# Evaluation of biodiversity metrics through environmental DNA metabarcoding outperforms visual and capturing surveys

Hideyuki Doi[1,*,†], Ryutei Inui[2,*,†], Shunsuke Matsuoka[1,†], Yoshihisa Akamatsu[2], Masuji Goto[2], and Takanori Kono[2]

[1] Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima Minami-machi, Chuo-ku, Kobe, 650-0047, Japan

[2] Graduate School of Science and Engineering, Yamaguchi University, 2-16-1 Tokiwadai, Ube, Yamaguchi, 755-8611, Japan

[†]These authors equally contributed to this study.

[*]Corresponding authors:

Hideyuki Doi (hideyuki.doi@icloud.com)

Ryutei Inui (inuiryutei@gmail.com)

**Abstract**

Information on alpha (local), beta (between habitats), and gamma (regional) diversity is fundamental to conserving biodiversity and the functions and stability of ecosystem processes. Robust methods like environmental DNA (eDNA) metabarcoding are currently considered useful to investigate biodiversity. However, the performance of eDNA methods in evaluating diversity has not been tested quantitatively. We compared the performance of eDNA metabarcoding and visual and capturing surveys in estimating alpha, beta, and gamma diversity in river fish communities, particularly considering community nestedness and turnover. In five rivers across west Japan, when compared with visual and capturing surveys, eDNA metabarcoding detected higher alpha and gamma diversity in local habitats, and indicated differences in beta diversity more clearly; this suggests the superiority of eDNA metabarcoding over visual/capturing surveys in estimating diversity. The statistical frameworks, particularly nestedness and turnover, can provide quantitative evidences needed to assess diversity component estimation by new survey methods.

38

39 The maintenance of biodiversity underpins the stability of ecosystem processes in
40 constantly changing environments[1,2,3]. Moreover, biodiversity loss affects ecosystem
41 functions and services and consequently human society[1,2,3]. Ecologists have made
42 efforts to conserve biodiversity based on essential biodiversity data, e.g., species
43 richness and distribution[1,2,3,4]. Biodiversity can be evaluated in different ways: viz., by
44 estimating alpha (local), beta (between habitats), and gamma (regional), diversity. To
45 conserve local communities, ecologists incorporated these diversity measurements into
46 management decision-making[1,2,5]. For example, beta diversity can quantify biodiversity
47 loss and inform the placement of protected areas and the management of biological
48 invasions and landscapes[5]. Thus, robust methods for monitoring biodiversity are
49 fundamental for biodiversity and environmental management.

50 Since recent advances in molecular techniques, environmental DNA (eDNA)
51 analysis has been considered a useful tool to investigate the distribution and richness of
52 aquatic and terrestrial organisms[6,7,8,9,10,11,12,13,14]. High-throughput sequencing (HTS)
53 derived from eDNA, called "eDNA metabarcoding", became a useful and powerful tool
54 for biodiversity surveys[13,16,17,18,19,20]. eDNA metabarcoding has recently been applied in
55 fish community surveys, e.g., Miya et al.[11] (2015) developed the universal PCR primers
56 for fish community (MiFish primers). To confirm the usefulness of eDNA
57 metabarcoding for community assessment, many studies conducted eDNA
58 metabarcoding comparing the observed species list with traditional direct surveys
59 including visual and capturing surveys[13,16,20,21,22]. However, evaluating the performance
60 of eDNA metabarcoding estimating alpha and beta diversity was still limited
61 quantitively and statistically, especially when evaluating beta diversity and comparing
62 to traditional surveys.

