

¹

Skagit qPCR main text

² Andrew Olaf Shelton, Ryan Kelly esquire, Good doctor Linda Park, Piper “P-Diddy” Schwenke, James L.

³ O’Donnell, Correigh Greene, others.

4 Abstract

5 Species of conservation and ecological interest are often rare or mobile, making it difficult to count them,
6 and making time-series sampling especially labor-intensive. Sampling genetic traces of such species from
7 environmental media such as water, air, or soil (environmental DNA or eDNA) is an increasingly common
8 means of surveying, because this technique is sensitive, noninvasive, and cost-effective. However, eDNA is a
9 completely different sampling regime than traditional methods such as visual, net, or quadrat surveys, and
10 as a result, it is not clear whether or how to compare the results of genetic and traditional surveys. Here,
11 we present the results of parallel seine-net and quantitative-PCR (qPCR) time-series surveys of Chinook
12 (*Oncorhynchus tshawytscha*) and Coho (*O. kisutch*) salmon from Skagit Bay, an estuarine nearshore habitat
13 of Puget Sound, Washington, USA. The former species is listed as threatened under the U.S. Endangered
14 Species Act, and is closely monitored by NOAA. We use a highly replicated survey design to identify sources
15 of variance in fish counts (net surveys) and eDNA copy number (qPCR) over space and time; we then use
16 these variance estimates to interpret trends in each survey result over space and time over the course of a
17 seasonal salmon migration through the estuarine habitat. We develop models for relating qPCR to seine-net
18 results at two different spatial scales and assess spatial autocorrelation for each survey technique over the
19 study area. [In general, we find that eDNA closely reflects the seasonal pulse in salmonids—explaining XX%
20 of the variance at the estuary scale—and that it is more smoothly distributed over the sampled area than are
21 observed fish.]

22 Introduction

23 Methods

24 General overview of methods.

25There are four discrete steps to our qPCR methodology: (1) Environmental sample collection, (2) isolation of
26particulates from water via filtration, (3) isolation of DNA from filter membrane, and (4) amplification of
27target locus via PCR. We provide brief overviews of these steps here, and encourage the reader to review the
28fully detailed methods presented in the supplementary material (CITE SUPPLEMENT). After the laboratory
29component

30 REPLICATION IS A BIG DEAL.

31 Field Sampling

32 Beach Seine and water collection

33FROM JIMMY: we collected water samples at the surface... (Figure *site_map*). To destroy residual DNA
34on equipment used for field sampling and filtration, we washed with a 1:10 solution of household bleach
35(8.25% sodium hypochlorite; 7.25% available chlorine) and deionized water, followed by thorough rinsing
36with deionized water. Each environmental sample was collected in a clean 1 liter high-density polyethylene
37bottle, the opening of which was covered with 500 micrometer nylon mesh to prevent entry of larger particles.
38Immediately after collecting the sample, the mesh was replaced with a clean lid and the sample was held on
39ice until filtering. To assess the extent of sample cross-contamination in the field, we filled one bottle with
40deionized water before each sampling day, opened and closed it in the field, and treated it identically to the
41samples for the remainder of the steps.

42 Filtration and DNA extraction

43FILTRATION 1 liter from each water sample was filtered in the lab on a sterile filter cup fitted with a 47
44millimeter diameter cellulose acetate membrane with 0.45 micrometer pores (Sterilitech #AF045W50). Filter
45membranes were moved into 1300 microliters of Longmire buffer (CITE LONGMIRE) using clean forceps
46and stored at room temperature (CITE Renshaw2015). To test for the extent of contamination attributable
47to laboratory procedures, we filtered three replicate 1 liter samples of deionized water. These samples were
48treated identically to the environmental samples throughout the remaining protocols.

