Environmental DNA analysis needs local reference data to inform taxonomy-based conservation policy – A case study from Aotearoa / New Zealand

Paul Czechowski[1a*], Michel de Lange[2,3,b] Michael Heldsinger[4c], Will Rayment[4,5d], Christopher Hepburn[4,5e], Monique Ladds[6f], and Michael Knapp[1,5g]

[1]Department of Anatomy, University of Otago, PO Box 56, Dunedin 9054, Aotearoa / New Zealand; [2]Biostatistics Centre, University of Otago, PO Box 56, Dunedin 9054, Aotearoa / New Zealand; [3]Pacific Edge, 87 Saint David Street, Dunedin 9016, Aotearoa / New Zealand; [4]Department of Marine Science, University of Otago, PO Box 56, Dunedin 9054, Aotearoa / New Zealand [5]Coastal People Southern Skies Centre of Research Excellence, University of Otago, PO Box 56, Dunedin 9054, Aotearoa / New Zealand; [6]Department of Conservation PO Box 10 420, Wellington 6143, Aotearoa / New Zealand.

[a]paul.czechowski@otago.ac.nz; [b]michel.delange@posteo.nz;

[c]michael.heldsinger@gmail.com;[d]will.rayment@otago.ac.nz;

[e]chris.hepburn@otago.ac.nz;[f] mladds@doc.govt.nz; [g]michael.knapp@otago.ac.nz

Running title: eDNA in marine reserves of Aotearoa

Keywords: BRUV, conservation, fish, environmental DNA, marine reserve, visual survey, sharks

Article Type: Perspectives

Number of words: 3195 (Introduction up to and including Acknowledgements)

Number of references: 29

Number of Figures and Tables: 4 + 1

Corresponding Author: Paul Czechowski, Department of Anatomy, University of Otago, PO Box 56, Dunedin 9054, Aotearoa / New Zealand; paul.czechowski@otago.ac.nz; Phone +64 021 024 91263

Abstract

26

27 Effective management of biodiversity requires regular surveillance of multiple species.

28 Analysis of environmental DNA by metabarcoding (eDNA) holds promise to achieve this

29 relatively easily. However, taxonomic inquiries into eDNA data need suitable molecular

30 reference data, which are often lacking. We evaluate the impact of this reference data void in

31 a case study of fish diversity in the remote fiords of New Zealand. We compared eDNA-

32 derived species identifications against Baited Remote Underwater Video (BRUV) data

33 collected at the same time and locations as the eDNA data. Furthermore, we cross referenced

34 both eDNA and BRUV data against species lists for the same region obtained from literature

35 surveys and the Ocean Biodiversity Information System (OBIS). From all four data sources,

36 we obtained a total of 116 species records (106 ray-finned fishes, 10 cartilaginous fishes; 59

37 from literature, 44 from eDNA, 25 from BRUV, 25 from OBIS). Concordance of taxonomies

38 between the data sources dissolved with lowering taxonomic levels, most decisively so for

39 eDNA data. BRUV agreed with local biodiversity information much better and fared better in

40 detecting regional biodiversity dissimilarities. We provide evidence that eDNA

41 metabarcoding will remain a powerful but impaired tool for species-level biodiversity

42 management without locally generated reference data.

Introduction

44 Marine reserve (MR) networks conserve biodiversity by stabilizing communities and

45 maintaining food web structure (Wing & Jack, 2013). Effective management of MR

46 biodiversity requires regular surveillance, for example to avoid overexploitation by fishing

47 (Wing & Jack, 2013), or to avoid damage through influx of non-indigenous species

48 (Cunningham, 2019). Fish surveillance is of particular interest due to their sensitivity to most

49 forms of human disturbance, their usefulness at all levels of biological organization, and the

50 favourable benefit-to-cost ratio of fish assessment programmes (Harris, 1995).

51 Analysis of environmental DNA (eDNA) metabarcoding data is a well-established molecular

52 technique for multispecies surveys (Cristescu & Hebert, 2018). Environmental DNA

53 metabarcoding holds promise for biodiversity surveys intended to inform biodiversity

54 management – associated techniques are regarded more cost-efficient than traditional

55 methods (such as baited remote underwater video surveys – BRUV), less dependent on expert

56 taxonomic knowledge, can be standardized, and are able to inform on a broad range of taxa

57 (Sigsgaard et al., 2020).

58   Reliable low-level taxonomic annotation is a prerequisite for useful biodiversity management
59   and biological surveillance (e.g., Jack & Wing 2013). For example, in a southern New
60   Zealand context, *Parapercis colias* (blue cod) is of high commercial interest, but three other
61   of the 79 cod species are known from New Zealand (Roberts et al., 2019), so that genus
62   information alone is ambiguous for determining blue cod presence or absence. Accordingly,
63   higher-level taxonomic classifications (e.g., family, and order levels) are even less
64   informative for species level conservation management. This reality translates into the desire
65   for obtaining perfect 100 bp to 200 bp alignments (Huson et al. 2007) between an unknown
66   eDNA-derived query sequence and a well described reference sequence derived from a valid
67   species. In practice, absence of such reference data necessitates relaxation of taxonomy-
68   assigning alignment parameters to retain sufficient eDNA data for analysis, and in
69   consequence the data's informative quality suffers.

70   Availability of suitable reference data for metabarcoding is highly variable depending on
71   taxonomic groups and geographic locations, with fish considered relatively well covered in
72   Barcode of Life Data Systems (BOLD) and NCBIs GenBank (Benson et al. 2011) for some
73   regions such as Europe (Weigand et al., 2019). Arguably, fewer reference data are available
74   for fish of southern New Zealand. For example, for six commonly used 12S primer pairs,
75   recognized as well suitable for fish multispecies surveys (Weigand et al., 2019), an average
76   of 36% of all northern European fish species are available as reference data, but only 26% of
77   southern New Zealand species (GAPeDNA v1.0.1 web interface, 11-Sep-2021; Marques et
78   al. 2021; also see SI Table 1).

79   In this study, we evaluate the impact of taxonomic data limitations on multispecies surveys
80   using the example of fish in the UNESCO World Heritage Site Te Wahipounamu (Fiordland)
81   in southern New Zealand. We compare the results of concurrent eDNA and BRUV surveys
82   and cross-reference these data against species lists for the same region obtained from
83   literature and the Ocean Biodiversity Information System (OBIS). The fish diversity of Te
84   Wahipounamu has been described based on a diverse range of mostly visual methods. If our
85   BRUV and eDNA approaches work optimally, we should see a strong overlap between these
86   field data and previously described fish diversity records of the region.