63 Beta diversity is a fundamental aspect for communities, and it is important for
64 evaluating community responses to environmental gradients and the spatial locations
65 across the different habitats and ecosystems[5,23,24,25]. Beta diversity itself and the trends
66 along environmental gradients such as productivity and disturbance can be affected by
67 alpha diversity[24]. Furthermore, beta diversity can decline even if alpha and gamma
68 diversity remain unchanged or even increase in the area, mainly due to biological
69 homogenization[26,27,28]. Therefore, beta diversity evaluation of communities would be
70 considerably different among the survey methods when alpha and gamma diversity
71 evaluations were different between the survey methods, i.e., eDNA metabarcoding and
72 traditional surveys.
73 Beta diversity is considered to reflect two different components: nestedness
74 and spatial turnover[29,30,31]. Nestedness of species assemblages in the communities
75 occurs when the community at the sites with less species are subsets of the community
76 at the sites with higher species richness[32,33]. Nestedness generally reflects a non-random
77 process of species loss as a consequence of any factor that promotes the orderly
78 disaggregation of assemblages[34]. Contrastingly, spatial turnover implies the replacement
79 of some species by others because of environmental sorting or spatial/historical
80 constraints[31]. Baselga[31] (2010) developed statistical separation methods to evaluate beta
81 diversity considering nestedness and spatial turnover and applied them to beta-diversity
82 evaluation in various systems[31,35]. However, the method has never been applied to
83 evaluate the performance of eDNA metabarcoding estimating beta diversity. Moreover,
84 Baselga's framework[31] (2010) can be applied to comparing the performance among
85 methods when evaluating alpha diversity via nestedness and species turnover.
86 Using statistical methods, we can quantitatively and statistically compare the
87 performance of eDNA metabarcoding and traditional surveys in community evaluation

using alpha, beta, and gamma diversity. Here, we tested the performance of eDNA metabarcoding in five river systems in different regions with various fish species. We conducted eDNA metabarcoding using the MiFish universal primer for fish and also identified the fish by visual snorkeling and hand-net capturing surveys. We evaluated the performance of eDNA metabarcoding by comparing the obtained fish community structure to that evaluated by visual/capturing survey with special regard to nestedness and spatial turnover. We finally validated the ability of eDNA metabarcoding to estimate biodiversity.

## Results

**Overview.** We detected 53 fish taxa, almost all identified to species or genus level, by eDNA metabarcoding in five rivers (Table S1, and S2) and visually observed 38 fish taxa in total. A MiSeq paired-end sequencing for the 40 libraries (30 samples plus 10 negative controls) yielded a total of 1,601,816 reads (53,351 ± 17,639; mean ± S. D. for each sample, Table S2). We confirmed very low reads from negative controls (Table S2).

**Alpha, beta, and gamma diversity between the methods.** We found significant differences in fish richness between eDNA metabarcoding and visual/capturing surveys (Fig. 1, GLMM, $t = -5.45$, $P = 0.000018$). Richness was not significantly different among river segments ($t = -5.85$, $P = 0.000004$) but not among rivers ($t = 1.737$, $P = 0.0942$), indicating higher alpha and gamma diversity estimated by eDNA metabarcoding than by visual/capturing surveys.

112 We found differences in community structures between the two methods by
113 NMDS ordination (Fig. 2). PERMANOVA results for the ordination suggested there
114 were differences in community composition evaluated by each method, eDNA
115 metabarcoding and visual/capturing survey (P = 0.003, Table S3 for full statistical
116 results). Moreover, communities from the combined results of eDNA metabarcoding
117 and visual/capturing survey were significantly different among rivers and segments (P =
118 0.001, Table S3). We found different patterns in ordinated river sites for each method
119 (Fig. 2). The PERMANOVA results determined that communities were significantly
120 different among the rivers by eDNA metabarcoding (P = 0.011) but not by
121 visual/capturing survey (P = 0.12, Table S4 for full statistical results). Conversely, the
122 communities were significantly different among river segments by visual/capturing
123 survey (P = 0.011) but marginally not significantly different by eDNA metabarcoding
124 (P = 0.061, Table S4). The differences in PERMANOVA results suggested that
125 differences in beta-diversity among rivers across regions can be detected by eDNA
126 metabarcoding but not by visual/capturing survey.
127 Indicator taxa analysis comparing the communities estimated by both
128 methods, eDNA metabarcoding and visual/capturing survey, detected significantly
129 different method preferences of several taxa, including Japanese eel (*Anguilla*
130 *japonica*), salmon (e.g., *Oncorhynchus keta)*, and Amur catfish (*Silurus asotus*),
131 between both methods ($p < 0.05$, Table 1 for statistically significant taxa, Table S5 for
132 all taxa).
133
134 **Nestedness and turnover**. We compared the pairwise standardized effect size (pSES)
135 between eDNA metabarcoding and visual/capturing survey in nestedness and species
136 turnover (Fig. 3). Nestedness pSES was significantly positive without overlapping the

137 zero-pSES 95% confidence interval, indicating that the visual/capturing survey
138 community was nested in that detected by eDNA metabarcoding. The significantly
139 negative pSES in species turnover indicated that the taxa did not significantly turnover
140 in the community evaluated by eDNA metabarcoding and visual/capturing survey
141 comparing with random communities.

