

Quantifying Impacts of an Environmental Intervention Using Environmental DNA: Supplemental Text 1

³ Elizabeth Andruszkiewicz Allan, Ryan P. Kelly, Erin D’Agnese,
⁴ Maya Garber-Yonts, Megan Shaffer, Zachary Gold, Andrew O. Shelton

5 2023

⁶ Supplemental material for “Quantifying Impacts of an Environmental Intervention Using Environmental
⁷ DNA” in *Ecological Applications*

8 Field Sampling

9 Site selection and study design

10 There were two culverts in the treatment creek (Padden) that were suspected to be partially impassable and
11 thus was removed and replaced during the course of the study; one of the control creeks had a bridge, which
12 allowed fish passage (Portage), one control creek had a culvert classified as having limited fish passability
13 (Squalicum), and two control creeks had culverts classified as preventing fish passage (Barnes and Chuckanut)
14 (Washington Department of Fish and Wildlife 2019). These creeks were chosen due to their comparable size,
15 flow, watersheds, and species presumed to be present to constrain as many ecological variables as possible.

16 Distance between sites and flow variability at sites

¹⁷ The average distance between upstream and downstream sampling within a creek was about 160 m; the
¹⁸ largest distance between downstream and upstream sampling was at Barnes Creek, which was approximately
¹⁹ 330 m, whereas the shortest distance between sampling was at Squalicum Creek at approximately 66 m.

Over the course of the year, flow within each creek varied. USGS flow gauges were located in three of the five creeks, relatively nearby to the sampling locations (Figure S1). The closest gauge to sampling locations was Padden Creek (~1.5 km); the gauge at Chuckanut Creek was ~5.5 km and the gauge at Squalicum Creek was ~7.9 km away (calculated using the Haversine distance in R).

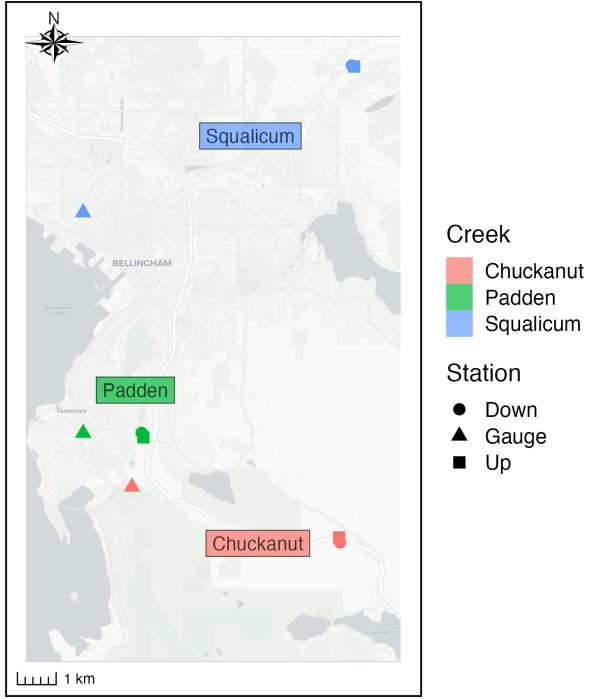


Figure S1. Location of flow gauges compared to sampling locations for Chuckanut, Padden, and Squalicum Creeks.

24 The flow meters at Squalicum Creek and Chuckanut Creek were offline from November 2021 for the remainder
 25 of the sampling period. The highest discharge seen during the course of the study from January to November
 26 2021 occurred in November 2021 at Squalicum Creek. The mean discharge in each creek was: $0.42 \text{ m}^3/\text{s}$ in
 27 Padden, $0.29 \text{ m}^3/\text{s}$ in Chuckanut, and $1.14 \text{ m}^3/\text{s}$ in Squalicum Creek. The lowest discharge registered by the
 28 flow meters is $0.0028 \text{ m}^3/\text{s}$, which occurred 8.5%, 1.6%, and 0.78% of the time in Padden, Chuckanut, and
 29 Squalicum, respectively.

30 Due to the lack of flow data in Squalicum and Chuckanut Creeks from November 2021 to February 2022, we
 31 used historical data from the three flow gauges to calculate the average discharge for each day of the year
 32 from about 2015-2021 (Figure S2). We then used the value for the day of the year that we sampled in either
 33 2021 or 2022 when the gauges were offline. For consistency, we also did this at Padden Creek despite the
 34 gauge there being online for the entire sampling period.

35 We compared the different ways one could use flow data to correct the eDNA concentrations. We included
 36 (1) the value from the closest time point from the gauge to the time point sampling, (2) the average flow
 37 on the day of sampling from the gauge, (3) the monthly average for the month of sampling from the gauge,
 38 and (4) the correction factor approach. For (4), the values for Padden Creek represent the same as (1) for
 39 Padden Creek, the value from the closest time point in the gauge to the time point of sampling. The values