Methods

88   For this study, we evaluated the presence of Actinopterygii (ray-finned fishes) and
89   Chondrichthyes (cartilaginous fishes) species in one MR, two commercial exclusion zones

90 (all "MR"), and corresponding control areas in southern Te Wahipounamu, New Zealand

91 (west coast, approximately from -44.3 to -46.25 Southern latitude; Fig. 1a). We obtained and

92 analysed eDNA and BRUV data as well as electronic records proximate to the field work

93 area from the Ocean Biodiversity Information System (OBIS; Ausubel 1999). Furthermore, a

94 reference list of ray-finned fishes and cartilaginous fishes that have been observed in

95 Fiordland was assembled from literature. All observations were formalized using NCBI

96 taxonomy (Federhen, 2012), including trivial names, and limited to classes Actinopterygii

97 and Chondrichthyes.

98 Literature data, itself obtained using various methods, were extracted from five sources,

99 including one meta-analysis (SI Table 2). OBIS data were downloaded for a 38 km radius

100 around all field work sites (centre point W 166.89°, S -45.80°), as well as for smaller areas

101 surrounding individual field work sites, (2.5 km radius; Fig. 1a).

102 For a detailed description of field and laboratory work please refer to SI. For eDNA

103 collection and BRUV filming we visited three locations in southern Te Wahipounamu

104 (Moana Uta / Wet Jacket Arm, Taumoana / Five Fingers, and Te Tapuwae a Hua / Long

105 Sound; henceforth WJ MR, FF MR, and LS MR), and accompanying control areas outside

106 those MRs (henceforth WJ CTRL, FF CTRL, and LS CTRL), from 12.–22. December 2019

107 (Fig. 1a.) Within each sampling location, at randomised sites, we collected eDNA (mean

108 depth 14.05 m, med.: 15, sd.: 1.4 m), and subsequently deployed BRUV assemblies (mean

109 depth 15.6 m, med.: 16, sd.: 2.6 m; SI). We considered data from 21 sites (FF: 2 FF MR and

110 3 FF CTRL, WJ: 4 WJ MR and 4 WJ CTRL, and LS: 4 LS MR and 4 LS CTRL). We

111 collected two 900 ml water samples with eDNA at each site, filtered them alongside negative

112 controls, then sealed and stored them until further processing. BRUV footage was obtained

113 for one hour and analysed by eye.

114 Environmental DNA was isolated in a PCR-free facility alongside extraction and cross-

115 contamination controls (SI: four species of tropical freshwater fish). After *in silico* PCR (SI),

116 we amplified our extracts with the well-established and widely used 12S MiFish primers

117 ("MiFish-U"; Miya et al. 2015; see SI Table 1 for primer comparison), targeting

118 Actinopterygii. Chondrichthyes were targeted with slightly altered derivatives ("Elas02",

119 Taberlet et al. 2018). Our single-step PCRs were cycled 45 times, with annealing

120 temperatures of 45 °C (MiFish-U) or 40 °C (Elas02). Amplified eDNA was then pooled,

121 visualised, purified, combined equimolarly, diluted to 4.5 pmol, and sequenced on an

122 Illumina MiSeq (Illumina, San Diego, US-CA; kit v2, 300 cycles, single-ended).

123   We defined Amplicon Sequence Variants (ASVs; Callahan et al. 2017) from eDNA after
124   demultiplexing with Cutadapt v3.0 (Martin, 2011), using Qiime2 2020-08 (Bolyen et al.,
125   2019) and DADA2 1.10.0 (Callahan et al., 2016). To yield high quality sequence data we did
126   not allow any mismatches, nor Expected Errors (Edgar & Flyvbjerg, 2015) during
127   demultiplexing. Taxonomic annotation of denoised data was obtained using Blast 2.10.0+
128   (Camacho et al., 2009) and a local full copy of the NCBI nucleotide collection (April 2020;
129   Benson et al., 2011) while excluding environmental samples. To yield a maximum of
130   taxonomically annotated ASVs, we chose relaxed taxonomic assignment parameters in
131   combination with an e-value to retain only the most significant alignments. We required a
132   minimum identity of 75% among all alignments and kept five high-scoring pairs for each
133   eDNA query, each of which needed a minimum coverage of 95% to be retained. The minimal
134   acceptable e-value was set to 10^{-10}. We retained the best high-scoring alignment of each
135   query-reference pair based on the highest bit score. We removed data contained in negative
136   controls, alongside ASVs covered by fewer than 15 reads (see SI).

137   To investigate how well the literature- and OBIS-derived biodiversity information were
138   resolved by eDNA and BRUV, we checked the concordance of all data sources on order,
139   family, genus and on species levels. To judge sampling effort and total species diversity
140   based on BRUV and eDNA observations, we inspected species accumulation curves and
141   calculated Good Turing estimators (giving the number of all species based on species already
142   seen in a small sample; Good, 1953), and then compared those values to combined Te
143   Wahipounamu literature and OBIS species records.

144   To verify the credibility of eDNA information, we checked all eDNA species lists against a
145   comprehensive list of all New Zealand fish (Roberts et al., 2019) and evaluated species
146   assignment and alignment qualities.

147   To investigate how useful BRUV, eDNA and OBIS records are in detecting regional
148   differences between fish biodiversity, we used Analysis of Similarity (ANOSIM; Clarke
149   1993). Thereby analysing Jaccard distances (Jaccard, 1912), we looked for significant
150   differences in taxon (species, genus, family, order) overlap depending on various factor
151   combinations, hence checking whether a particular observation method fared better in
152   detecting taxon composition differences either between different field work areas (WJ, FF,
153   LS), or according to protection status (MR or CNTRL).

Results

155 Each of our four data sources yielded different species counts. In total, we yielded 116
156 species (106 Actinopterygii, 10 Chondrichthyes), comprised of 59 species from previously
157 published Te Wahipounamu works, 44 unique species derived from eDNA, 25 from BRUV
158 and 25 from OBIS (large area; see Table 1, Fig. 2, SI. Table 4). While 21 field work sites
159 (Fig. 1a) yielded environmental DNA and BRUV data (Fig. 1b, c) matching local OBIS data
160 could only be obtained for nine field work sites (Fig. 1a, small circles, LS CNTRL, FF, WJ),
161 and hence those finer scaled OBIS data were later excluded from ANOSIM as incomplete
162 data (Fig. 1a, small circles, Fig. 1d). For further comparisons with Good-Turing estimates, we
163 posit the local "real" species count to 68 as the number of unique species observed across
164 literature and OBIS (Fig. 3).