142 We compared the longitudinal nested structure (upstream to downstream) of
143 the fish community in the study rivers using nested pSES (Fig. 4a). The longitudinal
144 nested structure was not significantly different between both methods (GLMM, P =
145 0.302). The positive pSES indicated that downstream community significantly nested
146 the upstream ones. The species per segment nested as downstream > mid-stream >
147 upstream in both methods (Fig. S4, 5, NOFD, P < 0.001). We also compared the
148 longitudinal spatial turnover of fish community in the study rivers using turnover pSES
149 (Fig. 4b). The longitudinal spatial turnover was not significantly different between both
150 methods (GLMM, P = 0.280).

151

152 **Discussion**

153

154 We found that river fish communities estimated by eDNA metabarcoding significantly
155 nested the communities estimated by visual/capturing survey. Furthermore, the species
156 turnover in the communities was very weak between methods. Nestedness of species
157 assemblages occurs when the communities obtained by the method estimating lower
158 number of species are subsets of the communities estimated by other methods with
159 higher species richness. Although it is difficult to assume the effort of sampling was
160 enough to evaluate the species compositions, we assumed the efforts for the visual and
161 capturing survey are enough to confirm the species composition in the study sites

162 because we combined two high-detection surveys, i.e. visual observation and capturing
163 surveys[12,13]. Therefore, the performance of eDNA metabarcoding in estimating alpha
164 and gamma diversity outperformed that of visual/capturing survey for river fish
165 community. In fact, the fish local and regional richness (alpha and gamma diversity)
166 evaluated by eDNA metabarcoding was significantly higher, including almost all taxa
167 evaluated by visual/capturing survey.

168 eDNA metabarcoding has been reported to perform better than traditional
169 methods in evaluating species richness[13,16,19,20,21,22]. Nakagawa et al.[21] (2018)
170 investigated freshwater fish community in 100 rivers and confirmed that the community
171 detected by eDNA metabarcoding was similar to the species list observed in
172 governmental-authorized monitoring. Furthermore, several eDNA metabarcoding
173 studies on fish communities have been performed in other river systems[20], marine
174 habitats[22] , and freshwater lakes[13,19]. Deiner et al.[16] (2016) showed that river eDNA
175 metabarcoding can reflect the community in a watershed, indicating that eDNA
176 metabarcoding has high-performance gamma diversity evaluation. These studies
177 indicated the great potential of eDNA metabarcoding as a useful tool for alpha and
178 gamma diversity assessment by simply comparing the community data obtained from
179 eDNA metabarcoding and traditional surveys. However, previous studies did not
180 evaluate the performance by nestedness and species turnover between eDNA
181 metabarcoding and other community data. Here, we supported previous literature by
182 showing nestedness, species turnover, and species preferences between the communities
183 detected by eDNA metabarcoding and traditional methods, and we found the nestedness
184 structure eDNA metabarcoding > visual/capturing survey, with scarce species turnover
185 in the community.

186 We especially focused on beta-diversity for assessing the performance of
187 eDNA metabarcoding compared to the visual/capturing survey. Higher beta diversity of
188 the rivers were statistically detected by eDNA metabarcoding than by visual/capturing
189 surveys. Furthermore, the spatial nestedness and turnover were not significantly
190 different between eDNA metabarcoding and traditional methods. These results
191 suggested that eDNA metabarcoding also outperformed visual/capturing survey in
192 evaluating river fish community beta diversity, probably due to the lower alpha
193 diversity detected by visual/capturing survey, especially in the upstream segments.
194 Moreover, beta diversity between segments could be detected by visual/capturing
195 surveys but not by eDNA metabarcoding. These differences in beta diversity evaluation
196 may lead us to interpret beta diversity using the results from both survey methods,
197 eDNA and traditional survey. Deiner et al.[16] (2016) suggested that eDNA sampling
198 allows an estimate of catchment-level diversity and integrates this information across
199 space due to downstream transport of eDNA. We agree with the suggestion for estimate
200 of catchment-level diversity and further suggest that the catchment-level diversity
201 metrics including alpha, beta, and gamma diversity, can be evaluated similarly to the
202 direct survey.
203 We further compared the indicator taxa for the communities obtained from
204 both eDNA metabarcoding and visual/capturing survey and concluded that several taxa,
205 including eel, salmon, and catfish, were significantly better detected by eDNA
206 metabarcoding, whereas non-indicator taxa were detected by visual/capturing surveys.
207 These results indicated that eDNA metabarcoding outperformed visual/capturing
208 surveys in fish taxa detection. The community structures estimated by eDNA
209 metabarcoding and visual/capturing survey were slightly different, as reported in
210 previous studies[13,19], probably because of the differences in taxa-detection

