In Silico – Design and computer validation for PCR probes

This document provides a general checklist of steps to develop a PCR probe, but is not intended to be detailed guidance.

1) Obtain a reference mitochondrial sequence for the target species.
   a) Prefer full mitochondrial genome, but if that is not available pick a reference sequence of a mitochondrial gene.
   b) Prefer commonly sequenced regions of common genes, Ie COI, 12S… This helps ensure that we are selecting unique sequences, not just unique to the database sequences.
   c) Prefer genes that we have metabarcoding primers for. This will allow us to use metabarcoding as a directly comparative measure.
2) Run NCBI Primer-BLAST, with the reference sequence as the PCR Template
   a) PCR product size should be set to 70 – 140
   b) The specificity check should be run against the nr database.
      i) Organism should be left blank
   c) If there are multiple primer options which identically match the target, prefer the primer that is:
      i) On COI or 12S, or if not possible on another commonly used gene.
      ii) Has the fewest potentially unintended targets.
      iii) Has the fewest unintended targets closely related / geographically located near to the target species.
      iv) If there are unavoidable potentially unintended targets, record them for probe design.
3) Use IDT PrimerQuest, with the same reference sequence and the primers designed by Primer-BLAST.
   a) Enter sequence and select qPCR 2 Primers + probe
   b) Show custom design parameters and enter the chosen forward and reverse primers.
   c) Run it
   d) If the forward and reverse primers were not specific to just the target species;
      i) Take the probe and run a BLASTn against all of the unintended targets returned by Primer-BLAST
      ii) If the probe has mismatches with all of the unintended targets you are all set.
      iii) If the probe does not have mismatches with any of them, use BLASTn to align the target reference with the troublesome unintended target.
         (1) In that alignment identify the location of the primer suggested by IDT.
         (2) Identify the smallest shift of the location of the probe that would introduce a mismatch between the two sequences.
         (3) Return to " i) " with the new shifted probe, make sure the GC content didn't change too much, and that you didn't add in something like "GGGGG"…

In Vitro – Positive control, calibration, and basic negative control.

1) Order a synthetic gene fragment of the region containing the amplified sequence for the target species.
    a) Use the same reference sequence as was used to design the probes.
    b) gBlocks Gene Fragments, whatever is the cheapest price that will contain at least the target sequence.
    c) Select a region of the reference sequence that is the longest possible for the price point, currently 500 bp, try and center the amplified region as much as possible.
    d) If the ordering tool complains about regions on the edge of the sequence that can be easily removed do so.
2) dPCR calibration
    a) Run positive control, negative control
        i) Positive control, aim to "saturate" the chip with the synthetic sequence.
        ii) Negative control, water with no synthetic sequence.
    b) Run calibration chip
        i) Dilute synthetic reads to ~10,000 copies for the 20,000 well chip. We want between 10% and 90% occupancy. Amount (ng) = length (bp) * 1.1 * 10^(-7)
3) qPCR calibration based on Forootan et al.
    a) Run a positive control, negative control
        i) Positive control, use "a lot" of DNA
        ii) Negative control, run water with no synthetic sequence.
    b) 2 fold dilution series from 1 – 64 molecules, with 24 wells each, except for 1 with 48 wells
        i) Can be scaled, precision depends on the number of replicates and accuracy depends on the spacing of the dilutions. Forootan et al. suggests the most pragmatic approach is to do a course series to estimate the Limit of Detection, then hone in as needed.

In Vivo – Test with real samples

1) Begin with sample(s) taken with as many other methods of detection as well.
    a) Negative controls
        i) Lab blank
        ii) Location blank
    b) Positive control
        i) Spike a sample with the reference dna (idea from Perez et al.).
            (1) Aim for ~10-20x the Limit of Detection, Perez et al. used 100 molecules.
        ii) Run / compare to a calibrated pure synthetic positive control to see the inhibition factor.
    c) Suspected positives, from detection with other methods.
        i) "bucket" samples where we can be almost positive there is dna in the sample.
        ii) "Normal" samples
            (1) what volume is needed to get a positive?
            (2) How does the sensitivity compare to other sampling methods?
    d) Suspected negatives, from no detection with other methods.
        i) Where there is almost certainly none of the species, like a sample from Oregon for an east coast only fish…
        ii) Where we just did not get any hits from other methods.
            (1) How does the sensitivity compare to other sampling methods?