165 Obtaining community composition comparable to literature and OBIS data within our works'
166 spatial constraints worked better with BRUV than with eDNA. Nineteen out of 25 species
167 detected with BRUV (76%) were contained in the literature or on OBIS, but only one out of
168 44 species detected with eDNA (2%) were contained in Te Wahipounamu-specific literature
169 or OBIS (Fig. 1a, large circle). Concordance of taxonomic information between the four data
170 sources dissolved with lowering taxonomic levels, and most decisively for eDNA data (Fig.
171 3). At species level, only two taxonomic assignments from eDNA matched other data
172 sources, namely *Notorynchus cepedianus* (broadnose sevengill shark), also found with
173 BRUV, and *Aldrichetta forsteri* (yellow-eye mullet) also listed in the literature (Fig. 2).
174 BRUV agreed better with available local biodiversity information, with 11 species mentioned
175 both in the literature and OBIS, and eight detected species mentioned in the literature only
176 (Fig. 3). On BRUV we identified six species *(Bodianus unimaculatus, Chelidonichthys kumu,*
177 *Galeorhinus galeus, Mustelus lenticulatus, Notorynchus cepedianus, Scorpaena cardinalis)*
178 not mentioned in literature, and not in OBIS, but occurring in New Zealand waters (Roberts
179 et al., 2019). While our plateauing species accumulation curves suggested exhaustive
180 sampling (SI Fig. 5), Good-Turing estimates of eDNA data inferred a presence of 60 species
181 in the study area (assuming 27% missed after 44 observations), and a presence of 26 species
182 using BRUV (assuming 7% missed after 25 observations).

183 Alignment qualities associated with taxonomic annotation of eDNA data were variable.
184 Forty-four species assigned among eDNA were defined by 92 ASVs (across 142
185 observations) of which only six yielded flawless alignments with reference data (i.e. 14%,
186 with full query coverage, no alignment gaps). Eighty-six ASVs had variable query coverage

187　(37 families, Tab. 1 and SI), while 32 ASVs had variable gap counts (15 families, Tab. 1 and

188　SI). Mean query coverage was 93.2% (min: 78.6%, med: 97%, sd 6.5%), mean gap count was

189　1 (max: 10, med: 0, sd 1.83; Tab. 1). Nineteen species assignments among eDNA (43%) had

190　not been observed in New Zealand, and none of these species were found using BRUV,

191　across literature, or OBIS data (apart from *Bovichtus variegatus* – thornfish, not in Roberts et

192　al., 2019, but in Roberts, 2005; Fig. 2). Importantly, 25 species observed with eDNA (56.8%

193　of eDNA-observed species) were known from somewhere New Zealand (Roberts et al., 2019)

194　but were not observed in BRUV or found in Te Wahipounamu literature. Interestingly, using

195　eDNA, we obtained perfect alignments between few ASVs and reference data for

196　*Arctocephalus forsteri* (New Zealand fur seal), *Balaenoptera musculus* (blue whale), and

197　*Tursiops truncatus* (bottlenose dolphin).

198　In ANOSIM, only BRUV data, and not eDNA data, exhibited location-specific differences

199　among species' presence overlaps among the 21 sites – on species, genus, family, and order

200　levels. Significant differences were calculated in overlaps between the six field work areas

201　but not between marine reserve nor control areas (SI Table 3).

202　Investigation of the strikingly homogenous structure of eDNA data by regression analysis of

203　the 142 non-unique eDNA observations (Tjur's R^2 0.027) suggested each additional

204　alignment gap to be associated with a 39% increased probability of observing a non-native

205　species (Odds Ratio 1.39, 95% CI from 1.19 to 1.66, p <0.01). A 1% increase in alignment

206　concordance was associated with a 7% increased probability of non-native observation (OR

207　1.07, 95% CI 1.03–1.12, p <0.01). Null deviance was 572.60 on 141 degrees of freedom,

208　residual deviance was 552.31 on 139 degrees of freedom (SI Figs 7 and 8).

209　**Discussion**

210　What is a realistic estimate of the fish biodiversity in Te Wahipounamu? Based on literature

211　and OBIS alone, we estimate the currently described combined ray-finned and cartilaginous

212　fish species count of Te Wahipounamu to be 68, minding that we constrained OBIS data to

213　surround field sites (Fig. 1, large circle), and that those data are predominantly based on

214　visual observations (SI Table 2). If species counts obtained from literature and OBIS were

215　close to a real value of 68, and the same was true for eDNA and BRUV data, both respective

216　Good-Turing estimates would be 68. Our BRUV-based Good-Turing estimate of 26 species

217　diverges strongly from this number. This may have several reasons. Firstly, we only

218　inspected an isolated area in Te Wahipounamu, while the literature describes a larger area.

219   Secondly, bait in BRUV does not attract all fish for the camera, particularly if deployed at

220   limited depth range, as done here. For eDNA, the Good-Turing estimate of 60 species is more

221   like the literature-inferred species count, but this could be coincidental.

222   How credible are eDNA derived species assignments with currently available reference data?

223   We believe lacking eDNA reference data to restrict accurate species annotation of ASVs.

224   There are several observations from our data that appear to support this hypothesis. First,

225   while there is a reasonably good concordance between species identified in our BRUV

226   analyses and species known from the area as combined from publications and OBIS, the

227   dissimilarity between eDNA data on one side, and BRUV, OBIS and publication data on the

228   other side, increases with decreasing taxonomic level, culminating in only two out of 44

229   eDNA species being either identified in our BRUV analyses or known from previous

230   publications (Fig. 3).

231   Secondly, every approach to identify species diversity in a marine ecosystem has its biases,

232   and published observations are mostly based on visual approaches. Thus, one could argue for

233   the existence of a bias favouring similarity between our visual BRUV observations and

234   published species occurrences to the detriment of eDNA data's similarity. However, we do

235   not believe this circumstance alone to be responsible for a bias favouring BRUV data to be

236   more similar with literature and OBIS observations in comparison to eDNA observations.

237   Literature and OBIS observation methodologies extend well beyond the specific biases of

238   BRUV, including a multitude of different observation techniques (poison stations, seine net

239   fishing, spear fishing, diver surveys and others, SI Table 1). Collectively, all observation

240   techniques should have provided an appropriately comprehensive overview of fish diversity

241   in Te Wahipounamu, lacking biases inherent to BRUV.

242   Thirdly, some divergence between eDNA data and the other data sources may be explained

243   by the known ability of eDNA to detect "cryptic" species that are not easily discovered by

244   any visual surveying. The most obvious candidates for this category would be the 25 eDNA

245   species that had previously been reported from New Zealand but not yet from Te

246   Wahipounamu (Fig. 2). However, such a bias should not prevent a broad overlap between

247   eDNA and visual approaches for species that can easily be detected visually. Clearly, we did

248   not find such an overlap.

249   Crucially, of the 25 species we detected by BRUV and that were therefore present at the time

250   of our concurrent water sampling for eDNA analyses, 24 species are not present in the NCBI

251 reference database (Fig. 2) and could therefore not be detected by our eDNA approach. This
252 highlights one of the main limitations of eDNA multispecies surveys today.

