

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

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

2023

6 Field Sampling

7 Site selection and study design

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

14 Distance between sites and flow variability at sites

¹⁵ The average distance between upstream and downstream sampling within a creek was about 160 m; the
¹⁶ largest distance between downstream and upstream sampling was at Barnes Creek, which was approximately
¹⁷ 330 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 S1.1). 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).

²² The flow meters at Squalicum Creek and Chuckanut Creek were offline from November 2021 for the remainder
²³ of the sampling period. The highest discharge seen during the course of the study from January to November

²⁴ 2021 occurred in November 2021 at Squalicum Creek. The mean discharge in each creek was: 0.42 m³/s in
²⁵ Padden, 0.29 m³/s in Chuckanut, and 1.14 m³/s in Squalicum Creek. The lowest discharge registered by the
²⁶ flow meters is 0.0028 m³/s, which occurred 8.5%, 1.6%, and 0.78% of the time in Padden, Chuckanut, and
²⁷ Squalicum, respectively.

²⁸ Due to the lack of flow data in Squalicum and Chuckanut Creeks from November 2021 to February 2022, we
²⁹ used historical data from the three flow gauges to calculate the average discharge for each day of the year
³⁰ from about 2015-2021 (Figure S1.2). We then used the value for the day of the year that we sampled in either
³¹ 2021 or 2022 when the gauges were offline. For consistency, we also did this at Padden Creek despite the
³² gauge there being online for the entire sampling period.

³³ We compared the different ways one could use flow data to correct the eDNA concentrations. We included
³⁴ (1) the value from the closest time point from the gauge to the time point sampling, (2) the average flow
³⁵ on the day of sampling from the gauge, (3) the monthly average for the month of sampling from the gauge,
³⁶ and (4) the correction factor approach. For (4), the values for Padden Creek represent the same as (1) for
³⁷ Padden Creek, the value from the closest time point in the gauge to the time point of sampling. The values
³⁸ for Chuckanut and Squalicum Creek are based on the the correction factor from Padden Creek. First, five
³⁹ years of historical data (2015-2020) were used to find monthly averages for flow rates for each creek. Because
⁴⁰ the gauges in Squalicum and Chuckanut Creeks stopped metering in 2021, we solved for the ratio of the
⁴¹ monthly average of each of those creeks to Padden Creek. Then, we used the closest values from Padden
⁴² Creek (1) and multiplied by the monthly correction factor from the 5 years of historical data to find a value
⁴³ for Squalicum and Chuckanut Creeks to use for the year of sampling (2021-2022). For all three creeks, we
⁴⁴ demonstrate the relatively small changes in discharge depending on which way flow data were used (Figure
⁴⁵ S1.3). Though in the course of sampling, the discharge in Padden Creek ranged from no metered flow to
⁴⁶ 23 m³/s, the discharge on the dates of sampling only reached a maximum of 1.3 m³/s. For sites with no
⁴⁷ metered flow, half of the minimum verified discharge of the flow gauge was used (0.0014 m³/s).

⁴⁸ **Construction and Fish Exclusion**

⁴⁹ Fish were excluded from Padden Creek on August 30th, 2021 in preparation for the stream to be diverted
⁵⁰ on September 9th, 2021 and the diversion was removed October 7th, 2021. Water sampling occurred on
⁵¹ September 10th, 2021, the day after the diversion, and on October 12, 2021, just 5 days after reconnecting
⁵² the stream (Figure S1.4).

53 **Stocking in Lake Padden**

54 Padden Lake has historically been stocked with hatchery fish by the Washington Department of Fish and
55 Wildlife (Figure S1.4). Rainbow trout (*O. mykiss*) and occasionally cutthroat trout (*O. clarkii*) and kokanee
56 salmon (*O. nerka*) are stocked in Lake Padden, approximately 1.5 km upstream of the sampling sites. During
57 the course of the study, rainbow trout were stocked in April 2021 and April 2022, kokanee salmon were
58 stocked in May 2021 and October 2022, and cutthroat trout were stocked in November 2022 (Figure S1.4).
59 However, despite the stocking of 30,000 kokanee salmon in May in Lake Padden, *O. nerka* was not detected
60 by metabarcoding in May 2021 or at any point in 2021 until November (see main text Results). Importantly,
61 this suggests that Lake Padden is far enough upstream that the eDNA signal at the sampling sites by the
62 culvert is not a result of stocking the lake 1.5 km upstream (see main text Discussion for more information).

63 **Water Sampling**

64 Water samples were collected using Smith Root's eDNA Backpack (Thomas et al. 2018), a portable pumping-
65 and-filtering setup set to filter at 1 L/min at 82.7 kPa (12 psi). For most months, a trident sampler was used
66 to collect all 3 biological replicates at the exact same time, for a total sampling time of about 5 minutes.
67 Otherwise, the three replicates were collected consecutively, for a total sampling time of about 15 minutes.
68 Downstream sites were always sampled before upstream sites to ensure no potential DNA was introduced
69 into the stream before sampling. In some samples, less than 2 L of water was filtered due to clogging (mean
70 = 1.95 L).

71 **Laboratory Processing**

72 **DNA Extraction, Amplification, Sequencing**

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

83 eDNA data from water samples with mock community data as described in Shelton et al. (2022) (see
84 Supplemental Text 2 for more information).

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

89 **DNA Extractions**

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

98 **PCR Amplification**

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

109 Each month of samples was amplified on a single plate with the addition of a no template control (NTC;
110 molecular grade water in lieu of template) and a positive control (genomic DNA from kangaroo). After PCR
111 amplification, PCR products were visualized on a 1-2% gel. If no band was present for a given sample, a
112 new amplification was attempted with extracts diluted 1:10 iteratively until a band was detected. PCR

113 products were size-selected and cleaned using MagBind Beads (Omega Biotek, USA) at a sample:beads ratio
114 of 1.2. Bead-cleaned PCR products were eluted in 30 μ L of molecular grade water and quantified via Qubit
115 (Invitrogen, USA).

116 An indexing PCR reaction added a unique index to each sample using Nextera indices (Illumina, USA) to
117 allow pooling multiple samples onto the same sequencing run. For indexing, 10 ng of PCR product was used
118 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
119 was added despite being less than 10 ng of amplicon. Each sample received a unique index; Nextera index
120 sets A and B were used to avoid using the same index for more than one sample on a single sequencing run.
121 The PCR reaction included the 11.25 μ L of template, 12.5 μ L of Kapa HiFi MMX (Roche, USA), and 1.25
122 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
123 for 30 sec, 72°C for 3 min, and a final extension of 72°C for 5 min.

