**Supporting information for review and publication**

**Evaluating bioinformatics pipelines for population level inference using environmental DNA**



**Table S1.** Sequences of the intraspecific barcode obtained after amplification with the primer pair MS-DL06, sequencing and initial trimming for the 39 *Mullus surmuletus* individuals of the aquarium and used for the reference database of the aquarium and sea samples.

 **Table S2.** Sequences of the intraspecific barcode obtained after amplification with the primer pair MS-DL06, sequencing and initial trimming for the 21 *Mullus surmuletus* individuals sampled in different locations and used for the reference database of the sea samples.

**Table S3.** Parameters for each marker obtained with ECOPCR from the D-loop of *Mullus surmuletus*.

**Table S4.** Sensitivity, haplotype precision and read precision obtained for each aquarium, and mean sensitivity, mean haplotype precision and mean read precision obtained for each pipeline and their corresponding standard deviation.

**Figure S1.** *In silico* amplification of the target species *Mullus surmuletus* and of the non-target species with the primer pairs MS-DL01 to MS-DL05, according to the number of mismatches in the forward and reverse primers. Species amplified *in silico* which have a very low probability of being sampled with *M. surmuletus* are represented in blue (freshwater or terrestrial species). Circle diameter is proportional to the number of sequences from the dataset which are amplified.

This figure was created with the ROBITaxonomy (LECA, 2012), ROBIBarcodes (LECA, 2013) and ROBITools (LECA, 2012) R packages.

**Figure S2. (A)** *In silico* amplification of the target species *Mullus surmuletus* and of the non-target species with the primer pair MS-DL06, according to the number of mismatches in the forward and reverse primers. Species amplified *in silico* which have very low probability of being sampled with *M. surmuletus* are represented in blue (freshwater or terrestrial species). The circle diameter is proportional to the number of sequences from the dataset which were amplified. **(B)** Sequences of forward and reverse primers (5’->3’).
This figure was created using the ROBITaxonomy (LECA, 2012), ROBIBarcodes (LECA, 2013) and ROBITools (LECA, 2012) R packages.

**Figure. S3.** Results for ocean eDNA samples analyzed with pipeline B2 with a threshold of 1/1000 per site for the minimum relative read count in both samples. **(A)** Bar charts representing the number of haplotypes in the reference database, false positive haplotypes, and possibly true haplotypes returned for eDNA samples from Banyuls and Calvi, in comparison to the composition of the reference dataset built to design the barcode. **(B)** Scatterplot overlapping the number of reads for haplotypes in the reference database, false positive haplotypes, and “possibly true haplotypes”. A decimal logarithmic scale is used for the vertical axis.