253 Nevertheless, and despite the lack of reference data, eDNA still identified a larger number of
254 species than our concurrent BRUV analyses. From where do these species assignments
255 come? In most cases during taxonomic assignment, where no perfect match can be found
256 between eDNA query and reference subject sequence, the algorithm assigning ASVs to
257 species information (BLAST) chose the next-closest matching species contained in the
258 reference data collection, as encouraged by our taxonomic assignment parameters. Our
259 taxonomic assignments correspond with this hypothesis, as binomial regression showed that
260 each additional gap in a sequences' reference alignment associated with a 39% increased
261 probability of observing of a non-native species.

262 Interestingly, a 1% increase in alignment concordance increased the likelihood of a non-
263 native observation as well, by 7%. At first sight this seems counter-intuitive, however the
264 latter observation is also consistent with our hypothesis: A poorly matching sequence would
265 not be assigned to a matching species but rather to a higher matching taxon such as genus or
266 family. A better fit increases the likelihood of a species level assignment, but without native
267 species contained among reference data, the likelihood increases that the query sequence is
268 assigned to a closely related species not occurring in New Zealand. Similar observations have
269 been made in other regions of the world (Stoeckle, Das Mishu, & Charlop-Powers, 2020).

270 The large number of species detected by our eDNA approach – although probably
271 misassigned in several instances – is a testament to the potential power of eDNA methods.
272 Arguably, any detected effect of lacking reference data could be less pronounced by using
273 another, or multiple primer pairs. For example, our primer evaluations with the recently
274 released software GAPeDNA (Marques et al., 2021) show that the "Fish 16S" primer set by
275 McInnes et al. (2017) would have covered 249 instead of the 119 New Zealand marine fish
276 species covered by our MiFish 12S dataset (SI Table 1). However, the overall conclusion
277 remains. Of the over 1294 known New Zealand marine fish species, molecular reference data
278 of any kind is available only for 489 species in southern New Zealand, and for no available
279 primer pairs sufficient reference data is available. Hence without substantial effort into
280 generating suitable reference data for a carefully selected range of similar primers, eDNA
281 analysis here and everywhere else will remain an impaired tool for biodiversity management.
282 While this insight holds true after almost two decades of eDNA research (Hebert, Cywinska,
283 Ball, & DeWaard, 2003) we note that a growing number of researchers are working hard on

284   closing reference data voids around the globe (reviewed in Marques et al., 2021; Weigand et

285   al., 2019).

Acknowledgments and Data

References

297   Ausubel, J. (1999). Toward a Census of Marine Life. *Oceanography*, *12*(3), 4–5. doi:

298   10.5670/oceanog.1999.17

299   Benson, D. a, Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2011).

300   GenBank. *Nucleic Acids Research*, *39*(Database issue), D32-7. doi:

301   10.1093/nar/gkq1079

302   Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., …

303   Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome

304   data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. doi:

305   10.1038/s41587-019-0209-9

306   Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should

307   replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*,

308   *11*(12), 113597. doi: 10.1038/ismej.2017.119

309   Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S.

310   P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data.

311   *Nature Methods*, *13*(7), 581–583. doi: 10.1038/nmeth.3869

312   Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden,

313   T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1),

314   421. doi: 10.1186/1471-2105-10-421

315   Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community

316   structure. *Austral Ecology*, *18*(1), 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x

317   Cristescu, M. E., & Hebert, P. D. N. (2018). Uses and Misuses of Environmental DNA in

318   Biodiversity Science and Conservation. *Annual Review of Ecology, Evolution, and*

319   *Systematics*, *49*(1), 209–230. doi: 10/gfkkjq

320   Cunningham, S. (2019). Mitigating the threat of invasive marine species to Fiordland: New

321   Zealand's first pathway management plan. *Management of Biological Invasions*,

322   *10*(4), 690–708. doi: 10.3391/mbi.2019.10.4.07

323   Edgar, R. C., & Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for

324   next-generation sequencing reads. *Bioinformatics*, *31*(21), 3476–3482. doi:

325   10.1093/bioinformatics/btv401

326   Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Research*, *40*(D1), D136–

327   D143. doi: 10/c452q3

328   Good, I. J. (1953). The Population Frequencies of Species and the Estimation of Population

329   Parameters. *Biometrika*, *40*(3/4), 237. doi: 10/d7kccb

330   Harris, J. H. (1995). The use of fish in ecological assessments. *Austral Ecology*, *20*(1), 65–80.

331   doi: 10/d2pq6f

332   Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological

333   identifications through DNA barcodes. *Proceedings of the Royal Society B:*

334   *Biological Sciences*, *270*(1512), 313–321. doi: 10/fnjqv7

335   Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of

336   metagenomic data. *Genome Research*, *17*(3), 377–386. doi: 10.1101/gr.5969107

337   Inglis, G., MAF Biosecurity New Zealand, Post-border Directorate, & MAF Biosecurity New

338   Zealand. (2008). *Milford Sound: First baseline survey for non-indigenous marine*

339   *species (research project ZBS2005/19)*. Wellington, N.Z.: MAF Biosecurity New

340   Zealand.   Retrieved   from   https://niwa.co.nz/static/marine-

341   biosecurity/Inglis%20et%20al%202008%20milford%20resurvey%20report.pdf

342   Jaccard, P. (1912). The distribution of the flora in the Alpine zone. *New Phytologist*, *11*(2),

343   37–50. doi: 10/fvhsjd

344   Jack, L., & Wing, S. R. (2013). A safety network against regional population collapse:

345   Mature subpopulations in refuges distributed across the landscape. *Ecosphere*, *4*(5),

346   57. doi: 10.1890/ES12-00221.1

347   Marques, V., Milhau, T., Albouy, C., Dejean, T., Manel, S., Mouillot, D., & Juhel, J.-B.

348   (2021). GAPeDNA: Assessing and mapping global species gaps in genetic databases

349   for eDNA metabarcoding. *Diversity and Distributions*. doi: 10.1111/ddi.13142

350   Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing

351   reads. *EMBnet.Journal*, *17*(1), 10. doi: 10.14806/ej.17.1.200

352   McInnes, J. C., Jarman, S. N., Lea, M.-A., Raymond, B., Deagle, B. E., Phillips, R. A., …

353   Alderman, R. (2017). DNA Metabarcoding as a Marine Conservation and

354   Management Tool: A Circumpolar Examination of Fishery Discards in the Diet of

355   Threatened Albatrosses. *Frontiers in Marine Science*, *4*, 277. doi:

356   10.3389/fmars.2017.00277

357   Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., … Iwasaki, W. (2015).