49 DNA PURIFICATION DNA was purified from the membrane following a phenol:chloroform:isoamyl alcohol
50 protocol following Renshaw (CITE RENSHAW). Preserved membranes were incubated at 65C for 30 minutes
51 before adding 900 microliters of phenol:chloroform:isoamyl alcohol and shaking vigorously for 60 seconds. We
52 conducted two consecutive chloroform washes by centrifuging at 14,000 rpm for 5 minutes, transferring the
53 aqueous layer to 700 microliters chloroform, and shaking vigorously for 60 seconds. After a third centrifugation,
54 500 microliters of the aqueous layer was transferred to tubes containing 20 microliters 5 molar NaCl and 500
55 microliters 100% isopropanol, and frozen at -20C for approximately 15 hours. Finally, all liquid was removed
56 by centrifuging at 14000 rpm for 10 minutes, pouring off or pipetting out any remaining liquid, and drying
57 in a vacuum centrifuge at 45C for 15 minutes. DNA was resuspended in 200 microliters of ultrapure water.
58 Genomic DNA extracted from tissue of a species absent from the sampled environment (*Poecilia wingei*)
59 served as positive control for the remaining protocols.

60 **Statistical Methods for qPCR results**

61 Constructing a joint statistical model for all observations provides several advantages. First, joint models
62 appropriately propagate uncertainty through the qPCR analyses from the regression analysis of the DNA
63 dilution series to the estimates of DNA concentration in field samples. Standard analyses of qPCR data
64 typically ignore uncertainty in the relationship between the density of DNA in the dilution series standard,
65 resulting in estimates of field samples ignore that are less precise and potentially biased. Furthermore,
66 constructing a joint model enables distinct processes to be modeled explicitly and compared, allowing the
67 identification of processes that contribute uncertainty. Finally, we can construct a parallel model for data
68 derived from beach seine samples, enabling a formal comparison between sampling methods using parallel
69 statistical models.

70 Following general practice with quantitative PCR, for each of $q = 1, 2, \dots, 5$ quantitative PCR plates, we ran
71 a dilution series of known Chinook salmon concentrations (densities, D , between 10-1 and 10-7 $\mu\text{g } \mu\text{L}^{-1}$) to
72 estimate the density of Chinook salmon DNA in field collections. We observe the estimated PCR cycle, C_{iq} ,
73 at which replicate PCR reaction i crossed THIS THRESHOLD (appropriate verbiage here). Then we can
74 estimate regression coefficients **beta** that determine the relationship between known DNA concentration and
75 counts,

$$C_{iq} \sim \text{Normal}(\beta_{0q} + \beta_{1q}[\log_{10} D_{qi}], \sigma^2)$$

76 For some concentrations of the standard and some field samples, no PCR amplification was observed. We
77 wanted to include this information in our model as there is no logical reason to exclude the information held

78 in these observations, so we added a presence-absence component, using logistic regression to model the
79 probability of occurrence, θ_{iq} ,

$$Y_{iq} \sim Bernoulli(\theta_{iq})$$

$$\text{logit}(\theta_{iq}) = \phi_{0q} + \phi_{1q}[\log_{10} D_{qi}]$$

80 where ϕ are regression coefficients and $Y_{iq} = 1$ if PCR amplification was observed and $Y_{iq} = 0$ otherwise. For
81 the field samples, the concentration of Chinook salmon DNA is unknown and the object of interest. We
82 define γ_{st} as the true \log_{10} Chinook DNA concentration at each site, s , in month, t . Let δ_{bst} represent the
83 deviation for each replicate bottle, b , collected at each site-month combination so that the \log_{10} Chinook
84 DNA concentration is

$$\log_{10} D_{bst} = \gamma_{st} + \delta_{bst}$$

$$\delta_{bst} \sim Normal(0, \tau^2)$$

85 Here, τ^2 represents the small scale (10s of meters, 10s of minutes), among-bottle variance in DNA concentration
86 at a given site-month combination. In parallel to the likelihood for the dilution standard, we can connect the
87 estimated Chinook DNA concentration to the observed PCR counts,

$$C_{ibstq} \sim Normal(\beta_{0q} + \beta_{1q}[\log_{10} D_{bst}], \sigma^2 + \omega^2)$$