124 Indexed PCR products were also size-selected and purified using MagBind Beads (Omega Biotek, USA) at a
125 sample:beads ratio of 0.8. Bead-cleaned PCR products were eluted in 30 μ L of molecular grade water and
126 quantified via Qubit. Indexed and bead-cleaned products were normalized before pooling into libraries, which
127 were subsequently quantified via Qubit and visualized on a Bioanalyzer (Agilent, USA) before sequencing.
128 Samples were randomized in 3-month blocks and each block split across 3 sequencing runs, for a total of
129 12 sequencing runs. The loading concentration of each library was 4-8 pM and 5-20% PhiX was included
130 depending on the composition of the run. Sequencing was conducted using an Illumina MiSeq with v3 2x300
131 chemistry at the NOAA Northwest Fisheries Science Center and the University of Washington's Northwest
132 Genomics Center.

133 Species Specific qPCR

134 We quantified cutthroat trout (*O. clarkii*) DNA in each sample, targeting a 114 bp fragment of the
135 cytochrome b gene with a qPCR assay (Duda et al. 2021). The primer/probe sequences were: F 5'
136 CCGCTACAGTCCTCACCTTCTA 3', R 5' GATCTTGATGAGAAGTAAGGATGGAA 3', P 5' 6FAM-
137 TGAGACAGGATCCAAC-MGB-NFQ 3'. The qPCR assay was multiplexed with TaqMan Exogenous Internal
138 Positive Control Reagents (EXO-IPC) (Applied Biosystems, USA) to check for the presence of PCR inhibitors
139 (Duda et al. 2021). Each DNA sample was run in triplicate using Gene Expression Mastermix (ThermoFisher,
140 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
141 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
142 mix includes the primers and probe for the EXI-IPC DNA, with the probe having a VIC reporter, allowing it
143 to be multiplexed with the *O. clarkii* assay, which has a FAM reporter. All qPCRs were conducted on an

¹⁴⁴ Applied Biosystems StepOnePlus thermocycler.

¹⁴⁵ 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
¹⁴⁶ for 1 min. The cycle threshold (Ct) value determined for the EXO-IPC assay from the NTC was compared
¹⁴⁷ to the Ct value for the EXO-IPC assay in each of the environmental samples. If the Ct value was >0.5 Ct
¹⁴⁸ values from the mean Ct for the NTCs, the sample was deemed inhibited and diluted 1:10 and re-assayed
¹⁴⁹ until the Ct value fell within the accepted range. After converting Ct values to DNA concentrations using
¹⁵⁰ the standard curve (see below), the concentration was multiplied by the dilution factor.

¹⁵¹ Each plate included a 8-point standard curve created using synthetic DNA (gBlocks) at the following
¹⁵² concentrations: 100,000 copies/µL, 10,000 copies/µL, 1,000 copies/µL, 100 copies/µL, 10 copies/µL, 5
¹⁵³ copies/µL, 3 copies/µL, 1 copy/µL Additionally, six no template controls (NTCs) were included on each
¹⁵⁴ plate: 3 with the IPC DNA mix and 3 with molecular grade water instead of template or IPC DNA mix.
¹⁵⁵ Plates were re-run if efficiency as determined by the standard curve was outside of the range of 90-110%.

¹⁵⁶ To check for inhibition, the cycle threshold (Ct) value determined for the EXO-IPC assay from the NTC
¹⁵⁷ was compared to the Ct value for the EXO-IPC assay in each of the environmental samples. If the Ct value
¹⁵⁸ was >0.5 Ct values from the mean Ct for the NTCs, the sample was deemed inhibited and diluted 1:10
¹⁵⁹ and re-assayed until the Ct value fell within the accepted range. The majority of environmental samples
¹⁶⁰ (60%) were inhibited and accordingly diluted for analysis. In 80% of inhibited samples, a 1:10 dilution or less
¹⁶¹ remedied the inhibition, but some samples (11%) required dilution by a factor of up to 1000.

¹⁶² Bioinformatics Processing

¹⁶³ Primers were removed with cutadapt (Martin 2011) and then reads were de-noised, filtered, merged, and
¹⁶⁴ ASVs were generated using dada2 (Callahan et al. 2016). For each MiSeq run, the trimming lengths were
¹⁶⁵ determined by visually assessing the quality score plots. After ASVs were generated, taxonomy was assigned
¹⁶⁶ using the “classify” function in the insect package in R using the classifier published by the authors of the
¹⁶⁷ package (Wilkinson et al. 2018).

¹⁶⁸ Quality Controls

¹⁶⁹ Positive controls were included on each sequencing run to monitor for cross contamination that might have
¹⁷⁰ occurred in the laboratory or due to “tag jumping”. With 13 MiSeq runs, we included one sample of kangaroo
¹⁷¹ tissue on each run and then measured how many reads of kangaroo were found in environmental samples and
¹⁷² how many reads of non-kangaroo were found in kangaroo samples (Figure S1.5).

173 We can also check to make sure that no reads assigning to kangaroo were in the environmental samples. We
174 only found kangaroo in two environmental samples, both of which were a very small number (and proportion)
175 of reads (2 and 136 reads found in samples with 40,425 and 28,725 reads respectively) (Figure S1.6).

176 **Annotation**

177 We first used a tree-based annotation method (insect package) and then followed up with a BLAST search
178 for all ASVs that were not annotated to species level by insect. The percent of reads annotated did not
179 correlate with sample read depth, creek, station, or month of sampling. Read depth across samples ranged
180 from 1,011 to 311,879, with a mean of 79,709 and median of 75,967 reads (Figure S1.7). With a total of
181 ~565 samples, 93% of samples had >20,000 reads and 65% of samples had over 50,000 reads. There did
182 not seem to be a pattern with samples of low reads with creek or time. Additionally, of the low read depth
183 samples (<20,000 reads, 40 samples), there was only one sample in which all three replicates were low (March
184 2022 Squalicum Upstream), meaning that it is very unlikely that low read depth samples would have lead to
185 changing ecological results.