358   MiFish, a set of universal PCR primers for metabarcoding environmental DNA from

359   fishes: Detection of more than 230 subtropical marine species. *Royal Society Open*

360   *Science*, *2*(7), 150088. doi: 10/gmcj95

361   Roberts, CD, Stewart, A., Struthers, C., Barker, J., & Kortet, S. (2019). *Checklist of the*

362   *Fishes of New Zealand* (No. July; p. 219). Museum of New Zealand Te Papa

363   Tongarewa. Retrieved from Museum of New Zealand Te Papa Tongarewa website:

364   https://collections.tepapa.govt.nz/document/10564

365   Roberts, Clive (Ed.). (2005). *Regional diversity and biogeography of coastal fishes on the*

366   *West Coast South Island of New Zealand*. Wellington, N.Z: Dept. of Conservation.

367   Sigsgaard, E. E., Torquato, F., Frøslev, T. G., Moore, A. B. M., Sørensen, J. M., Range, P.,

368   … Thomsen, P. F. (2020). Using vertebrate environmental DNA from seawater in

369   biomonitoring of marine habitats. *Conservation Biology*, *34*(3), 697–710. doi:

370   10/gmcj83

371   Stoeckle, M. Y., Das Mishu, M., & Charlop-Powers, Z. (2020). Improved Environmental

372   DNA Reference Library Detects Overlooked Marine Fishes in New Jersey, United

373   States. *Frontiers in Marine Science*, *7*, 226. doi: 10.3389/fmars.2020.00226

374   Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For*

375   *biodiversity research and monitoring* (1st ed.). New York, NY: Oxford University

376   Press.

377   Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., … Ekrem, T.

378   (2019). DNA barcode reference libraries for the monitoring of aquatic biota in

379   Europe: Gap-analysis and recommendations for future work. *Science of the Total*

380   *Environment*, *678*, 499–524. doi: 10/c6tb

381   Wing, S. R., & Jack, L. (2013). Marine reserve networks conserve biodiversity by stabilizing

382   communities and maintaining food web structure. *Ecosphere*, *4*(11), art135. doi:

383   10.1890/ES13-00257.1
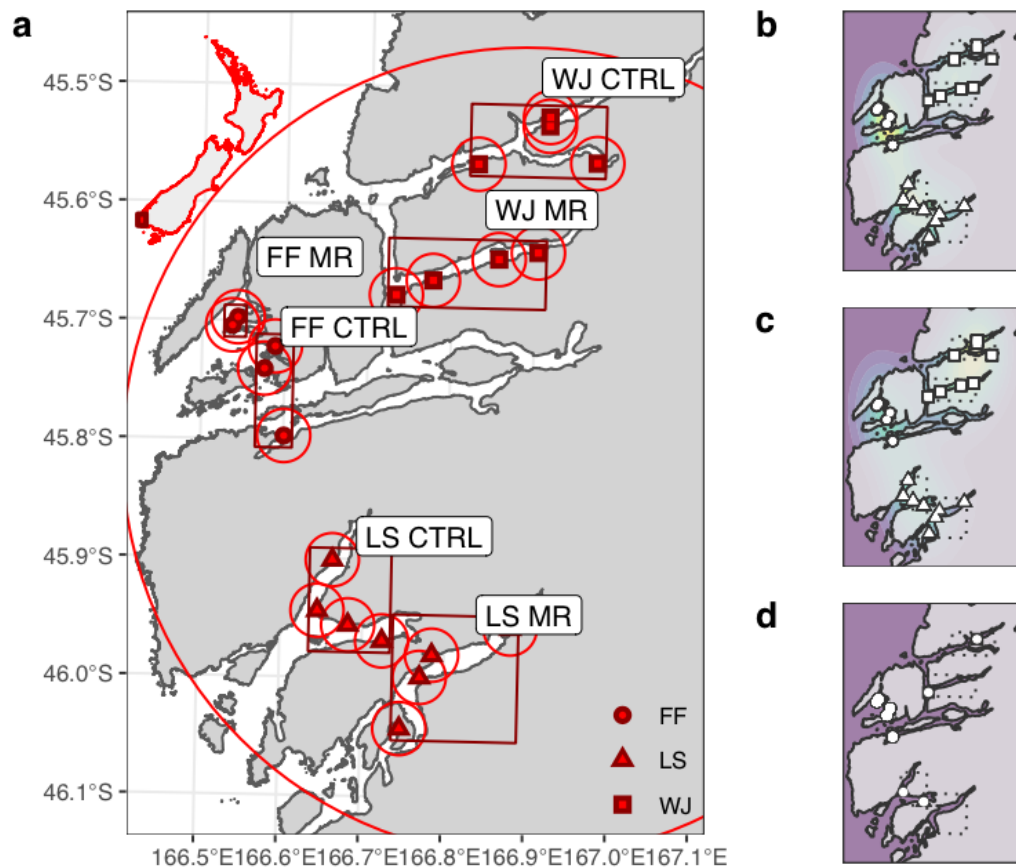
384

385 **Figures and Tables**

386

387 **Fig. 1**: Field work area, description, sites, and data coverage for eDNA, BRUV, and
388 OBIS data. **a:** We obtained biodiversity information from baited remote underwater video
389 (BRUV) footage and environmental DNA (eDNA) data from 21 field work sites across
390 three sampling regions (highlighted by rectangles) – Five Fingers (FF), Long Sound (LS)
391 and Wet Jacket (WJ). In each region we collected samples inside marine reserves /
392 commercial exclusion zones (MR) and outside in control areas (CTRL). To obtain
393 additional biodiversity information, we queried the Ocean Biodiversity Information
394 System (OBIS – https://obis.org/) for records within a 2.5 km radius of each field work
395 site (small circles) for the purpose of community structure analysis. Furthermore, we
396 obtained OBIS records for the entire sampling region (large circle) to extend our species
397 list alongside species mentioned across various literature sources (Table 1, SI Table 3). **b:**
398 Environmental DNA (eDNA), and **c:** BRUV data in a spatial context, lighter colour
399 indicates a higher density of distinct species observations (corresponding to numerical
400 values in Fig. 2). **d:** Species data for all filed work sites could not be obtained from OBIS,

401   necessitating the exclusion of this data in the statistical analyses of regional biodiversity

402   data. Graph created using R package *ggplot2 (3.3.5)*.