88 Here, ω^2 is the additional variance in counts attributable to laboratory processing of the field samples beyond
89 the variance contributed by the standards. We can write the contribution of the presence component of the
90 field samples as well

$$Y_{ibstq} \sim Bernoulli(\theta_{bstq})$$

$$\text{logit}(\theta_{bstq}) = \phi_{0q} + \phi_{1q}[\log_{10} D_{bst}]$$

91 Given this statistical model, we are primarily interested in the estimates of the Chinook DNA concentration
92 at each site-month combination, γ_{st} . Additionally, we are interested in the parameters that reveal the
93 contribution of distinct processes to variation in Chinook DNA density. For the qPCR analyses we have two
94 processes that contribute to the variability in observed qPCR counts σ^2 and ω^2 while the other

95 **Constructing indices of abundance**

96 **Estimation**

97 We estimated both the qPCR and beach seine model statistical model in STAN (REF) as implemented in the
98 R environment (rstan; v.2.16.2, REF). STAN is a Hamiltonian Markov Chain Monte Carlo sampler for (see

99 this ref for a description targeted at Ecologists... UW guy) For both the beach seine and qPCR analyses
100 we used 5 parallel chains with diffuse starting locations and examined Gelman-Rubin diagnostics and We
101 used diffuse prior distributions for all parameters in the online supplemnt and provide all code necessary for
102 analysis and plotting in THIS ONLINE LOCATION.