186 A total of 81 unique species were identified in the environmental samples by the MiFish primers, including 25
187 fish, 25 mammals, 23 birds , and 8 amphibians (Figure S1.8 and S1.9; see also Supplemental Table 1). Of the
188 81 species, 17 only were found in a single environmental sample. The three most commonly found species
189 were coho salmon (*O. kisutch*), cutthroat trout (*O. clarkii*), and rainbow trout (*O. mykiss*).

190 **Correcting metabarcoding data for amplification bias**

191 Using our six mock communities (three different taxa compositions at two different proportions [even and
192 skewed]), we can first check how well the quantitative metabarcoding model corrects for amplification bias.
193 In one case, we consider the even mock communities as the mock community data and the skewed mock
194 communities as unknown. We can then re-create what the model believes to be the original starting proportions
195 of the skewed mock community given the proportions of reads found in the skewed mock communities and
196 the proportion of DNA as compared to the proportion of reads found in the even mock communities. We can
197 also do the same treating the skewed mock communities as known and even mock communities as unknown
198 (Figure S1.10).

199 We can also check how well the calibration is working by comparing the alpha values by using different subsets
200 of mock community data as true and unknown (Figure S1.11). We can then use the mock communities to
201 correct the data from the MiSeq to account for the different alpha values. The corrected results are shown in
202 the main text as Figure 4.

203 **Species-specific Effect of Culverts**

204 In the main text, we show the effect of culverts averaged over creeks and species (Figure 8). Here, we show
205 them separated by species and creek (Figure S1.12).

206 **References**

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

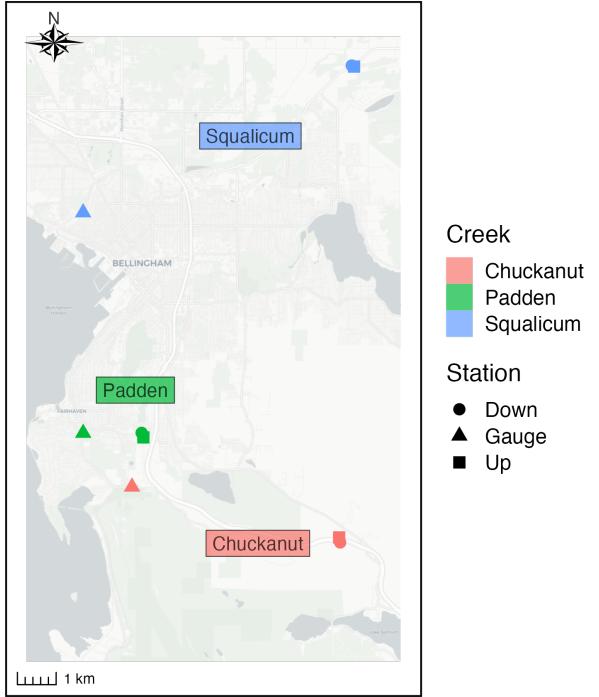


Figure S1.1. Location of flow gauges compared to sampling locations for Chuckanut, Padden, and Squalicum Creeks.

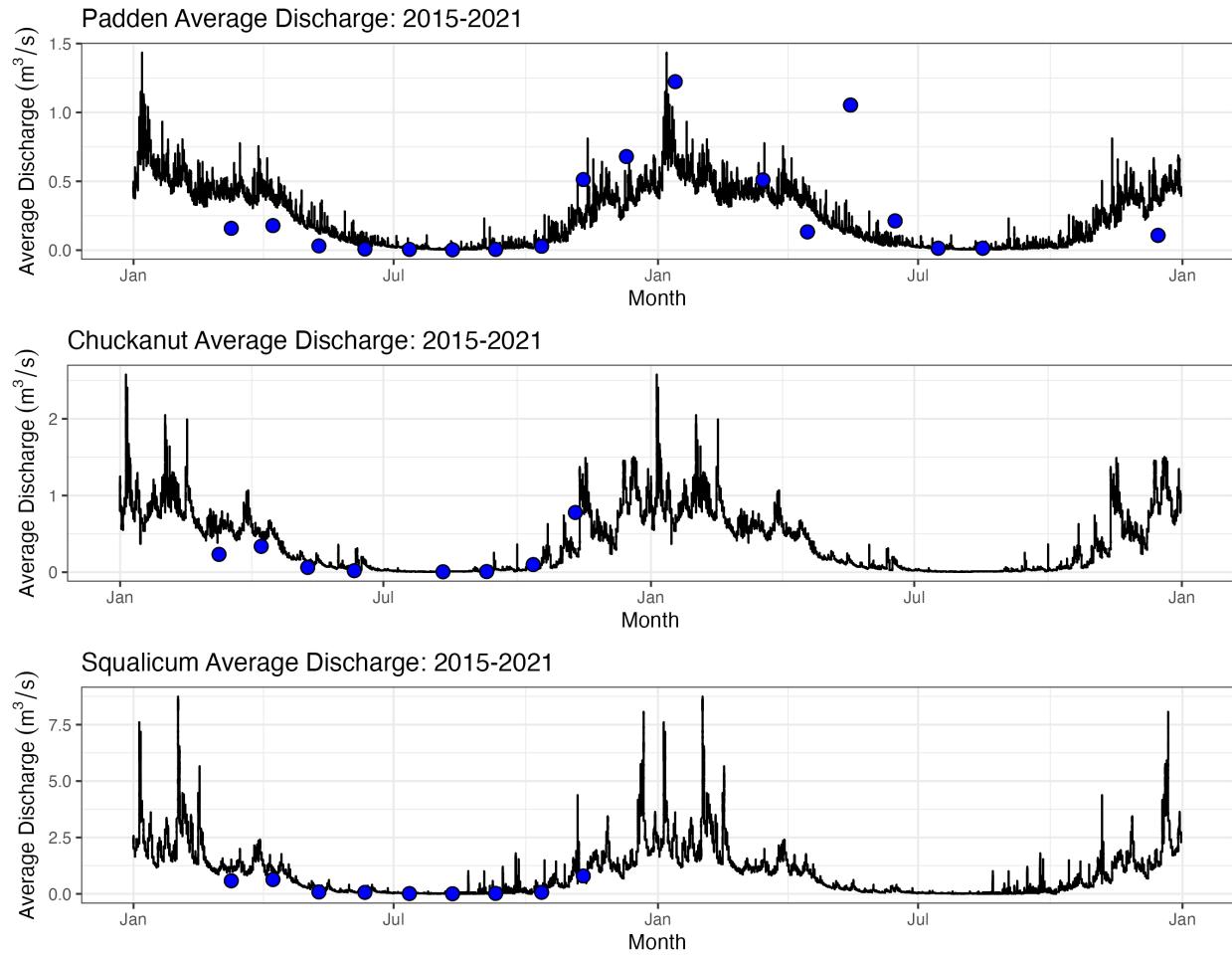


Figure S1.2. Daily average discharge from 2015-2020 in creeks from USGS gauges. Blue dots show the discharge time of sampling during the course of this study (2021-2022) for time points where gauges were online (note the missing data points after December through February in Chuckanut and Squalicum Creeks).