403
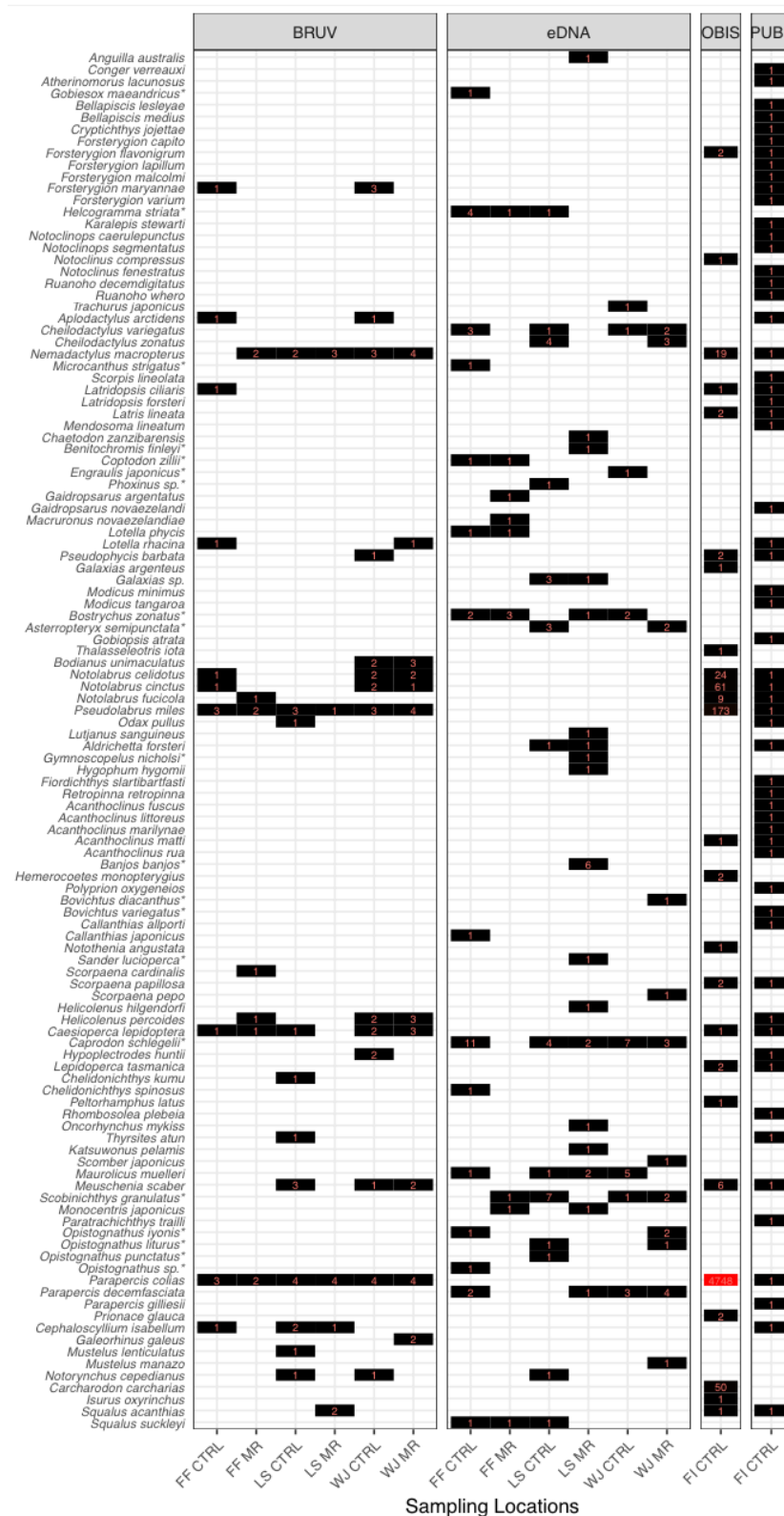
404   **Fig. 2**: Distinct species observations across data sources and field work locations.

405   Observation types: **BRUV** – Observations from baited remote underwater surveys; **eDNA**

406   – environmental DNA observations; **OBIS** – data retrieved from the Ocean Biodiversity

407   Information System (https://obis.org/) for the area surrounding field work sites (large

408   circle in Fig.1); **PUBL** – Fiordland fish species collated from multiple literature records

409   as summarized by Inglis (2008). Sampling Locations: FF – Five Fingers area; LS – Long

410   Sound area; WJ – Wet Jacket area; MR – marine reserve or commercial exclusion zone;

411   CTRL – neither marine reserve nor commercial exclusion zone. Species list: Order

412   follows Table 1, species not listed as New Zealand Species in (CD Roberts et al., 2019)

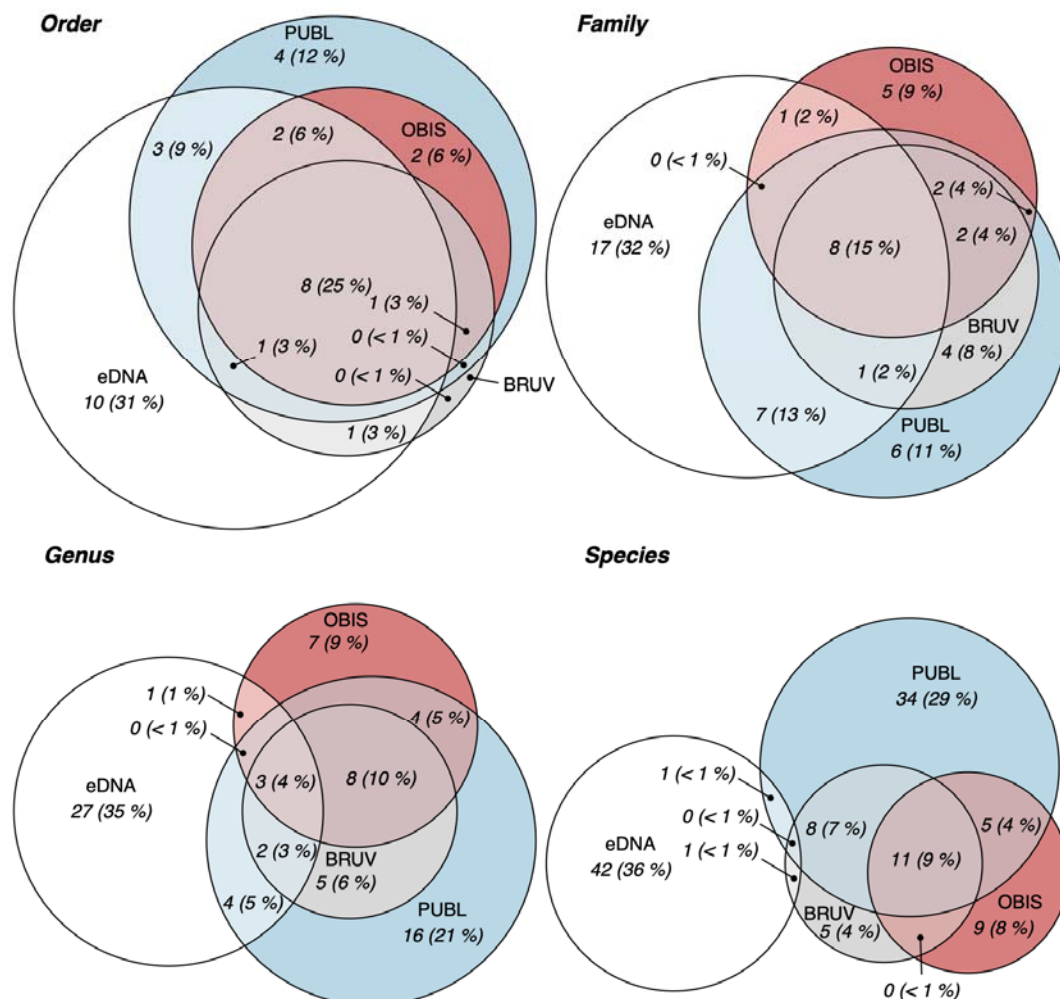413   are marked with an asterisk (*). Graph created using R package *ggplot2 (3.3.5)*.