211 performances. Discriminated taxa in this analysis included eel, salmon, and catfish,
212 which mostly had larger body size and lower abundances in these rivers[36,37]. In fact, the
213 Japanese eel *Anguilla japonica*, was difficult to find by visual observation, probably due
214 to its hiding behavior[38]. Such endangered species would be important as top-predator[37].
215 eDNA metabarcoding can evaluate the distribution of such rare and important taxa in
216 fish communities better than traditional surveys.

217 To evaluate the comparing among the local sites and the rivers, we examined
218 the performances of eDNA metabarcoding at three sites of five rivers with eDNA
219 sampling from only two habitats. Thus, our understanding of details in the fish
220 community spatial structure in the rivers and the performance of community evaluation
221 in local habitats, such as backwater, was still limited. In fact, Bylemans et al.[20] (2018)
222 found that river morphology in the habitats influenced the optimal sampling strategy for
223 eDNA metabarcoding. Moreover, in backwater lakes, the performance of eDNA
224 metabarcoding varied with different lake morphology[13]. However, testing the usefulness
225 of eDNA metabarcoding for assessing river fish community biodiversity has been
226 limited. Further research is needed to evaluate fish community spatial structure in
227 rivers.

228 In conclusion, eDNA metabarcoding outperformed the visual/capturing
229 survey in evaluating alpha, beta, and gamma diversity of fish communities in five rivers
230 across the west Japan region. Biodiversity testing using statistical frameworks,
231 especially community nestedness and turnover, provided the quantitative evidence
232 needed to support the advantages of eDNA metabarcoding against traditional surveys.
233 eDNA methods for biodiversity assessment may provide more information to us, as
234 shown here, but we should also pay attention to the unknown characteristics of eDNA,
235 such as the origins, degradation, and transport of eDNA in the water, which are still

236 unknown in various habitats[39,40]. To routinely use eDNA methods to assess biodiversity,
237 we encourage testing the effect of eDNA degradation and transport on the performance
238 of eDNA methods for biodiversity evaluation.
239
240 **Methods**
241
242 **Study sites.** In 2016, we conducted field surveys in five river systems across Japan
243 (river map in Fig. S1): the Kyuragi River on 10th October, the Koishiwara River on 21st
244 October, the Yato River on 25th October, the Hazuki River on 2nd November, and the
245 Oze River on 6th November. The survey sites were set at three sites in each of three
246 segments (upstream, mid-stream, and downstream, the internal distances ranged from
247 4.5 to 25.8 km, Fig. S1) for each river. Each site was set so that the length in the up-
248 down direction is approximately 100 m and with a riffle at the downstream end (e. g.,
249 Fig. S2).
250
251 **Water collection for eDNA survey.** In each site, we collected 1 L of surface water in
252 bleached bottles at two points, the stream near the downstream end and the shore with
253 static or semi-static water (Fig. S2). The bottles and filtering equipment removed the
254 DNA using 10% commercial bleach (approximately 0.6% hypochlorous acid) and
255 washed it using DNA-free distilled water. One milliliter of benzalkonium chloride
256 (BAC, 10% w/v) was added per liter of water sample to avoid a decrease in eDNA
257 concentration in the samples[41]. During transport, samples were stored in a cooler with
258 ice packs. The 'cooler blank' contained 1 L DNA-free water, which we brought to the
259 field and was treated identically to the other water samples, except that it was not
260 opened at the field sites.

261

262 **Visual observation and capturing methods.** After water sampling, the fish fauna
263 survey was conducted by visual observation with snorkel and collection with hand net.
264 For visual observation, we observed and recorded fish species by snorkeling in 100-m
265 area (1 person × 1 h, Fig. S2). We observed at various micro habitats including riffle,
266 pool, and shore bank from downstream end to upstream end. We also conducted the
267 hand-net capturing survey (1 person × 1 h) using a D-flame net (net opening: 0.16 $m^2$)
268 in the various habitats in the river including riffle, pool, and shore bank. In this study,
269 Because of the impossibility of sampling some habitats such as behind stones, we did
270 not use electrofishing to observe species. Fishes were identified according to Nakabo et
271 al.[37] (2013) at the survey site. We used the combined taxa list from both surveys to
272 compare with that from eDNA metabarcoding. In order to prevent contamination of
273 DNA, the investigator who collected and identified the fish and the investigator who
274 sampled water were different.