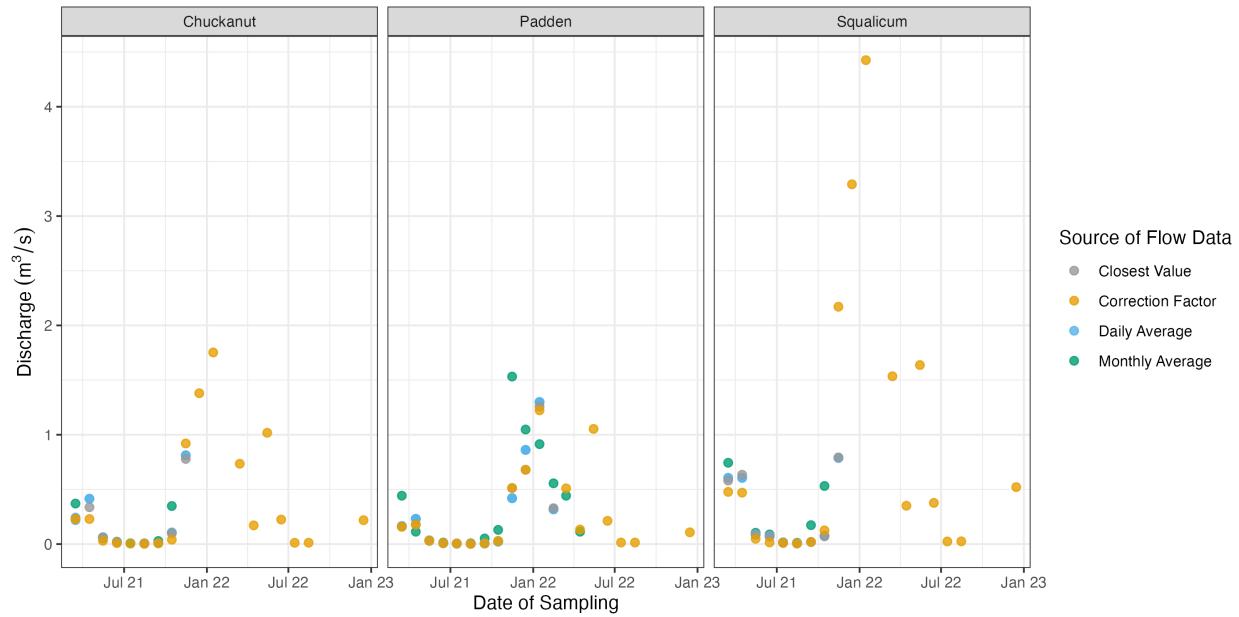


Figure S1.3. Comparison of different ways flow data can be used to correct eDNA concentrations. In the main text, the correction factor is used. Note that for Padden Creek, the “Correction Factor” method is the same as the “Closest Value” method. Also note that for Chuckanut and Squalicum Creeks, no data exist for closest value, daily average, or monthly average after November 2021 when the gauges went offline.

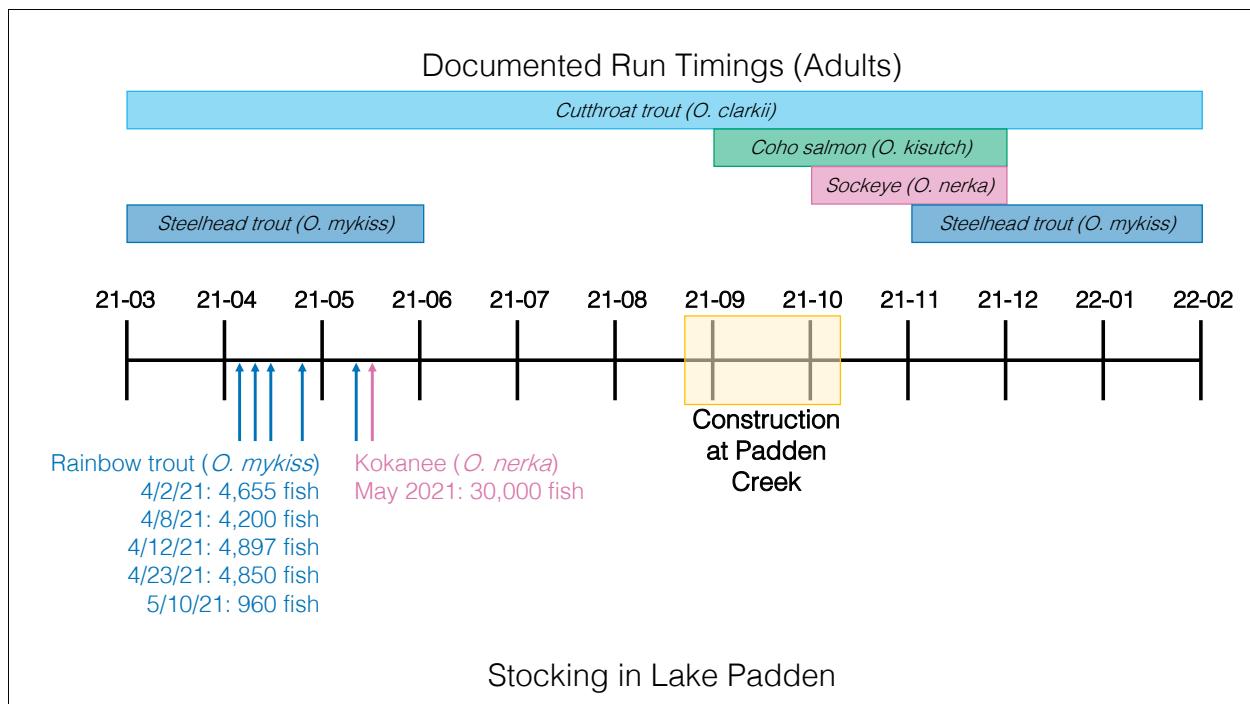


Figure S1.4. Timeline of runs for migrating species, stocking of Lake Padden, and construction at the intervention site (Padden Creek). Note, the dates of kokanee stocking do not include days, just the month and year. For plotting purposes, they are shown as the 15th.