414

415   **Fig. 3**: Concordance of taxonomic information across four data sources of Fiordland fish

416   biodiversity. Biodiversity data (Table 1, SI Table 2) is summarized at four different

417   taxonomic levels, shown are unique observation counts at each level, as well as the

418   corresponding percentage of those counts in comparison to all data. Circle sizes

419   proportional to observation count. Observation types: BRUV (grey) – Observations from

420   baited remote underwater surveys; eDNA (white) – environmental DNA observations;

421   OBIS (red) – data retrieved from the Ocean Biodiversity Information System

422   (https://obis.org/) for the area surrounding field work sites (large circle in Fig.1); PUBL

423   (blue) – Fiordland fish species collated from multiple literature records as summarized by

424   (Inglis et al., 2008). Graph created using R package *eulerr (6.1.0)*.

425 **Table 1**: Details on taxonomic observations across data sources. Taxonomic hierarchies conform with NCBI taxonomy where available, thus
426 allow analysis in relation to environmental DNA (eDNA) data and are sorted alphabetically – the resulting species order is identical to Fig. 2.
427 Taxa not listed as New Zealand species by Roberts et al., (2019) are highlighted with asterisk (*). Trivial names are indicated where
428 available from NCBI. For all taxonomic assignments also yield from eDNA we provide the alignment coverage and alignment gaps. Since
429 identical species were assigned to multiple Amplicon Sequence Variants (ASV's; Callahan et al., 2017) in some instances, ranges are
430 provided for alignment coverages and gap counts for species-specific alignments.

431

Phylum	Class	Order	Family	Genus	Species	Common name	Algn. covrg.	Algn. gaps
Chordata	Actinopteri	Anguilliformes	Anguillidae	*Anguilla*	*Anguilla australis*	Australian shortfin eel	100%	0
			Congridae	*Conger*	*Conger verreauxi*	conger eel		
		Atheriniformes	Atherinidae	*Atherinomorus*	*Atherinomorus lacunosus*	hardyhead silverside		
		Blenniiformes	Gobiesocidae	*Gobiesox**	*Gobiesox maeandricus**	northern clingfish	79.3%	2
			Tripterygiidae	*Bellapiscis*	*Bellapiscis lesleyae*	mottled twister		
					Bellapiscis medius	twister		
				Cryptichthys	*Cryptichthys jojettae*			
				Forsterygion	*Forsterygion capito*	spotted robust triplefin		
					Forsterygion flavonigrum	yellow-and-black triplefin		
					Forsterygion lapillum	common triplefin		
					Forsterygion malcolmi			

			Forsterygion maryannae			
			Forsterygion varium	striped triplefin		
		Helcogramma *	*Helcogramma striata* *		88.2-92.3%	1
		Karalepis	*Karalepis stewarti*			
		Notoclinops	*Notoclinops caerulepunctus*			
			Notoclinops segmentatus			
		Notoclinus	*Notoclinus compressus*			
			Notoclinus fenestratus			
		Ruanoho	*Ruanoho decemdigitatus*			
			Ruanoho whero	spectacled triplefin		
Carangiformes	Carangidae	*Trachurus*	*Trachurus japonicus*	Japanese jack mackerel	99.4%	0
Centrarchiformes	Aplodactylidae	*Aplodactylus*	*Aplodactylus arctidens*			

	Cheilodactylidae	*Cheilodactylus*	*Cheilodactylus variegatus*		97.6-98.8%	0
			Cheilodactylus zonatus	blackbarred morwong	97-97.6%	0
		Nemadactylus	*Nemadactylus macropterus*	tarakihi		
	Kyphosidae	*Microcanthus**	*Microcanthus strigatus**	stripey	85.3%	3
		Scorpis	*Scorpis lineolata*	silver sweep		
	Latridae	*Latridopsis*	*Latridopsis ciliaris*	blue moki		
			Latridopsis forsteri	bastard trumpeter		
		Latris	*Latris lineata*	striped trumpeter		
		Mendosoma	*Mendosoma lineatum*			
Chaetodontiformes	Chaetodontidae	*Chaetodon*	*Chaetodon zanzibarensis*		81.5%	4
Cichliformes	Cichlidae	*Benitochromis**	*Benitochromis finleyi**		79.4%	0
		Coptodon*	*Coptodon zillii**	redbelly tilapia	90.6%	3

Clupeiformes	Engraulidae	*Engraulis**	*Engraulis japonicus**	Japanese anchovy	98.8%	0
Cypriniformes	Cyprinidae	*Phoxinus**	*Phoxinus sp.**		97.1%	0
Gadiformes	Gaidropsaridae	*Gaidropsarus*	*Gaidropsarus argentatus*	Arctic rockling	90.1%	1
			Gaidropsarus novaezelandi			
	Merlucciidae	*Macruronus*	*Macruronus novaezelandiae*	blue grenadier	100%	0
	Moridae	*Lotella*	*Lotella phycis*		95.9%	0
			Lotella rhacina	rock cod		
		Pseudophycis	*Pseudophycis barbata*	southern bastard codling		
Galaxiiformes	Galaxiidae	*Galaxias*	*Galaxias argenteus*			
			Galaxias sp.		96.5%	0
Gobiesociformes	Gobiesocidae	*Modicus*	*Modicus minimus*			
			Modicus tangaroa			
Gobiiformes	Eleotridae	*Bostrychus**	*Bostrychus zonatus**	barred gudgeon	85.4-86%	5