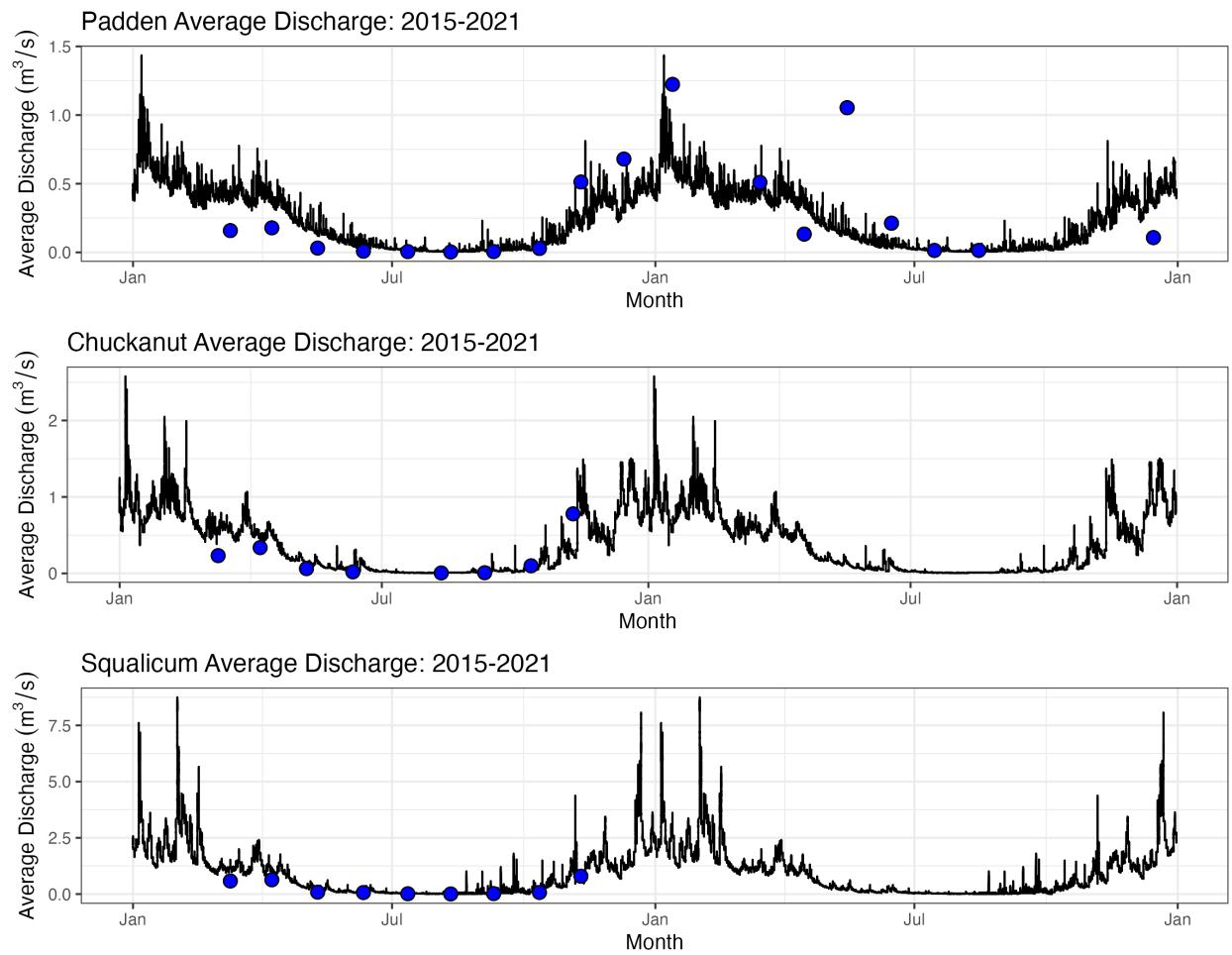


Figure S2. Daily average discharge from 2015-2020 in creeks from USGS gauges. Blue dots show the discharge time of sampling during the course of this study (2021-2022) for time points where gauges were online (note the missing data points after December through February in Chuckanut and Squalicum Creeks).

for Chuckanut and Squalicum Creek are based on the the correction factor from Padden Creek. First, five years of historical data (2015-2020) were used to find monthly averages for flow rates for each creek. Because the gauges in Squalicum and Chuckanut Creeks stopped metering in 2021, we solved for the ratio of the monthly average of each of those creeks to Padden Creek. Then, we used the closest values from Padden Creek (1) and multiplied by the monthly correction factor from the 5 years of historical data to find a value for Squalicum and Chuckanut Creeks to use for the year of sampling (2021-2022). For all three creeks, we demonstrate the relatively small changes in discharge depending on which way flow data were used (Figure S3). Though in the course of sampling, the discharge in Padden Creek ranged from no metered flow to 23 m^3/s , the discharge on the dates of sampling only reached a maximum of 1.3 m^3/s . For sites with no metered flow, half of the minimum verified discharge of the flow gauge was used (0.0014 m^3/s).

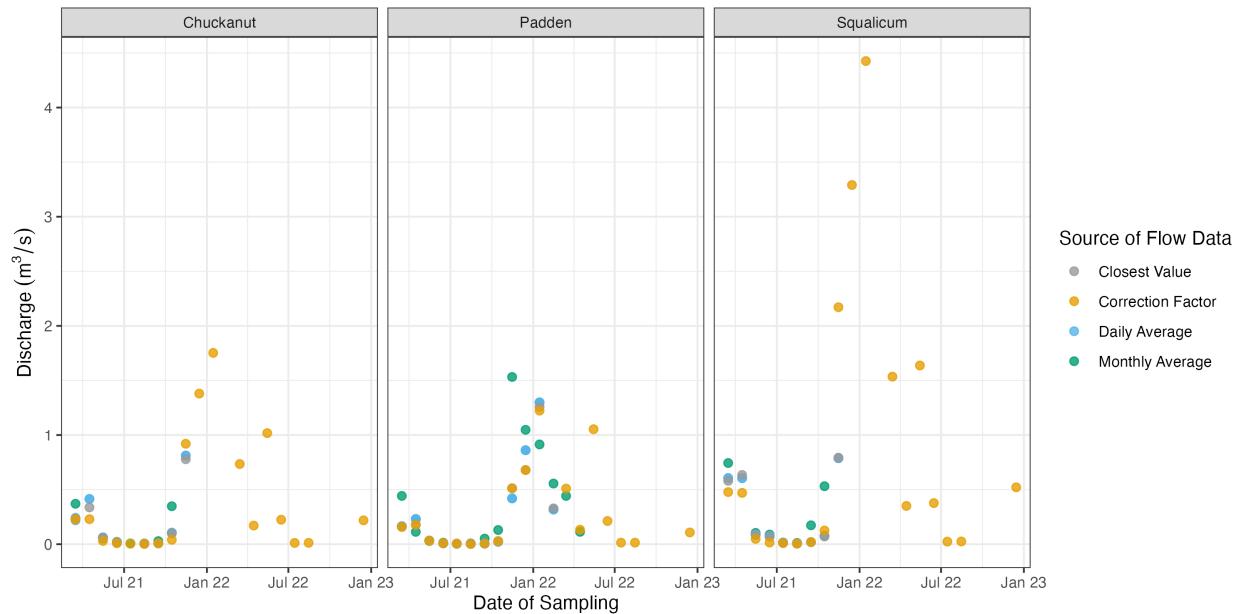


Figure S3. Comparison of different ways flow data can be used to correct eDNA concentrations. In the main text, the correction factor is used. Note that for Padden Creek, the “Correction Factor” method is the same as the “Closest Value” method. Also note that for Chuckanut and Squalicum Creeks, no data exist for closest value, daily average, or monthly average after November 2021 when the gauges went offline.

50 Construction and Fish Exclusion

Fish were excluded from Padden Creek on August 30th, 2021 in preparation for the stream to be diverted on September 9th, 2021 and the diversion was removed October 7th, 2021. Water sampling occurred on September 10th, 2021, the day after the diversion, and on October 12, 2021, just 5 days after reconnecting the stream (Figure S4).

55 **Stocking in Lake Padden**

56 Padden Lake has historically been stocked with hatchery fish by the Washington Department of Fish and
57 Wildlife (Figure S4). Rainbow trout (*O. mykiss*) and occasionally cutthroat trout (*O. clarkii*) and kokanee
58 salmon (*O. nerka*) are stocked in Lake Padden, approximately 1.5 km upstream of the sampling sites. During
59 the course of the study, rainbow trout were stocked in April 2021 and April 2022, kokanee salmon were
60 stocked in May 2021 and October 2022, and cutthroat trout were stocked in November 2022 (Figure S4).
61 However, despite the stocking of 30,000 kokanee salmon in May in Lake Padden, *O. nerka* was not detected
62 by metabarcoding in May 2021 or at any point in 2021 until November (see main text Results). Importantly,
63 this suggests that Lake Padden is far enough upstream that the eDNA signal at the sampling sites by the
64 culvert is not a result of stocking the lake 1.5 km upstream (see main text Discussion for more information).

