Hi Mark.  I've been frantically trying to analyze the data we have at present and get it to the point of being worth sharing. I'm still trying to parse through the results, so apologies if the following is long-winded.  I'm still trying to figure out what is going on.  Please let me know if this doesn't make sense.

I think the bullets are fine. I might add one more which would be something along the lines of:

• Assess the spatial and temporal scales of variability for environmental DNA and conventional net sampling.

**Progress to date:**

There are really 7 components of this project:   
1) sample environment (with nets and water sample),   
2) extract and clean DNA for each water sample  
3) qPCR for species (Chinook salmon and potentially shiner surfperch)

4) multi-species sequencing

5) analyze qPCR results  
6) analyze multi-species sequencing results

7) combine results for qPCR and multi-species sequencing

**Part 1** is done: We sampled water in replicate alongside with both beach seine and fyke net sampling monthly over five months at 8 beach seine sites and 8 fyke net sites. We have around 400 water samples in total.  Beach seines are separated by a minimum of 3-5 km as the crow flies.

**Part 2** is done for a representative set of samples. We will finish extracting and cleaning samples in the near future.

**Part 3.** We have refined and tested qPCR protocols for Chinook so that they can be rapidly applied.  We have results for a majority of beach seine sites for qPCR (discussed below).  We have tested primers for shiner surf perch but have not reached production stage.  Without Jimmy (or replacement) we may not get to any further with shiner surf perch.

**Part 4.** We have run several test samples for multi-species sequencing. Production of results for the set of samples that have been used with qPCR almost sequenced.  I expect them to be put on the sequencer in the next week or two.

**Part 5.** I have been processing the qPCR results for Chinook salmon over the last few days and comparing them with beach seine samples.

**Basics:**

Chinook salmon were above the detection threshold in approximately 75% of the samples with a detection threshold of approximately 0.06 picograms / microliter. For comparison, Chinook salmon were detected in 63% of beach seine sets, so these are ballpark similar.  
  
**Variability across scales:**  
Our data allows us to look at the amount of variation due to molecular processing (e.g., PCR error) to variation within sites (among sample variation at a given site and time) to spatial variation (among site variation in each month) and to temporal variation (among month variation at each site). As expected, we can show that there is relatively low variation due to molecular processing: replicated processing of a given sample are quite consistent. Surprisingly, within site variation also had low variation: within site variation is of the same magnitude as variability due to the molecular processing.  In contrast, both among site and among month variation were much more variable (3 to 8 times as variable as within site variability). This is important because it suggests that the Chinook salmon DNA is sufficiently well mixed over small scales that it provides a detectable, consistent signal  and that this signal varies in space and time.  This is a necessary condition (but not a sufficient condition) for qPCR being able to provide estimates of abundance, so I think this is a nice preliminary result.

**Comparisons to seine samples:**

Comparing catches from beach seine samples to qPCR results directly is typically a bad idea... but people like to see it so here it is. Each dot is the value for each site at each time. If you squint and were in a positive mood, you might say there is a relationship.  But it is not overwhelming.   


Also, if we believe that the spatial distribution of the physical fish and their DNA have different pattern - i.e., if DNA can be present when fish are not - we'd actually expect this relationship to not be at the least messy and potentially for there to be no relationship at all.  
  
Additionally, we don't really care if these map perfectly onto each other, because each of these sites are supposed to be representative samples of a broader population.  We care how the eDNA results and net sampling compare at the sampling scale of interest.  In our case, we are interested at the scale of Skagit Bay. We can look at Skagit wide averages by month and this looks more promising:



We'll need to add a couple more sites and samples, but this is relatively positive. It suggests that even if the individual samples do not reflect the catches in nearest beach seine, at the broader scale eDNA and beach seine sample may provide similar answers about total Chinook salmon abundance.

An alternative way to look at this is by constructing an index of abundance.  This is classic index standardization from the stock assessment world. I standardized to the samples taken in February (i.e., February is an index of 1) and all subsequent months are relative to February (e.g., a value of 2 indicates a doubling of abundance).  
  


There are three interesting things here.   
  
First, the pattern of variability in the mean is broadly similar (peaking in June for both, lowest values in April and February). This suggests that both are picking up aspects of the true Chinook abundance signal.   
  
Second,the uncertainty about these estimates are roughly equivalent for eDNA and beach seines (the width of the shaded region are about the same at a given y-axis value) indicating that there are not likely strong gains in precision from eDNA over beach seine sampling.   
  
Third, the seine sampling is far more variable than the eDNA results (~25 fold variability for beach seines, ~8 fold variation for eDNA). I need to think about this more, but I think this might be indicative of how responsive the eDNA is to pulses of fish moving through the nearshore.  It also might be part of a way to determine something about the ability to see extreme high and low abundance areas or times using eDNA.

**Parts 6 and 7**. These require results from the sequencing run. We have a statistical framework for this but we can't apply it until the data comes in.

Best

Ole