103 at each site in each month is the object of interest

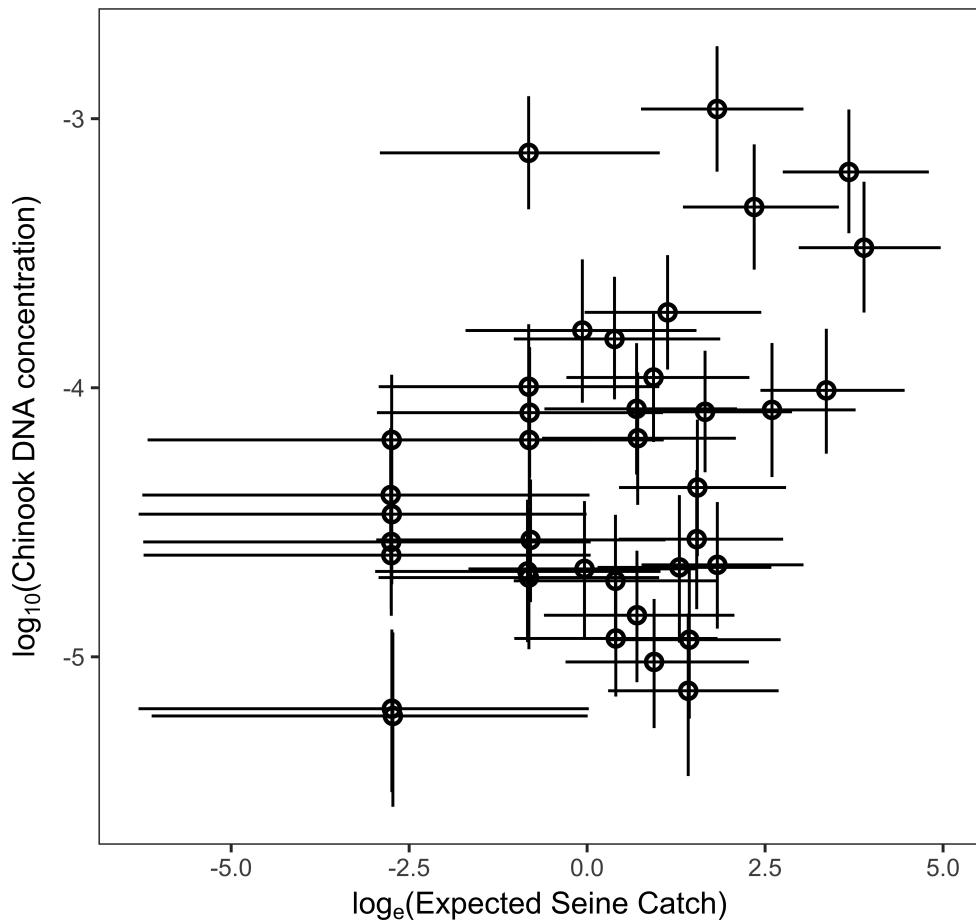
104 Results

105 ## NULL

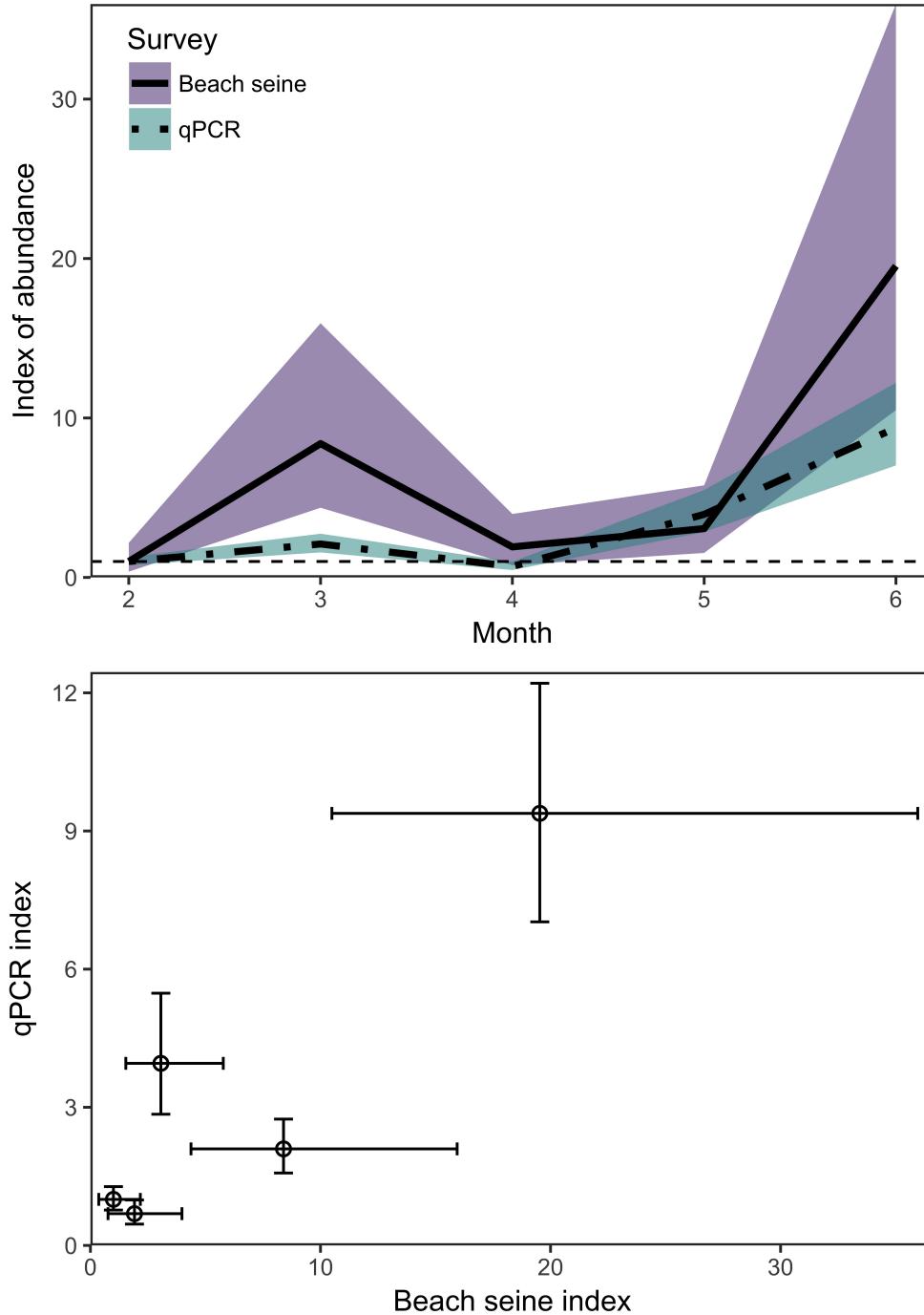