275

276 **Water filtering and eDNA extraction from filter samples.** Bottled water samples
277 were vacuum-filtered in the laboratory into 47 mm GF/F glass filters (pore size: 0.7 μm,
278 GE Healthcare, Little Chalfont, UK) within 24 h after sampling. After filtration, all
279 filters were stored at -20 °C before eDNA extraction. The cooler blank was also
280 processed in the same manner. A liter of Milli-Q water was used as the equipment
281 control to monitor contamination during filtering in each site and during subsequent
282 DNA extraction.

283 To extract the DNA from the filters, we followed the methods described in
284 Uchii, Doi & Minamoto[42] (2016). We incubated the filter by submerging the mixed
285 buffer of 400 μL of Buffer AL (Qiagen, Hilden, Germany) and 40 μL of Proteinase K

286 (Qiagen, Hilden, Germany), using a Salivette tube (Sarstedt, Nümbrecht, Germany) at
287 56 °C for 30 min. The Salivette tube with filters was centrifuged at 5000 × g for 5 min
288 and then we added 220 µL of TE buffer (10 mM Tris-HCl and 1 mM EDTA, pH: 8.0)
289 onto the filter and again centrifuged at 5000 × g for 5 min. The DNA in the eluted
290 solution was purified using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany)
291 following the manufacturer's protocol. The final volume of the extracted sample was
292 200 µL with Buffer AE of DNeasy Blood & Tissue Kit. Samples were stored at -20 °C
293 until qPCR assay.
294
295 **Library preparation and MiSeq sequencing.** The procedures used for molecular
296 methods are described in Fujii et al.[13] (2019). A two-step PCR-procedure was used for
297 library preparation of Illumina MiSeq sequencing. As the first step, a fragment of the
298 mitochondrial 12S rRNA gene was amplified using the MiFish-U-F and MiFish-U-R
299 primers[11] (Miya et al. 2015) which were designed to contain Illumina sequencing primer
300 regions and 6-mer Ns;
301 Forward: 5′-*ACACTCTTTCCCTACACGACGCTCTTCCGATCT* NNNNNN
302 GTCGGTAAAACTCGTGCCAGC-3′,
303 Reverse: 5′-*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT* NNNNNN
304 CATAGTGGGGTATCTAATCCCAGTTTG-3′
305 The italicized and normal letters represent MiSeq sequencing primers and
306 MiFish primers, respectively, and the six random bases (N) were used to enhance
307 cluster separation on the flow cells during initial base call calibrations on MiSeq. We
308 used a KOD FX Neo polymerase (Toyobo, Osaka, Japan) for the first PCR to facilitate
309 amplifications of DNA from crude extracts. The first PCR was performed with a 12 µL
310 reaction volume containing 1× PCR Buffer for KOD FX Neo, 0.4 mM dNTP mix, 0.24

311 U KOD FX Neo polymerase, 0.3 μM of each primer, and 2 μL template (Sato et al.
312 2017). The thermal cycles of this step were as follows: initial denaturation at 94 °C for
313 2 min, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 65 °C for 30
314 s, and elongation at 68 °C for 30 s, followed by final elongation at 68 °C for 5 min. The
315 first PCRs were performed using eight replicates to mitigate false negatives (PCR
316 dropouts). Thereafter, individual first PCR replicates were pooled. Each PCR product
317 (30 μL) was purified using AMPure XP (Beckman Coulter, Brea CA, USA) and eluted
318 with 30 μL of sterilized water. The purified first PCR products were used as templates
319 for the second PCR. The Illumina sequencing adaptors and the eight bp identifier
320 indices (XXXXXXXX) were added to the subsequent PCR process using a forward and
321 reverse fusion primer:
322 Forward: 5′-*AATGATACGGCGACCACCGAGATCTACA* XXXXXXXX
323 ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
324 Reverse: 5′-*CAAGCAGAAGACGGCATACGAGAT* XXXXXXXX
325 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′
326 The italicized and normal letters represent MiSeq P5/P7 adapter and
327 sequencing primers, respectively. The 8X bases represent dual-index sequences inserted
328 to identify different samples[43] (Hamady 2008). The second PCR was conducted with 12
329 cycles of a 12 μL reaction volume containing 1× KAPA HiFi HotStart ReadyMix, 0.3
330 μM of each primer, and 1.0 μL template from the first PCR production. The thermal
331 cycle profile after an initial 3 min denaturation at 95 °C was as follows: denaturation at
332 98 °C for 20 s, annealing, and extension combined at 72 °C (shuttle PCR) for 15 s, with
333 the final extension at the same temperature for 5 min. The second PCR products were
334 pooled in equal volumes and purified using AMPure XP as in the first PCR.

