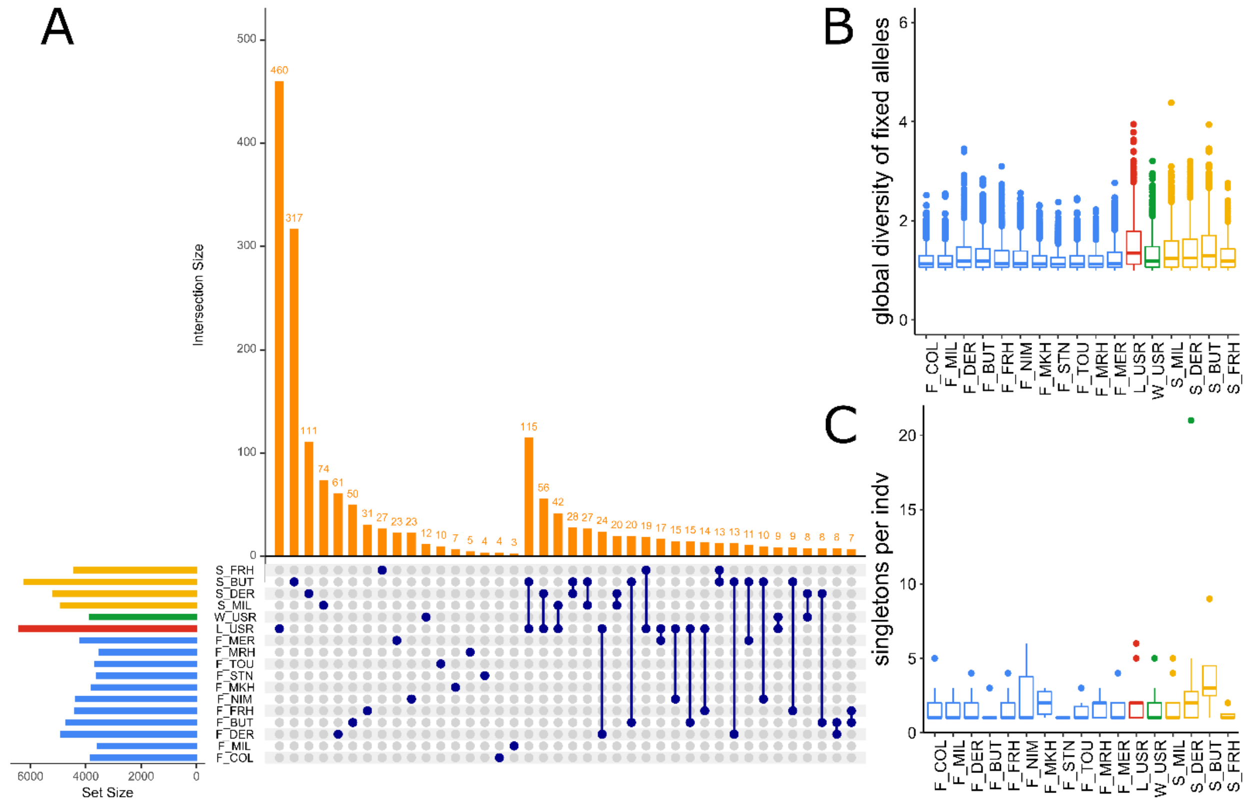
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



793

794 Figure 3: Assessment of the distribution of genomic diversity for individuals grouped by run type within
795 tributaries using heterozygosity-based parameters (A. observed heterozygosity Ho, B. expected
796 heterozygosity Hs), measures of allele diversity (C. Allelic richness, D. Evenness), and sequence-based
797 parameters (E. observed nucleotide diversity). Fall individuals are in blue, Late-Fall in red, Winter
798 depicted in green, and Spring in yellow.