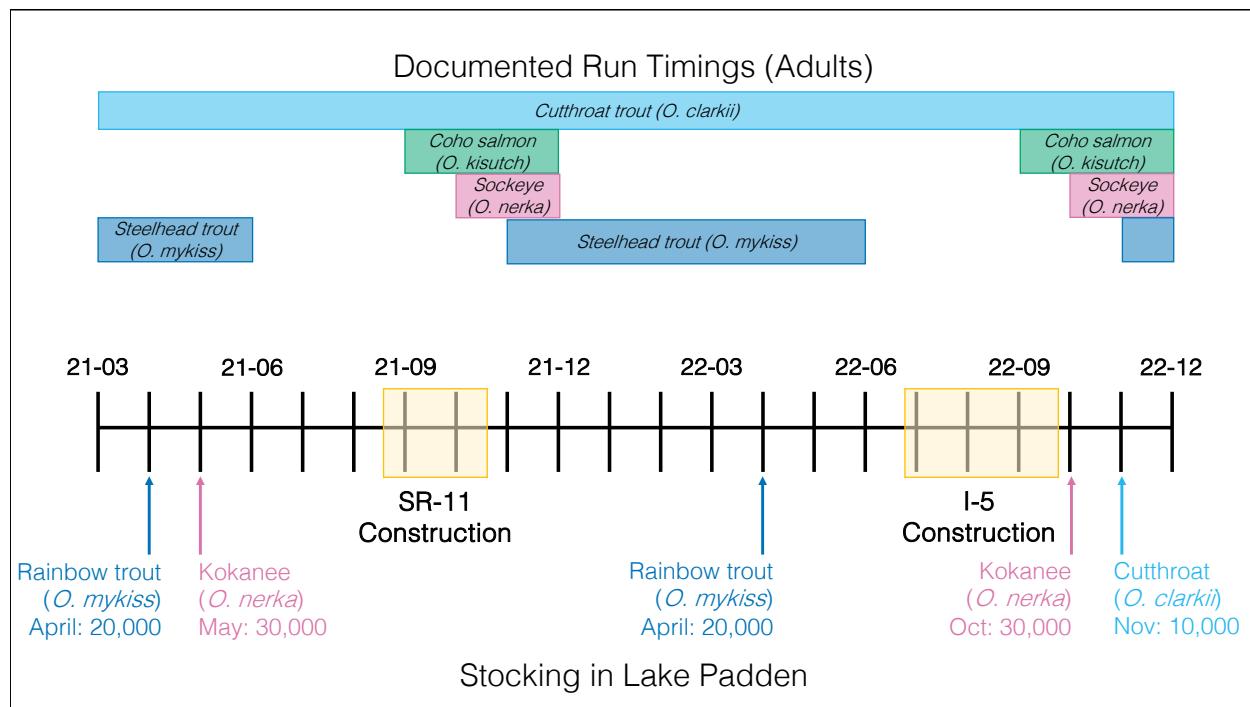


Figure S4. Timeline of runs for migrating species, stocking of Lake Padden, and construction at the intervention site (Padden Creek). Note, the dates of kokanee stocking do not include days, just the month and year. For plotting purposes, they are shown as the 15th.

65 **Water Sampling**

66 Water samples were collected using Smith Root's eDNA Backpack (Thomas et al. 2018), a portable pumping-
67 and-filtering setup set to filter at 1 L/min at 82.7 kPa (12 psi). For most months, a trident sampler was used
68 to collect all 3 biological replicates at the exact same time, for a total sampling time of about 5 minutes.
69 Otherwise, the three replicates were collected consecutively, for a total sampling time of about 15 minutes.

70 Downstream sites were always sampled before upstream sites to ensure no potential DNA was introduced
71 into the stream before sampling. In some samples, less than 2 L of water was filtered due to clogging (mean
72 = 1.95 L).

73 **Laboratory Processing**

74 **DNA Extraction, Amplification, Sequencing**

75 Here, we used mock communities to determine the species-specific amplification efficiencies for each salmonid
76 in the study. Briefly, we constructed five communities with known proportions of starting DNA from different
77 species (total DNA as measured by Qubit). The communities ranged from having a total of 12 to 20
78 species, but six salmonid species were included in all five mock communities to have more information on the
79 amplification efficiencies of salmonids (Supplemental Table 2). We sequenced these communities using the
80 same metabarcoding primers and thermocycling conditions above and then determined the species-specific
81 amplification rates given the discrepancy between the known starting proportion and the proportion of
82 reads after sequencing. The mock community data were then used to correct the sequencing reads from the
83 environmental samples to estimate the starting DNA proportions of each species in environmental samples,
84 which is the metric of interest (Figure 3, green boxes). This is the first application of the model to correct
85 eDNA data from water samples with mock community data as described in Shelton et al. (2022) (see
86 Supplemental Text 2 for more information).

87 All molecular work prior to sequencing was performed at the University of Washington. Bench-tops were
88 cleaned with 10% bleach for 10 minutes and then wiped with 70% ethanol. Molecular work was separated
89 onto pre- and post-PCR benches; all DNA extractions and PCR preparation was conducted on a bench where
90 no PCR product was handled.

91 **DNA Extractions**

92 We followed a protocol developed for extracting DNA off the self-preserving Smith Root filters (Thomas et
93 al. 2019). Filters were removed from their housing with sterile tweezers and cut in half using sterile razor
94 blades. One half was archived and the other half was used for extraction. DNA was extracted from half of
95 each filter using a Qiashredder column (Qiagen, USA) and the DNeasy Blood and Tissue Kit (Qiagen, USA)
96 with an overnight incubation (Thomas et al. (2019)), such that the effective filtering effort was 1 L/sample;
97 the remaining half of each filter was archived at -20°C. Extracts were eluted in 100 µL of molecular grade
98 water, quantified via Qubit (Invitrogen, USA) and stored at -20°C until PCR amplification within 2 months
99 of extraction.

100 **PCR Amplification**

101 For the metabarcoding approach, we targeted a ~186 bp hypervariable region of the mitochondrial DNA 12S
102 rRNA gene for PCR amplification (MiFish; Miya et al. 2015), but using modified primer sequences as given
103 in Praebel and Wangensteen (unpublished; via personal communication) and including the Illumina Nextera
104 overhang sequences for subsequent indexing. The primers used were as follows: F 5' *TCGTCGGCAGCGTCA-*
105 *GATGTGTATAAGAGACAGGCCGGTAAA*ACTCGTGCCAGC 3', R 5' *GTCTCGTGGGCTCGGAGAT-*
106 *GTGTATAAGAGACAGCATAGTGGGTATCTAATCCCAGTTG* 3' (*italics* indicate Nextera overhang).
107 PCR reactions included 10 μ L of 5X Platinum ii Buffer, 0.4 μ L of Platinum ii Taq, 1.25 μ L of 8 mM dNTPS,
108 1.25 μ L of 10 μ M F primer, 1.25 μ L of 10 μ M R primer, 5 μ L of template, and 30.85 μ L of molecular grade
109 water, for a total reaction volume of 50 μ L. Cycling conditions were as follows: 95°C for 2 min, 35 cycles of
110 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec, followed by a final extension of 72°C for 5 min.