335 The purified PCR products were loaded on a 2% E-Gel SizeSelect (Thermo
336 Fisher Scientific, Waltham, MA, USA) and the target size of the libraries
337 (approximately 370 bp) was collected. The library concentration and quality were
338 estimated by a Qubit dsDNA HS assay kit and a Qubit 2.0 (Thermo Fisher Scientific).
339 The amplicon libraries were sequenced by 2 × 250 bp paired-end sequencing on the
340 MiSeq platform using the MiSeq v2 Reagent Kit according to the manufacturer's
341 instructions. Note that the sequencing run contained a total of 339 libraries including 40
342 of our libraries (30 samples plus ten negative controls) and 299 libraries from other
343 research projects. The MiSeq sequencing was conducted in the Department of
344 Environmental Solution Technology, Faculty of Science and Technology, Ryukoku
345 University. Sequence data were deposited in the Sequence Read Archive of DNA Data
346 Bank of Japan (DRA, Accession number: DRA008090).
347
348 **Bioinformatic analysis for MiSeq sequencing.** The procedures used for
349 bioinformatics analysis are described in Fujii et al.[13] (2019). The processing formality
350 of the MiSeq reads was evaluated by the FASTQC program
351 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). After confirming a lack of
352 technical errors in the MiSeq sequencing, low-quality tails were trimmed from each
353 read using "DynamicTrim.pl" in the SOLEXAQA software package[44] (Cox et al. 2010)
354 with a cut-off threshold set at a Phred score of 10. The trimmed paired-end reads (reads
355 1 and 2) were then merged. The assembled 1,823,446 reads were further filtered by
356 custom Perl scripts to remove reads with either ambiguous sites (Ns) or those exhibiting
357 unusual lengths with reference to the expected size of the PCR amplicons (297 ± 25 bp).
358 The software TagCleaner[45] (Schmieder et al. 2010) was used to remove primer
359 sequences with a maximum of three-base mismatches and transform the FASTQ format

360 into FASTA. The pre-processed reads with an identical sequence (i.e., 100% sequence
361 similarity) were assembled using UCLUST[46] (Edgar 2010) and the number of identical
362 reads was added to the header line of the FASTA formatted data file. The sequences
363 represented by 10 or more identical reads were subjected to the downstream analyses.
364 Processed reads were subjected to local BLASTN searches on the previously
365 established comprehensive reference database of fish species[11] (Miya et al. 2015).
366 The top BLAST hit with a sequence identity ≥97% and the E-value threshold of $10^{-5}$
367 was applied to species detection of each sequence but the species were mostly identified
368 with ≥99% match. From the BLAST results, we identified the species using methods
369 previously described[19] (Sato et al. 2017).
370
371 **Statistical analyses**. All statistical analyses and graphics were conducted in R ver.
372 3.4.4[47] (R Core Team, 2018). All statistics were set at the significance level $\alpha = 0.05$. To
373 compare between eDNA metabarcoding and visual survey data, the taxonomic levels in
374 the species list from visual survey were adjusted to the lists from eDNA metabarcoding
375 (Table S1, S2) in reference to previous studies using MiFish primer[13,19] (Sato et al.
376 2017, Fujii et al. 2019). Before statistical analysis, we confirmed that the sequencing
377 depth was sufficient to detect alpha diversity in the samples by "iNEXT" and
378 "ggiNEXT" functions in "iNEXT" package (Fig. S3). We merged the community data
379 from two points, the stream near the downstream end and the shore, to compare with
380 visual/capturing surveys.
381 We tested the differences in fish richness of sites, segments, and rivers
382 between both methods by general linear mixed model (GLMM) with "lmer" function in
383 "lme4" package. In the GLMM models, the method was treated as a fixed effect, and
384 the rivers and segments were treated as the random effects.