Positive Control Samples

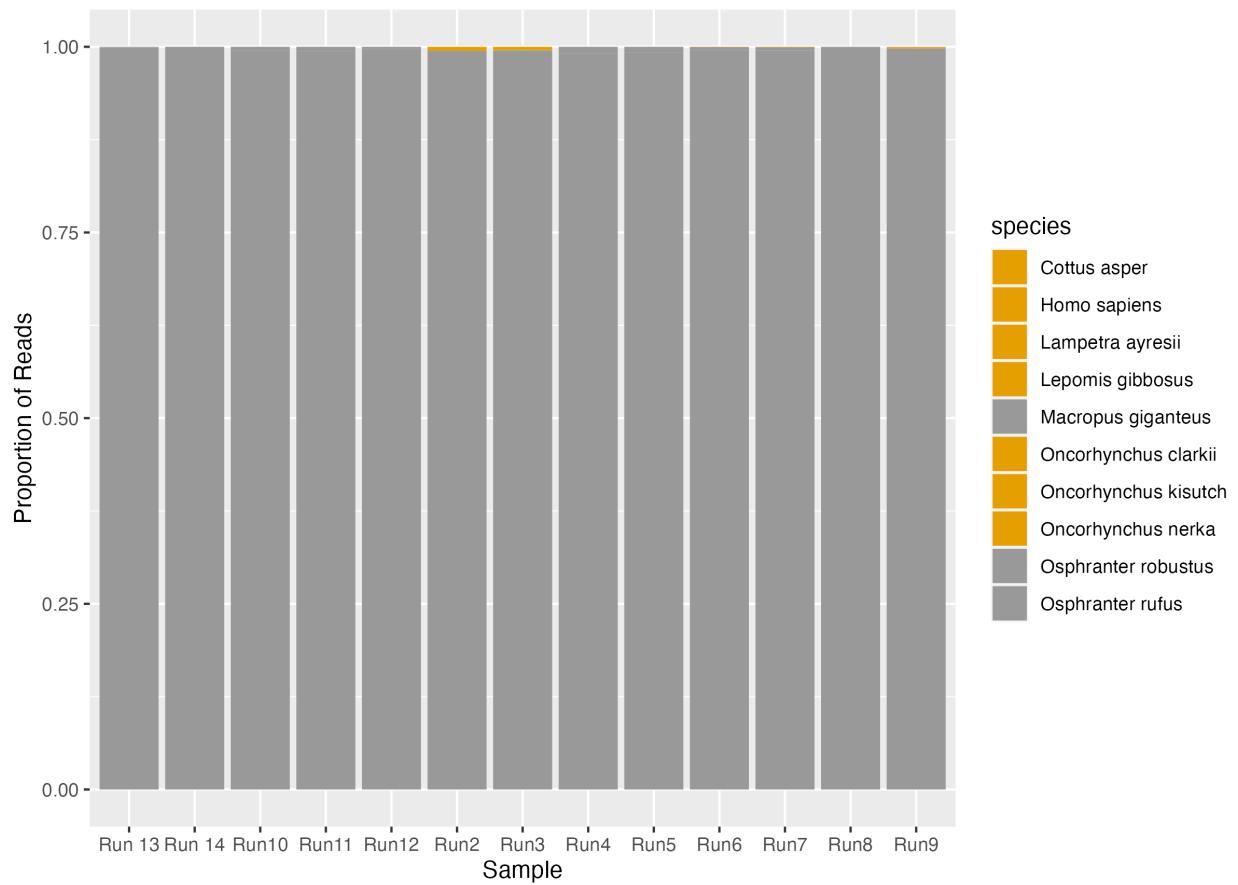


Figure S1.5. Proportion of annotated reads found in positive controls. Grey colors are the three species of kangaroo used for positive controls and are what should be in each sample. Orange species should not be in the positive controls and indicate low level contamination from environmental samples.