111 Each month of samples was amplified on a single plate with the addition of a no template control (NTC;
112 molecular grade water in lieu of template) and a positive control (genomic DNA from kangaroo). After PCR
113 amplification, PCR products were visualized on a 1-2% gel. If no band was present for a given sample, a
114 new amplification was attempted with extracts diluted 1:10 iteratively until a band was detected. PCR
115 products were size-selected and cleaned using MagBind Beads (Omega Biotek, USA) at a sample:beads ratio
116 of 1.2. Bead-cleaned PCR products were eluted in 30 μ L of molecular grade water and quantified via Qubit
117 (Invitrogen, USA).

118 An indexing PCR reaction added a unique index to each sample using Nextera indices (Illumina, USA) to
119 allow pooling multiple samples onto the same sequencing run. For indexing, 10 ng of PCR product was used
120 as template in a final volume of 11.25 μ L. For samples with concentrations less than 0.88 ng/ μ L, 11.25 μ L
121 was added despite being less than 10 ng of amplicon. Each sample received a unique index; Nextera index
122 sets A and B were used to avoid using the same index for more than one sample on a single sequencing run.
123 The PCR reaction included the 11.25 μ L of template, 12.5 μ L of Kapa HiFi MMX (Roche, USA), and 1.25
124 μ L of indexed primer. Cycling conditions were as follows: 95°C for 5 min, 8 cycles of 98°C for 20 sec, 56°C
125 for 30 sec, 72°C for 3 min, and a final extension of 72°C for 5 min.

126 Indexed PCR products were also size-selected and purified using MagBind Beads (Omega Biotek, USA) at a
127 sample:beads ratio of 0.8. Bead-cleaned PCR products were eluted in 30 μ L of molecular grade water and
128 quantified via Qubit. Indexed and bead-cleaned products were normalized before pooling into libraries, which
129 were subsequently quantified via Qubit and visualized on a Bioanalyzer (Agilent, USA) before sequencing.
130 Samples were randomized in 3-month blocks and each block split across 3 sequencing runs, for a total of

131 12 sequencing runs. The loading concentration of each library was 4-8 pM and 5-20% PhiX was included
132 depending on the composition of the run. Sequencing was conducted using an Illumina Miseq with v3 2x300
133 chemistry at the NOAA Northwest Fisheries Science Center and the University of Washington's Northwest
134 Genomics Center.

135 **Species Specific qPCR**

136 We quantified cutthroat trout (*O. clarkii*) DNA in each sample, targeting a 114 bp fragment of the
137 cytochrome b gene with a qPCR assay (Duda et al. 2021). The primer/probe sequences were: F 5'
138 CCGCTACAGTCCTCACCTCTA 3', R 5' GATCTTGTATGAGAAGTAAGGATGGAA 3', P 5' 6FAM-
139 TGAGACAGGATCCAAC-MGB-NFQ 3'. The qPCR assay was multiplexed with TaqMan Exogenous Internal
140 Positive Control Reagents (EXO-IPC) (Applied Biosystems, USA) to check for the presence of PCR inhibitors
141 (Duda et al. 2021). Each DNA sample was run in triplicate using Gene Expression Mastermix (ThermoFisher,
142 USA), a final concentration of 0.375 μ M F primer, 0.375 μ M R primer, and 0.105 μ M probe, as well as 1X
143 EXO-IPC mix, 1X EXO-IPC DNA, 3.5 μ L of template for a final reaction volume of 12 μ L. The EXO-IPC
144 mix includes the primers and probe for the EXI-IPC DNA, with the probe having a VIC reporter, allowing it
145 to be multiplexed with the *O. clarkii* assay, which has a FAM reporter. All qPCRs were conducted on an
146 Applied Biosystems StepOnePlus thermocycler.

147 Thermocycling was as follows: 50°C for 2 min, 95°C for 10 min, followed by 45 cycles of 95°C for 15 sec, 60°C
148 for 1 min. The cycle threshold (Ct) value determined for the EXO-IPC assay from the NTC was compared
149 to the Ct value for the EXO-IPC assay in each of the environmental samples. If the Ct value was >0.5 Ct
150 values from the mean Ct for the NTCs, the sample was deemed inhibited and diluted 1:10 and re-assayed
151 until the Ct value fell within the accepted range. After converting Ct values to DNA concentrations using
152 the standard curve (see below), the concentration was multiplied by the dilution factor.

153 Each plate included a 8-point standard curve created using synthetic DNA (gBlocks) at the following
154 concentrations: 100,000 copies/ μ L, 10,000 copies/ μ L, 1,000 copies/ μ L, 100 copies/ μ L, 10 copies/ μ L, 5
155 copies/ μ L, 3 copies/ μ L, 1 copy/ μ L Additionally, six no template controls (NTCs) were included on each
156 plate: 3 with the IPC DNA mix and 3 with molecular grade water instead of template or IPC DNA mix.
157 Plates were re-run if efficiency as determined by the standard curve was outside of the range of 90-110%.

158 To check for inhibition, the cycle threshold (Ct) value determined for the EXO-IPC assay from the NTC
159 was compared to the Ct value for the EXO-IPC assay in each of the environmental samples. If the Ct value
160 was >0.5 Ct values from the mean Ct for the NTCs, the sample was deemed inhibited and diluted 1:10
161 and re-assayed until the Ct value fell within the accepted range. The majority of environmental samples

162 (60%) were inhibited and accordingly diluted for analysis. In 80% of inhibited samples, a 1:10 dilution or less
163 remedied the inhibition, but some samples (11%) required dilution by a factor of up to 1000.

164 Bioinformatics Processing

165 Primers were removed with cutadapt (Martin 2011) and then reads were de-noised, filtered, merged, and
166 ASVs were generated using dada2 (Callahan et al. 2016). For each MiSeq run, the trimming lengths were
167 determined by visually assessing the quality score plots. After ASVs were generated, taxonomy was assigned
168 using the “classify” function in the insect package in R using the classifier published by the authors of the
169 package (Wilkinson et al. 2018).

170 Quality Controls

171 Positive controls were included on each sequencing run to monitor for cross contamination that might have
172 occurred in the laboratory or due to “tag jumping”. With 13 MiSeq runs, we included one sample of kangaroo
173 tissue on each run and then measured how many reads of kangaroo were found in environmental samples and
174 how many reads of non-kangaroo were found in kangaroo samples (Figure S5).

175 We can also check to make sure that no reads assigning to kangaroo were in the environmental samples. We
176 only found kangaroo in two environmental samples, both of which were a very small number (and proportion)
177 of reads (2 and 136 reads found in samples with 40,425 and 28,725 reads respectively) (Figure S6).

178 Annotation

179 We first used a tree-based annotation method (insect package) and then followed up with a BLAST search for
180 all ASVs that were not annotated to species level by insect. The percent of reads annotated did not correlate
181 with sample read depth, creek, station, or month of sampling. Read depth across samples ranged from 1,011
182 to 311,879, with a mean of 79,709 and median of 75,967 reads (Figure S7). With a total of ~565 samples, 93%
183 of samples had >20,000 reads and 65% of samples had over 50,000 reads. There did not seem to be a pattern
184 with samples of low reads with creek or time. Additionally, of the low read depth samples (<20,000 reads, 40
185 samples), there was only one sample in which all three replicates were low (March 2022 Squalicum Upstream),
186 meaning that it is very unlikely that low read depth samples would have lead to changing ecological results.