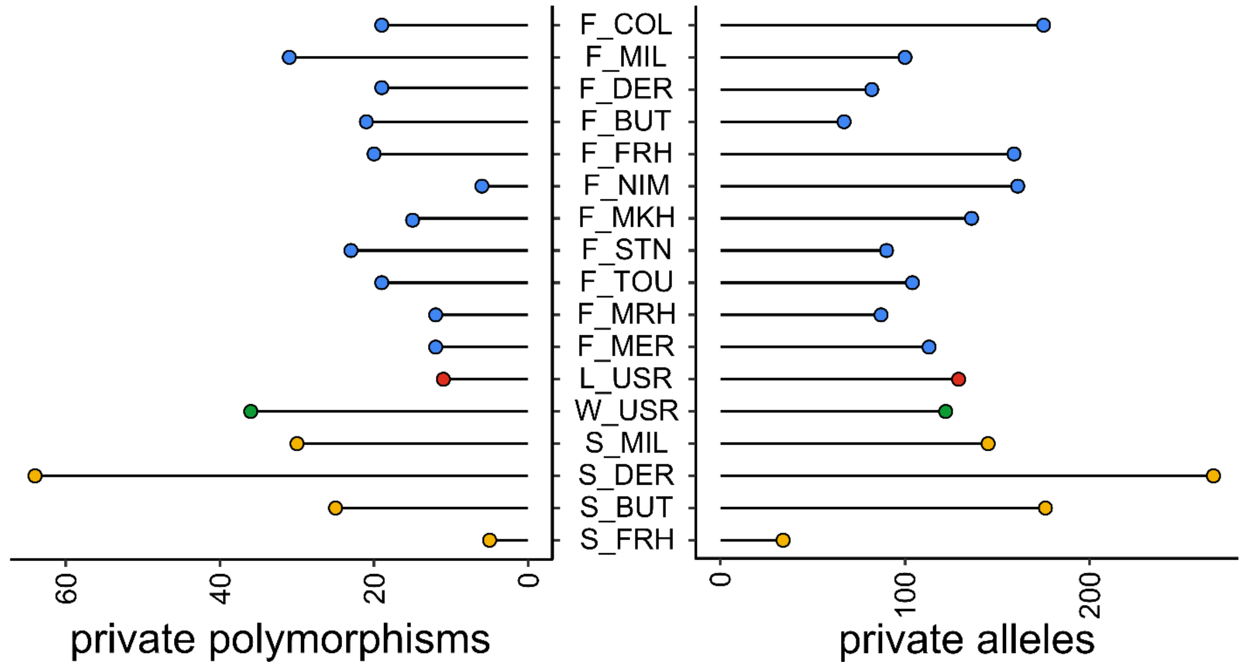
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



799

800 Figure 4: Assessment of fixed loci and singletons. A. Comparisons of fixed loci across run/tributary
801 groups. The set size (horizontal bars) indicate the total number of fixed loci in a given group, the intersect
802 size (vertical orange bars) the number of loci fixed in a single group (single blue dot) or in two (blue dots
803 connected by line). B. Distribution of global allelic richness of loci fixed in a given group. C. Distribution
804 of the number of singletons per individual for each run/tributary group. Fall individuals are in blue, Late-
805 Fall in red, Winter depicted in green, and Spring in yellow.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437213; this version posted August 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



806

807 Figure 5: Assessment of unique genomic diversity. Left axis indicates the number of private

808 polymorphisms (loci only variable in a single group), right axis indicates the number of private alleles

809 (alleles found only in a single group) for individuals grouped by run and tributary. Fall individuals are in

810 blue, Late-Fall in red, Winter depicted in green, and Spring in yellow.

811 **Supplementary material**

812 *Note: supplementary files will not render properly during conversion to pdf for review – html files that*

813 *can be opened in any browser can be downloaded from the research compendium at*

814 *https://github.com/sjoleary/ONC_GenDiv*.

815 Supplementary Material 1 (OLeary_SupMat_Genotyping.html): Data acquisition and processing

816 (Genotyping).

817 Supplementary Material 2 (OLeary_SupMat_Genotyping.html): Standalone extended methods and results

818 describing the assessment of population differentiation and patterns of genomic diversity. Contains R

819 code used for data analysis and supplementary figures and tables cited in the manuscript.