106 Discussion

107 Figures

108 ALL ERROR BARS ARE 90% Credible intervals

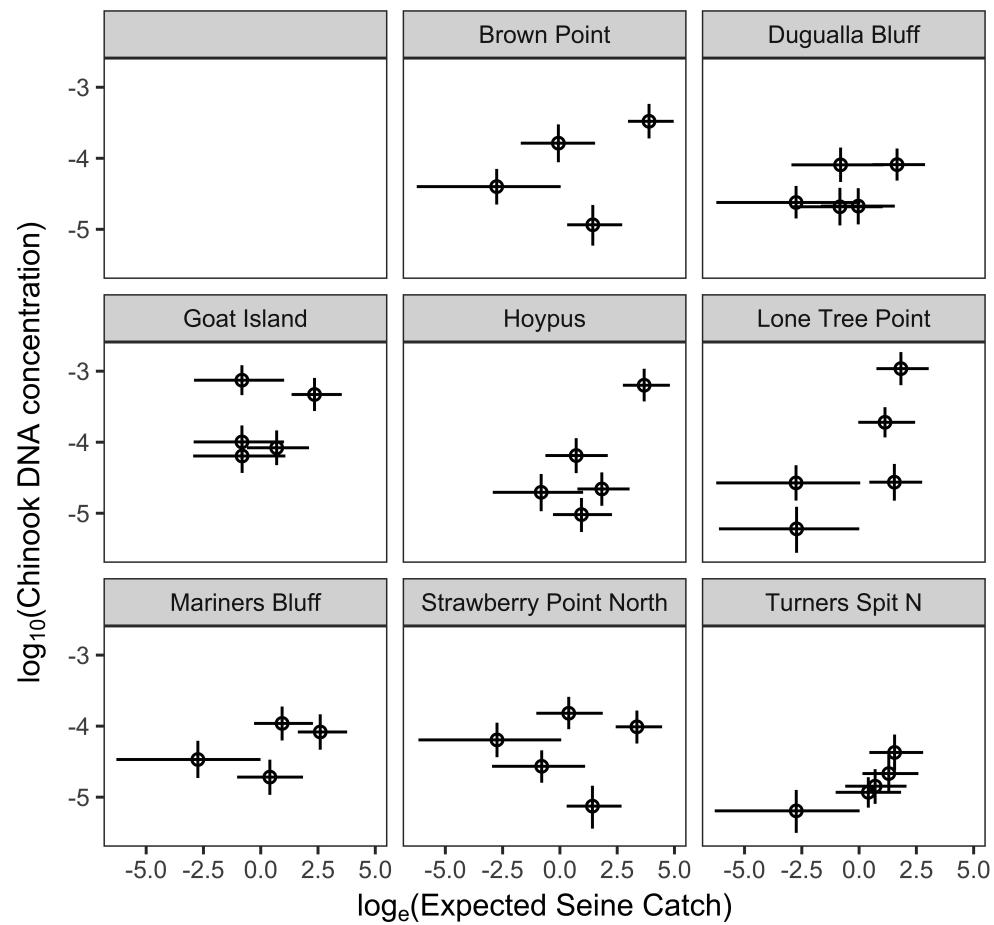


109



110

111 Supplementary Figures



112