187 A total of 81 unique species were identified in the environmental samples by the MiFish primers, including 25
188 fish, 25 mammals, 23 birds , and 8 amphibians (Figure S8 and S9; see also Supplemental Table 1). Of the 81
189 species, 17 only were found in a single environmental sample. The three most commonly found species were
190 coho salmon (*O. kisutch*), cutthroat trout (*O. clarkii*), and rainbow trout (*O. mykiss*).

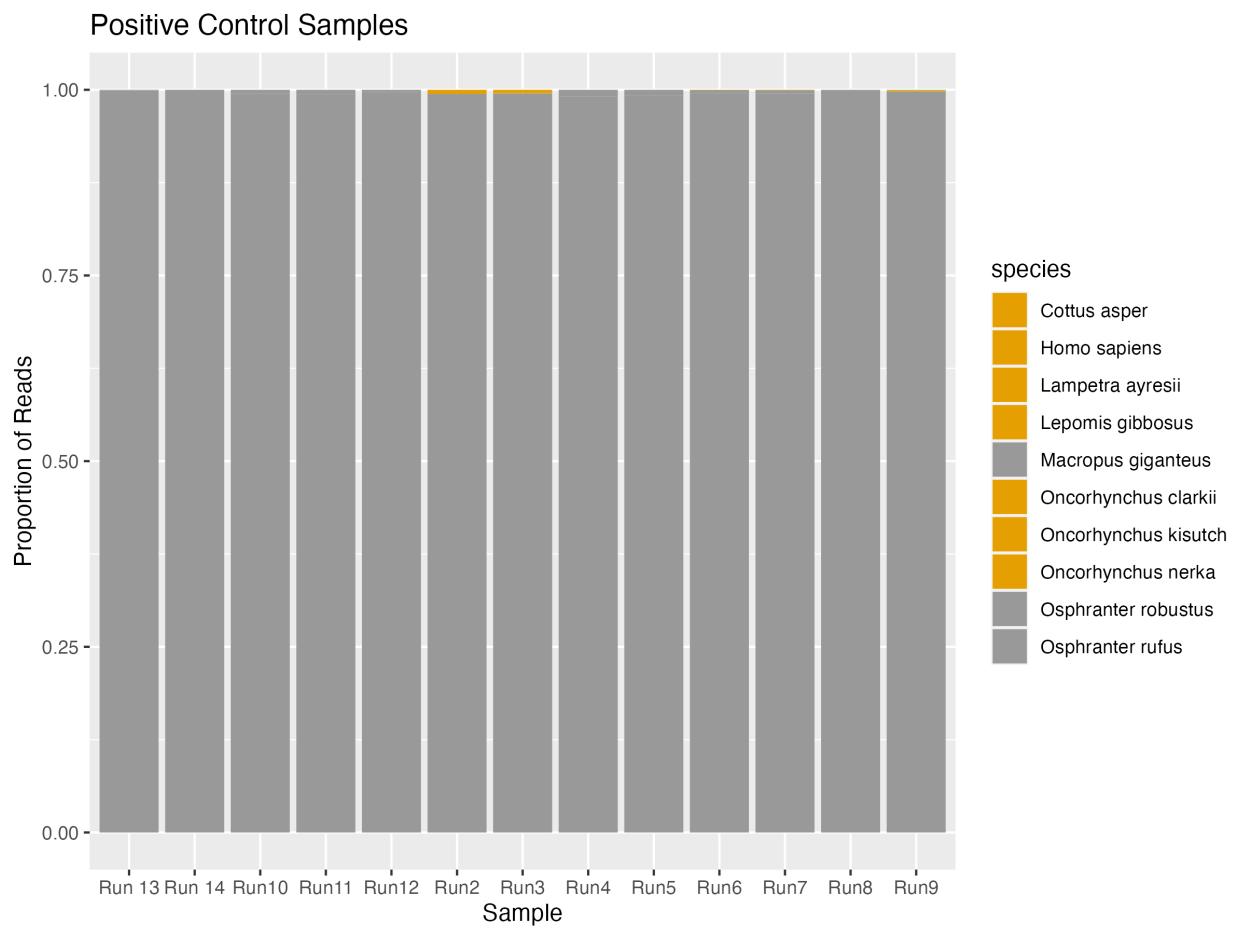


Figure S5. Proportion of annotated reads found in positive controls. Grey colors are the three species of kangaroo used for positive controls and are what should be in each sample. Orange species should not be in the positive controls and indicate low level contamination from environmental samples.