Environmental Samples with Positive Control

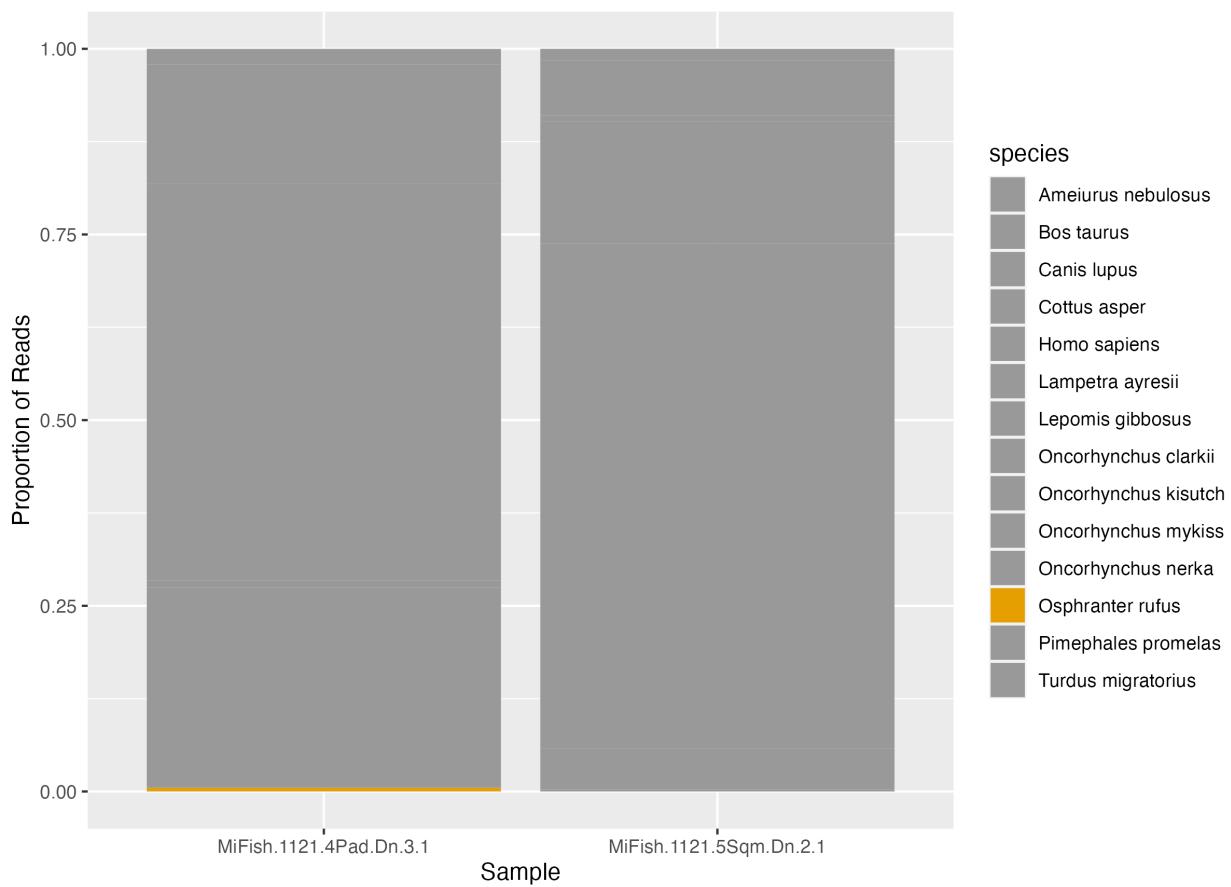


Figure S1.6. Proportion of annotated reads found in environmental samples with positive control. Grey colors are non-kangaroo reads and therefore are what should be in each sample. Orange species are kangaroo reads and therefore should not be in the environmental samples and indicate low level contamination from positive controls.

Read Depth of Environmental Samples

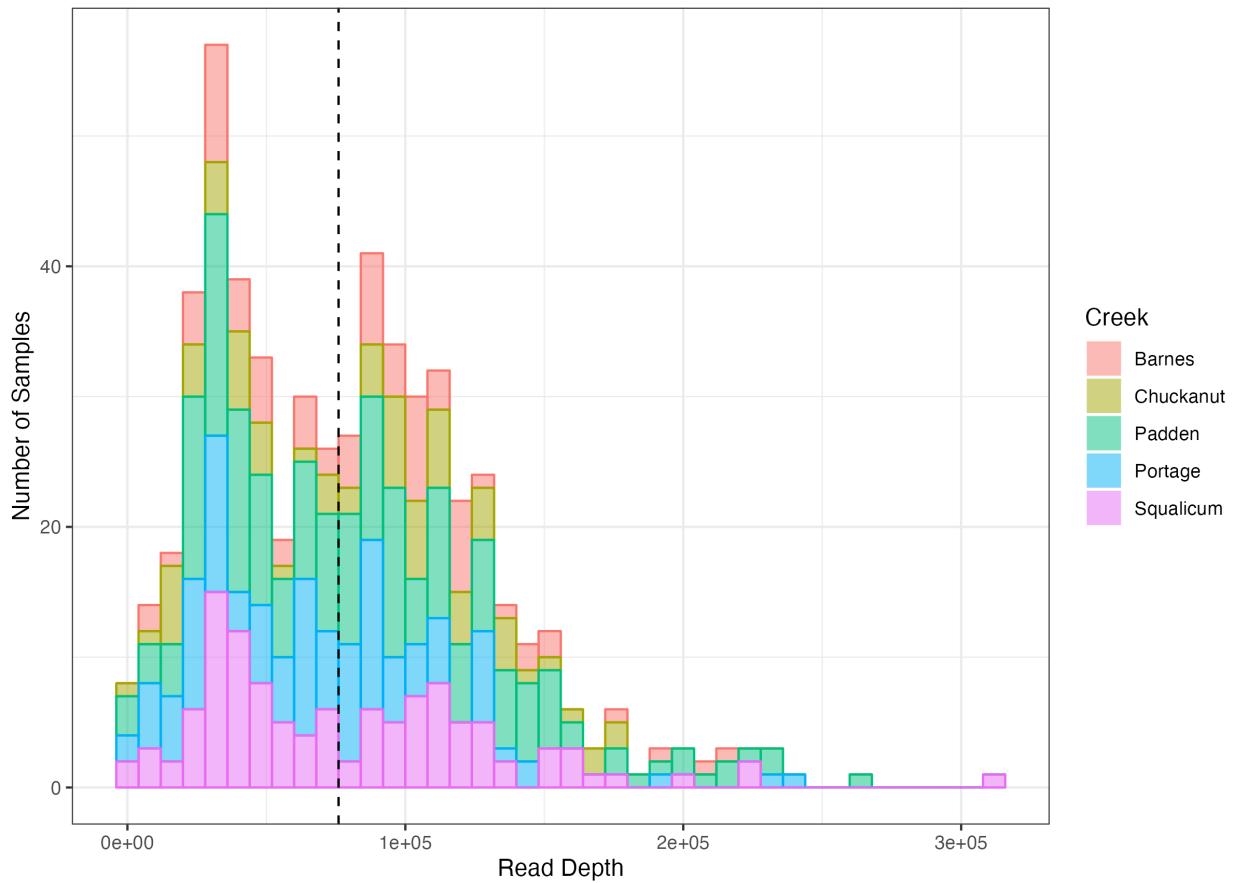


Figure S1.7. Read depth of samples colored by creek. Dashed line shows median read depth (87,698 reads).