	Gobiidae	*Asterropteryx**	*Asterropteryx semipunctata**	starry goby	81.9-82.5%	6
		Gobiopsis	*Gobiopsis atrata*			
	Thalasseleotrididae	*Thalasseleotris*	*Thalasseleotris iota*			
Labriformes	Labridae	*Bodianus*	*Bodianus unimaculatus*	red pigfish		
		Notolabrus	*Notolabrus celidotus*	New Zealand spotty		
			Notolabrus cinctus			
			Notolabrus fucicola	yellow-saddled wrasse		
		Pseudolabrus	*Pseudolabrus miles*			
	Odacidae	*Odax*	*Odax pullus*	greenbone		
Lutjaniformes	Lutjanidae	*Lutjanus*	*Lutjanus sanguineus*	humphead snapper	80.8%	6
Mugiliformes	Mugilidae	*Aldrichetta*	*Aldrichetta forsteri*	yellow-eye mullet	96-100%	0
Myctophiformes	Myctophidae	*Gymnoscopelus**	*Gymnoscopelus nicholsi**		79.8%	1

		Hygophum	*Hygophum hygomii*		78.7%	3
Ophidiiformes	Bythitidae	*Fiordichthys*	*Fiordichthys slartibartfasti*			
Osmeriformes	Retropinnidae	*Retropinna*	*Retropinna retropinna*	cucumberfish		
Ovalentaria	Plesiopidae	*Acanthoclinus*	*Acanthoclinus fuscus*			
			Acanthoclinus littoreus			
			Acanthoclinus marilynae			
			Acanthoclinus matti			
			Acanthoclinus rua			
Pempheriformes	Banjosidae	*Banjos**	*Banjos banjos**		84.7-85.2%	0
	Percophidae	*Hemerocoetes*	*Hemerocoetes monopterygius*			
	Polyprionidae	*Polyprion*	*Polyprion oxygeneios*			
Perciformes	Bovichtidae	*Bovichtus**	*Bovichtus diacanthus**		94.1%	0

		Bovichtus variegatus	thornfish		
Callanthiidae	*Callanthias*	*Callanthias allporti*			
		Callanthias japonicus		95.2%	1
Nototheniidae	*Notothenia*	*Notothenia angustata*	Maori chief		
Percidae	*Sander**	*Sander lucioperca**	pikeperch	81.4%	3
Scorpaenidae	*Scorpaena*	*Scorpaena cardinalis*	red rock cod		
		Scorpaena papillosa			
		Scorpaena pepo	pumpkin scorpionfish	85.7%	0
Sebastidae	*Helicolenus*	*Helicolenus hilgendorfi*		95.4%	0
		Helicolenus percoides			
Serranidae	*Caesioperca*	*Caesioperca lepidoptera*			
	*Caprodon**	*Caprodon schlegelii**	sunrise perch	90.6-98.2%	0-2

		Hypoplectrodes	*Hypoplectrodes huntii*			
		Lepidoperca	*Lepidoperca tasmanica*			
	Triglidae	*Chelidonichthys*	*Chelidonichthys kumu*	bluefin gurnard		
			Chelidonichthys spinosus	red gurnard	99.4%	0
Pleuronectiformes	Rhombosoleidae	*Peltorhamphus*	*Peltorhamphus latus*	speckled sole		
		Rhombosolea	*Rhombosolea plebeia*	New Zealand flounder		
Salmoniformes	Salmonidae	*Oncorhynchus*	*Oncorhynchus mykiss*	rainbow trout	100%	0
Scombriformes	Gempylidae	*Thyrsites*	*Thyrsites atun*	snoek		
	Scombridae	*Katsuwonus*	*Katsuwonus pelamis*	skipjack tuna	95.9%	0
		Scomber	*Scomber japonicus*	chub mackerel	100%	0
Stomiiformes	Sternoptychidae	*Maurolicus*	*Maurolicus muelleri*	pearlsides	86.3-99.4%	0-10
Tetraodontiformes	Monacanthidae	*Meuschenia*	*Meuschenia scaber*	velvet leatherjacket		

			Scobinichthys*	Scobinichthys granulatus*	rough leatherjacket	98.8-99.4%	0
	Trachichthyiformes	Monocentridae	Monocentris	Monocentris japonicus		94.4-97.6%	2-3
		Trachichthyidae	Paratrachichthys	Paratrachichthys trailli	sandpaper fish		
	undefined	Opistognathidae	Opistognathus*	Opistognathus iyonis*		89.9-90.5%	2-3
				Opistognathus liturus*	seto-amadai	89.3-90.5%	2
				Opistognathus punctatus*	finespotted jawfish	81.7%	5
				Opistognathus sp.*		85.7%	2
	Uranoscopiformes	Pinguipedidae	Parapercis	Parapercis colias	New Zealand blue cod		
				Parapercis decemfasciata		80.9%	1
				Parapercis gilliesii	yellow weaver		
Chondrichthyes	Carcharhiniformes	Carcharhinidae	Prionace	Prionace glauca	blue shark		
		Scyliorhinidae	Cephaloscyllium	Cephaloscyllium isabellum			

	Triakidae	*Galeorhinus*	*Galeorhinus galeus*	tope shark		
		Mustelus	*Mustelus lenticulatus*	spotted estuary smooth-hound		
			Mustelus manazo	starspotted smooth-hound	98.9%	0
Hexanchiformes	Hexanchidae	*Notorynchus*	*Notorynchus cepedianus*	broadnose sevengill shark	98.4%	0
Lamniformes	Alopiidae	*Carcharodon*	*Carcharodon carcharias*	great white shark		
		Isurus	*Isurus oxyrinchus*	shortfin mako shark		
Squaliformes	Squalidae	*Squalus*	*Squalus acanthias*	spiny dogfish		
			Squalus suckleyi	Puget Sound dogfish	100%	0

432

a

WJ CTRL

WJ MR

FF MR

FF CTRL

LS CTRL

LS MR

● FF
▲ LS
■ WJ

b

c

d

Sampling Locations

Order

PUBL
4 (12 %)

3 (9 %)

2 (6 %)

OBIS
2 (6 %)

eDNA
10 (31 %)

8 (25 %)

1 (3 %)

0 (< 1 %)

1 (3 %)

0 (< 1 %)

BRUV

1 (3 %)

Family

OBIS
5 (9 %)

1 (2 %)

0 (< 1 %)

eDNA
17 (32 %)

2 (4 %)

2 (4 %)

8 (15 %)

BRUV
4 (8 %)

1 (2 %)

7 (13 %)

PUBL
6 (11 %)

Genus

OBIS
7 (9 %)

1 (1 %)

0 (< 1 %)

eDNA
27 (35 %)

4 (5 %)

3 (4 %)

8 (10 %)

2 (3 %)

BRUV
5 (6 %)

4 (5 %)

PUBL
16 (21 %)

Species

PUBL
34 (29 %)

1 (< 1 %)

0 (< 1 %)

1 (< 1 %)

eDNA
42 (36 %)

8 (7 %)

5 (4 %)

11 (9 %)

BRUV
5 (4 %)

OBIS
9 (8 %)

0 (< 1 %)