Environmental Samples with Positive Control

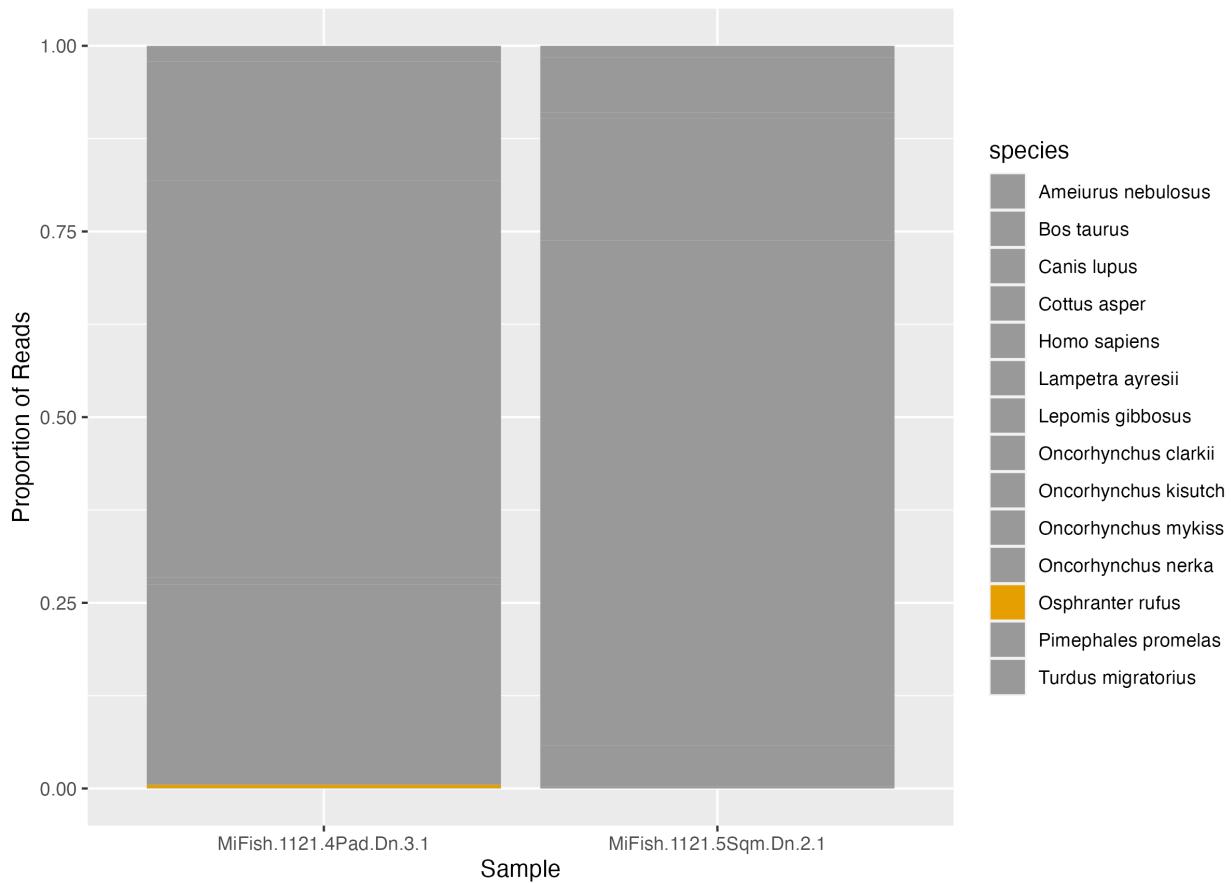


Figure S6. Proportion of annotated reads found in environmental samples with positive control. Grey colors are non-kangaroo reads and therefore what should be in each sample. Orange species are kangaroo reads and therefore should not be in the environmental samples and indicate low level contamination from positive controls.

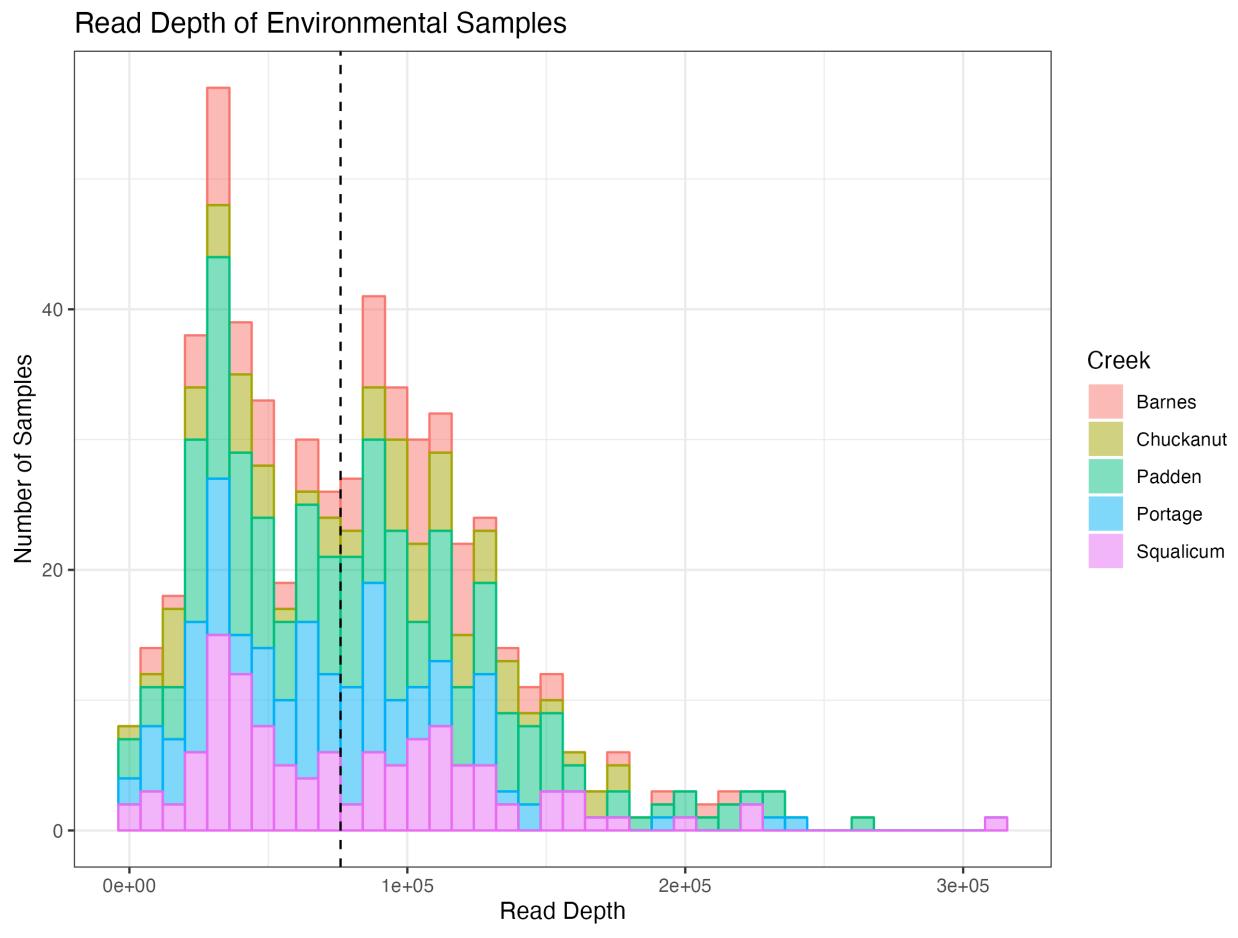


Figure S7. Read depth of samples colored by creek. Dashed line shows median read depth (87,698 reads).