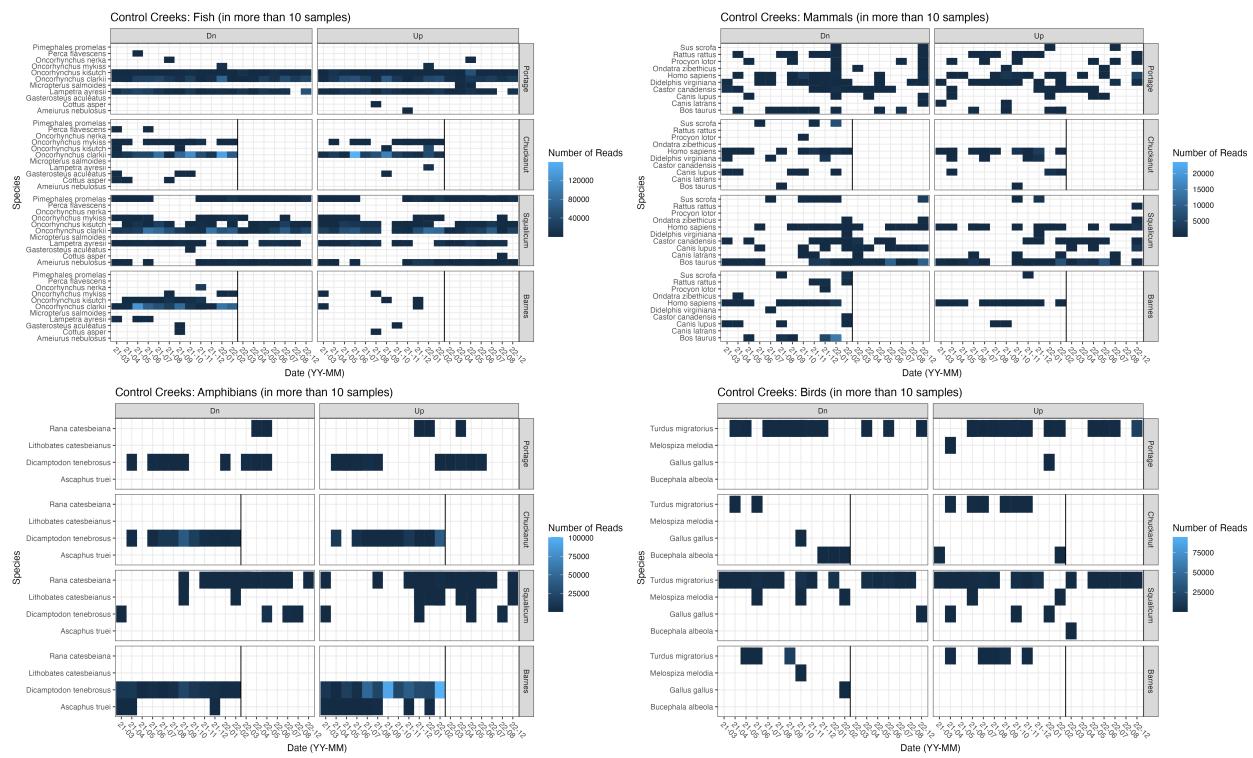


Figure S1.8. Heat map of all species found in control creeks in at least ten environmental samples.

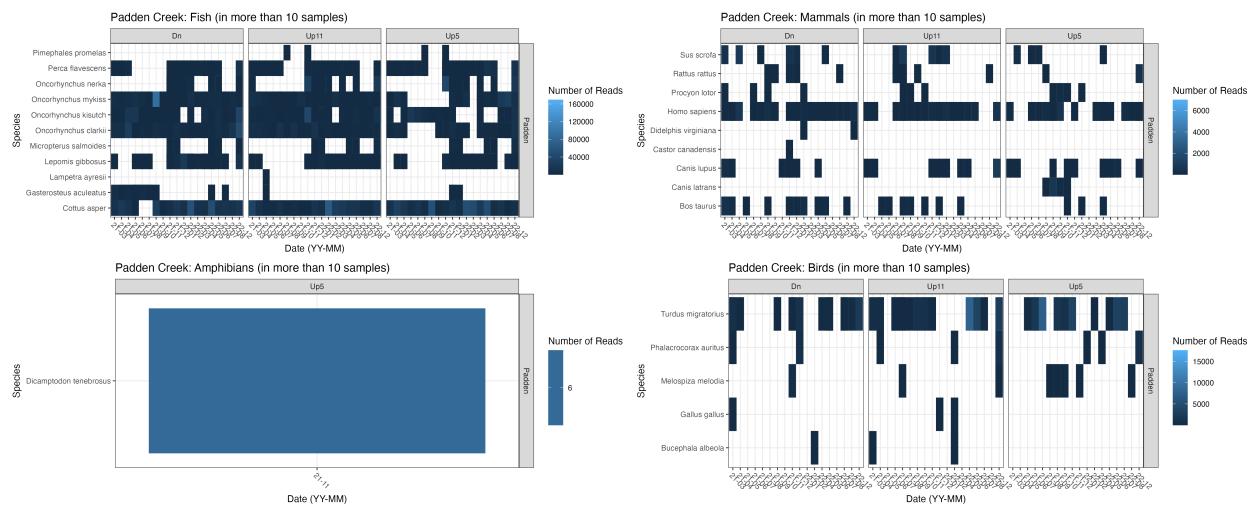


Figure S1.9. Heat map of all species found in Padden Creek in at least ten environmental samples.