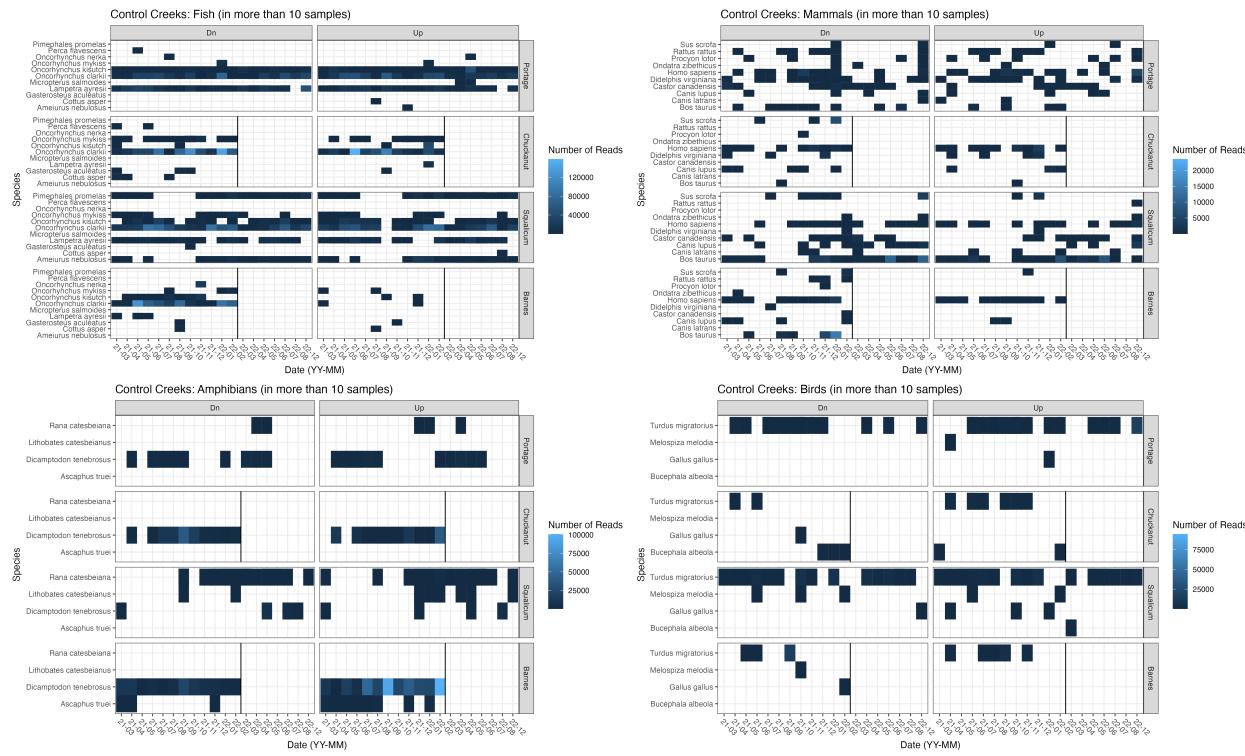


Figure S8. Heat map of all species found in control creeks in at least ten environmental samples.

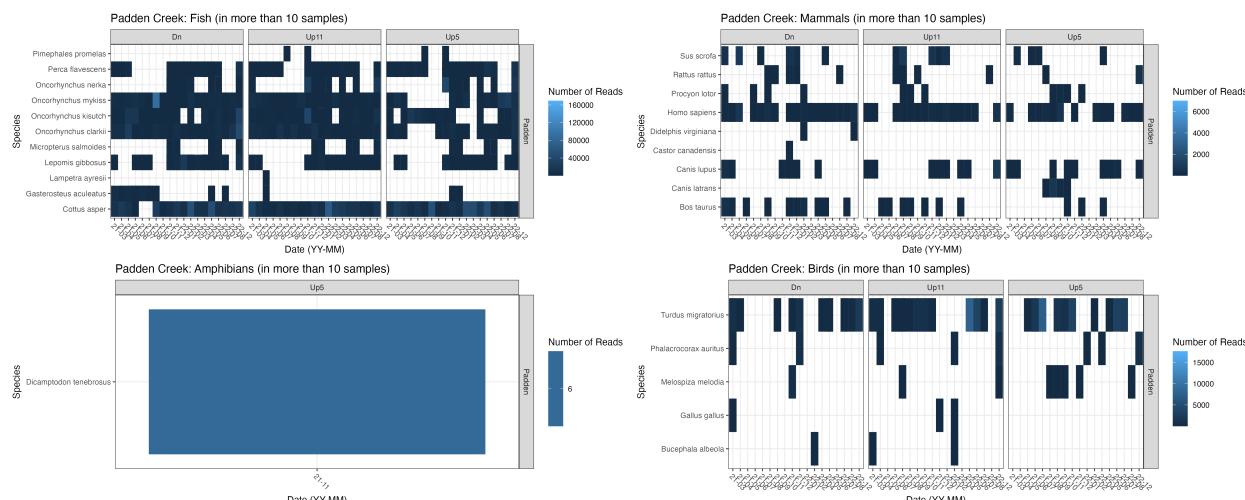


Figure S9. Heat map of all species found in Padden Creek in at least ten environmental samples.

191 **Correcting metabarcoding data for amplification bias**

192 Using our six mock communities (three different taxa compositions at two different proportions [even and
193 skewed]), we can first check how well the quantitative metabarcoding model corrects for amplification bias.
194 In one case, we consider the even mock communities as the mock community data and the skewed mock
195 communities as unknown. We can then re-create what the model believes to be the original starting proportions
196 of the skewed mock community given the proportions of reads found in the skewed mock communities and
197 the proportion of DNA as compared to the proportion of reads found in the even mock communities. We can
198 also do the same treating the skewed mock communities as known and even mock communities as unknown
199 (Figure S10).

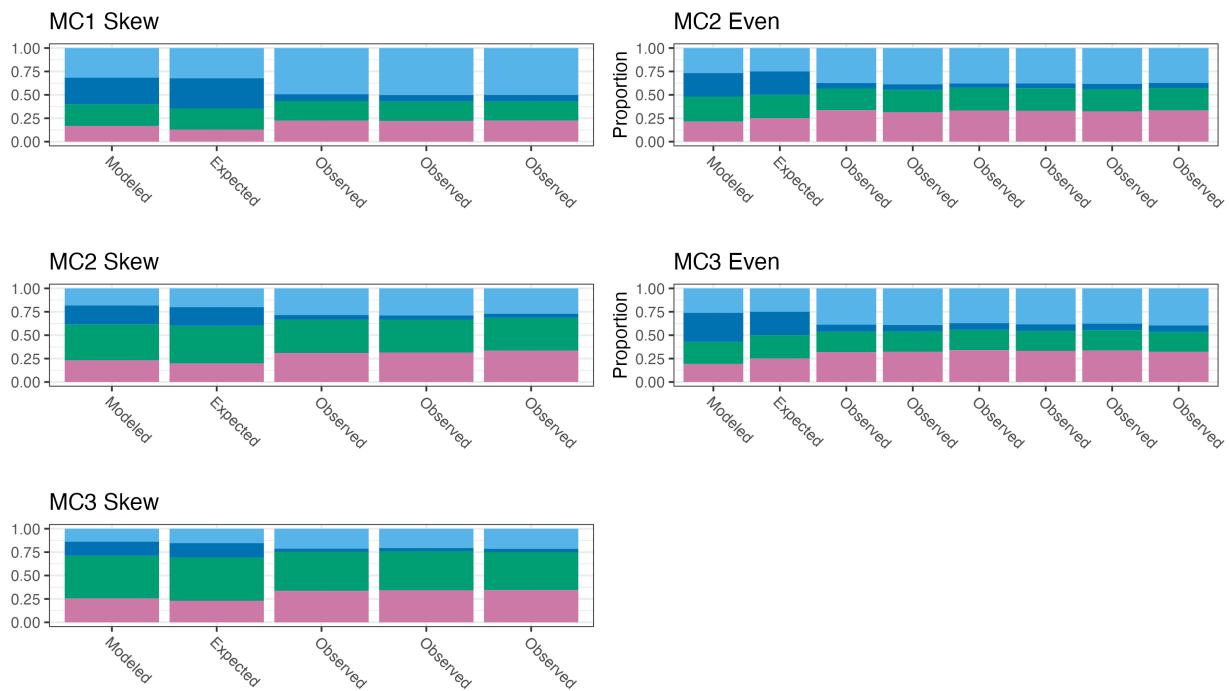


Figure S10. Intercalibration of mock communities used to correct environmental samples for amplification bias.

200 We can also check how well the calibration is working by comparing the alpha values by using different
201 subsets of mock community data as true and unknown (Figure S11). We can then use the mock communities
202 to correct the data from the MiSeq to account for the different alpha values. The corrected results are shown
203 in the main text as Figure 4.