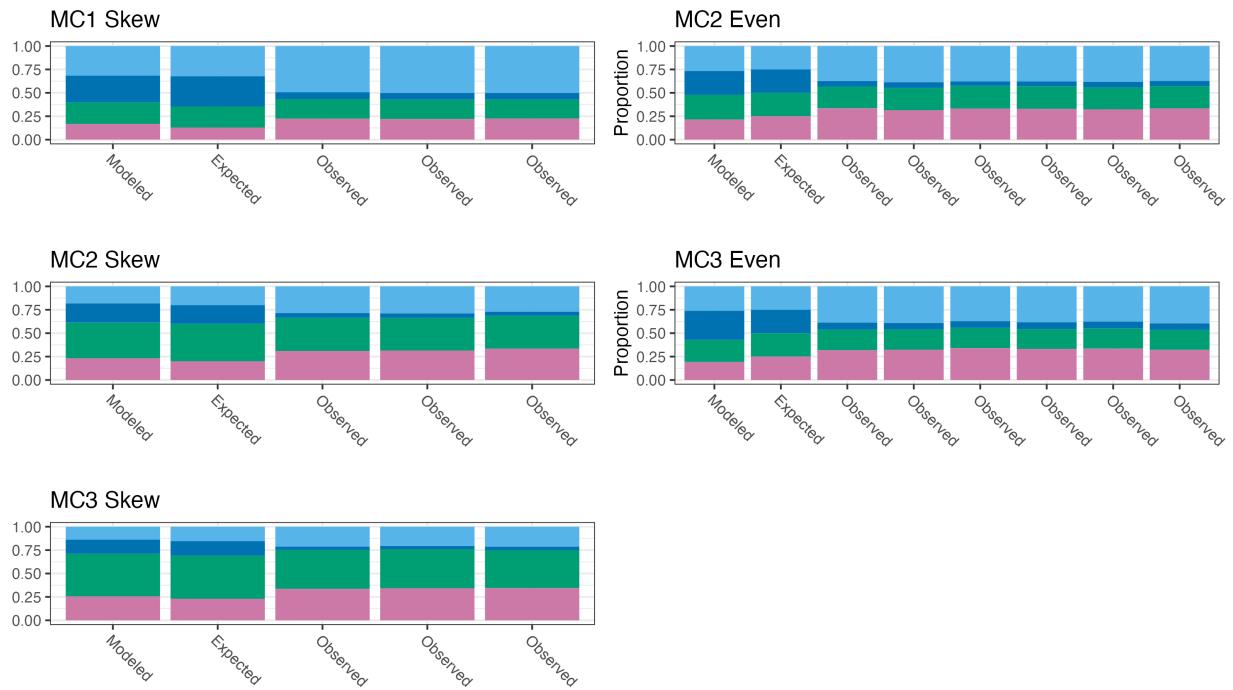


Figure S1.10. Intercalibration of mock communities used to correct environmental samples for amplification bias.

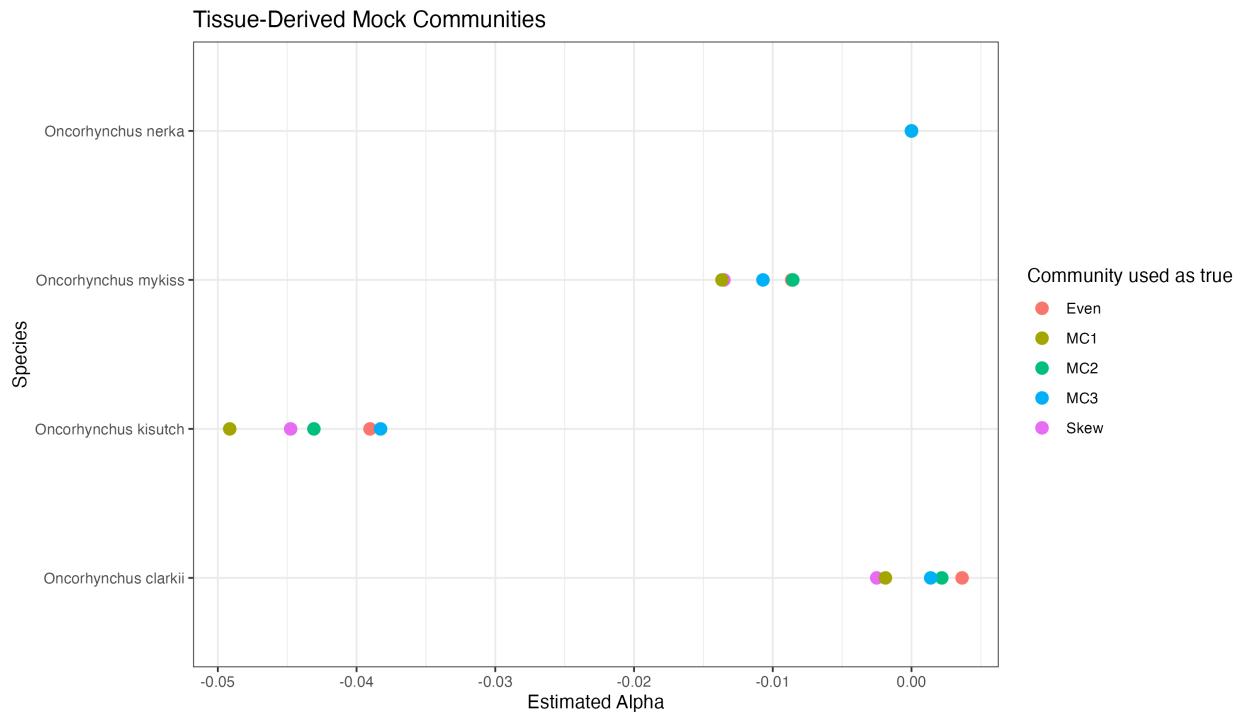


Figure S1.11. Estimated alpha values of salmonid species with different calibrations of the mock communities. Each color represents a different subset of mock community data treated as ‘true’ to calibrate the remainder of the mock community data.

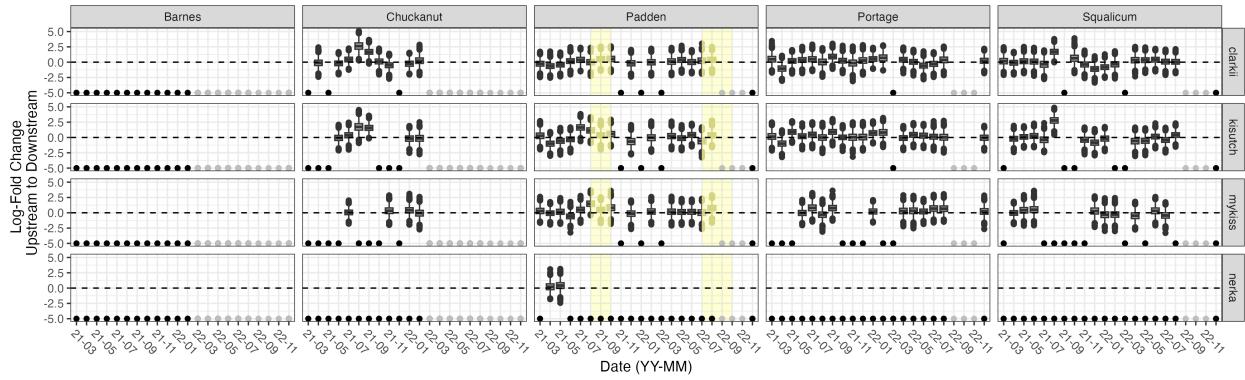


Figure S1.12. The effect of culvert on salmonid abundance separated by species and creeks across time. The y-axis shows the log-fold change in eDNA mass flow rate (copies/s) between upstream and downstream, normalized by upstream mass flow rate. The box boundaries correspond to the 25th and 75th percentiles; 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.