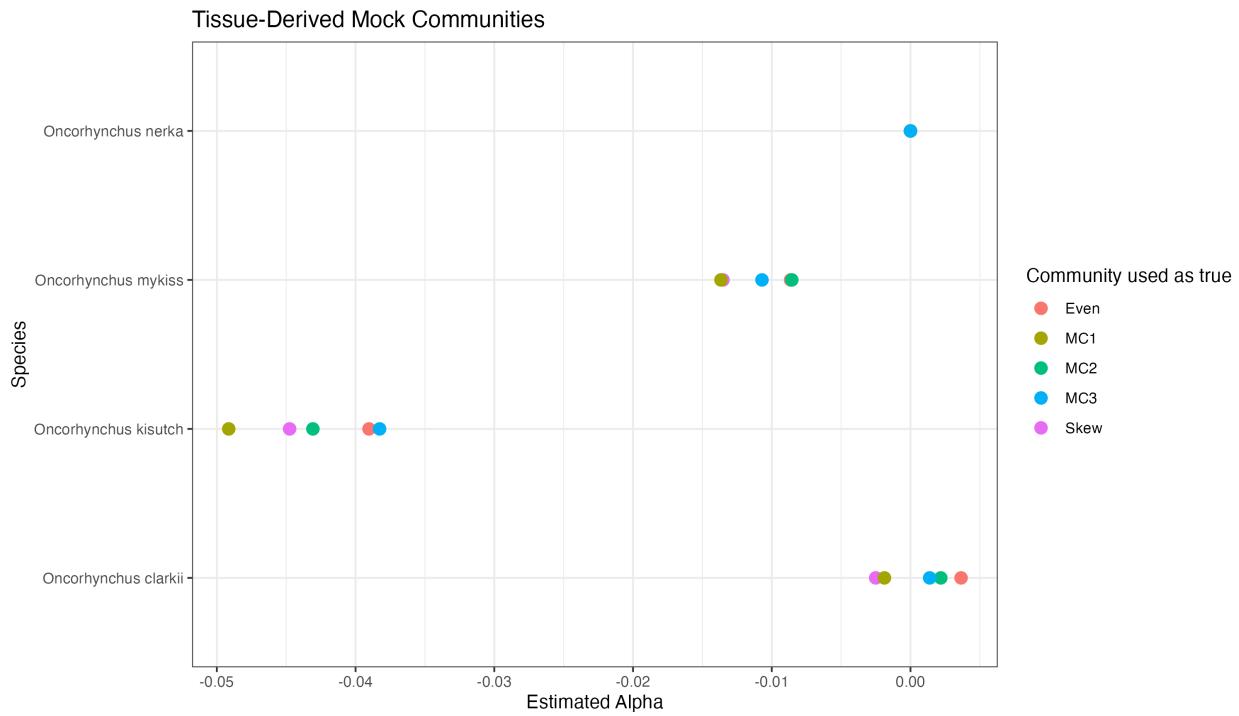


Figure S11. Estimated alpha values of salmonid species with different calibrations of the mock communities. Each color represents a different subset of mock community data treated as ‘true’ to calibrate the remainder of the mock community data.

204 Species-specific Effect of Culverts

205 In the main text, we show the effect of culverts averaged over creeks and species (Figure 8). Here, we show
206 them separated by species and creek (Figure S12).

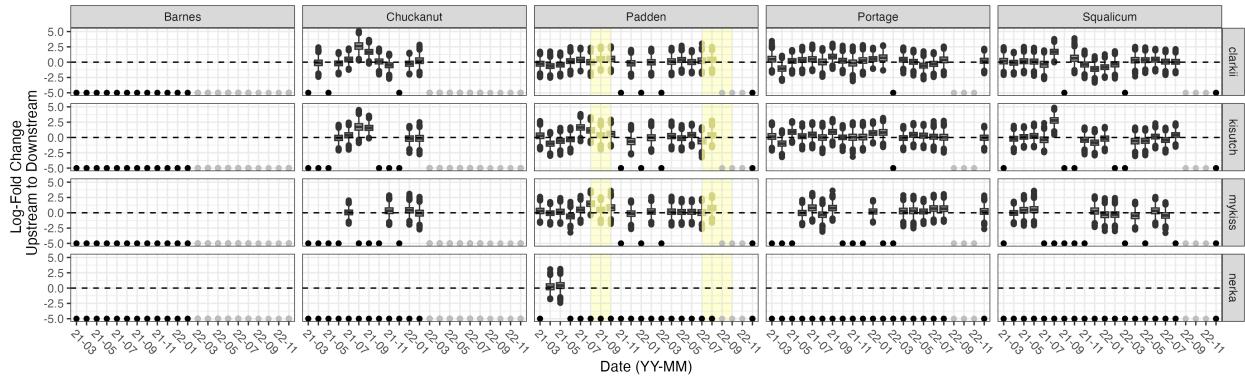


Figure S12. The effect of culvert on salmonid abundance separated by species and creeks across time. The y-axis shows the log-fold change in eDNA mass flow rate (copies/s) between upstream and downstream, normalized by upstream mass flow rate. The box boundaries correspond to the 25th and 75th percentiles; the whiskers correspond to 1.5 times the interquartile range. Here, negative values imply that eDNA mass flow rates are higher downstream than upstream. Samples with very low eDNA mass flow rates (< 150 copies/s) were removed before plotting to remove extreme proportional values due to large denominators. Grey points indicate times when no samples were taken. Black points indicate times when samples were taken, but no target DNA was found in either upstream or downstream samples and therefore the log-fold change can not be calculated.

207 References

- 208 Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2:
209 High resolution sample inference from illumina amplicon data. *Nature methods* 13:581–583.
- 210 Duda, J. J., M. S. Hoy, D. M. Chase, G. R. Pess, S. J. Brenkman, M. M. McHenry, and C. O. Ostberg. 2021.
211 Environmental DNA is an effective tool to track recolonizing migratory fish following large-scale dam
212 removal. *Environmental DNA* 3:121–141.
- 213 Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB-*
214 *net.journal* 17:10.
- 215 Shelton, A. O., Z. J. Gold, A. J. Jensen, E. D’Agnese, E. Andruszkiewicz Allan, A. Van Cise, R. Gallego, A.
216 Ramón-Laca, M. Garber-Yonts, K. Parsons, and R. P. Kelly. 2022. Toward quantitative metabarcoding.
217 *Ecology* n/a:e3906.
- 218 Thomas, A. C., J. Howard, P. L. Nguyen, T. A. Seimon, and C. S. Goldberg. 2018. ANDe™: A fully
219 integrated environmental DNA sampling system. *Methods in Ecology and Evolution* 9:13791385.
- 220 Thomas, A. C., P. L. Nguyen, J. Howard, and C. S. Goldberg. 2019. A self-preserving, partially biodegradable
221 eDNA filter. *Methods in Ecology and Evolution* 10:1136–1141.
- 222 Washington Department of Fish and Wildlife. 2019. Fish passage inventory, assessment, and prioritization
223 manual.
- 224 Wilkinson, S. P., S. K. Davy, M. Bunce, and M. Stat. 2018. Taxonomic identification of environmental DNA

²²⁵ with informatic sequence classification trees.