385 The differences in community compositions were visualized using nonmetric
386 multidimensional scaling (NMDS) with 500 separate runs of real data. For NMDS, the
387 community dissimilarity was calculated based on incidence-based Jaccard indices. We
388 evaluated the differences in community structures between methods and sites using
389 Permutational multivariate analysis of variance (PERMANOVA). For PERMANOVA,
390 we used Jaccard similarity matrix and calculated the statistical values with 999
391 permutations. We used "metaMDS" and "adonis" functions in vegan package for
392 NMDS ordination and PERMANOVA, respectively.

393 For the communities evaluated by both methods, indicator taxa analysis[48]
394 (Cáceres & Legendre, 2009) was performed to determine which taxa had significantly
395 different frequencies between both methods. The analysis was performed using the
396 "signassoc" function in the "indicspecies" package on the present/absence data. We
397 used mode = 1 (group-based) and calculated the p-values with 999 permutations after
398 Sidak's correction of the multiple testing.

399 To compare community composition (i.e., beta diversity) focusing on nested
400 and turnover structures between eDNA metabarcoding and visual/capturing survey, we
401 calculated the pairwise indices for nestedness and turnover[31,49] (Baselga 2010; Baselga
402 2012) using the "beta.pair" function in the "betapart" package. Standardized effective
403 size (SES) of nestedness and turnover indices were calculated to show the degree of
404 nestedness and turnover structure. The significance was defined by deviation from zero,
405 the expectation of random assembly (a null model) estimated with 999 random
406 sampling replicates. The SES was defined as: $(\beta_{obs} - \beta_{null}) / \beta_{sd}$, where $\beta_{obs}$ is the
407 observed beta diversity, $\beta_{null}$ is the mean of the null distribution of beta diversity, and $\beta_{sd}$
408 is the standard deviation of the null distribution. SES values greater than zero indicate
409 statistically stronger nestedness or turnover structure than expected under a random

410 model of community assembly, while negative values indicate weaker nestedness or
411 turnover than expected. The randomized community data were generated with
412 independent swap algorithm[50] (Gotelli 2000) using “randomizeMatrix" function in
413 “picante” package. Then, to evaluate the fish community longitudinal nestedness and
414 turnover structure along with river flow (i.e., upstream to downstream) by each method,
415 a nestedness metric[51] (NODF, Almeida-Neto et al., 2008) and pairwise indices were
416 calculated. First, to test whether the communities show significant longitudinal
417 nestedness, SES value of NODF and their SES were calculated with 999 randomizations
418 using “oecosimu” function in “vegan” package. Then, the SES of pairwise nestedness
419 and turnover were calculated for each sample pair within the river. Finally, we tested
420 the differences in SES of pairwise indices between the survey methods (eDNA
421 metabarcoding vs. visual/capturing survey) by general linear mixed model (GLMM)
422 with “lmer” function in “lme4” package. In the GLMM models, the SES of pairwise
423 indices was treated as a fixed effect, and the rivers-pairs and segment-pairs (i.e., three
424 pairs in each river) were treated as the random effects.
425 For the map graphics in Supplemental Materials, we used the stream and coast line data
426 from the National Land Information Division, Ministry of Land, Infrastructure,
427 Transport and Tourism of Japan (http://nlftp.mlit.go.jp/ksj/).
428
429 **Data availability**
430 All data of the MiSeq sequencing was shared in DRA (Accession number:
431 DRA008090), and the observed species data was shared in Table S2 in Supplemental
432 Materials.
433
434 **References**

435 1. Primack, R. B. (1993). Essentials of conservation biology (Vol. 23). Sunderland,
436 Massachusetts: Sinauer Associates.
437 2. Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature*, **405**, 243
438 (2000).
439 3. Pecl, G. T., et al. Biodiversity redistribution under climate change: Impacts on
440 ecosystems and human well-being. *Science*, **355**, 9214 (2017).
441 4. Doi, H. & Takahara, T. Global patterns of conservation research importance in
442 different countries of the world. *PeerJ*, **4**, e2173 (2016).
443 5. Socolar, J. B., Gilroy, J. J., Kunin, W. E. & Edwards, D. P. How should beta-
444 diversity inform biodiversity conservation? *Trends Ecol. Evol.*, **31**(1), 67-80 (2016).
445 6. Takahara, T., Minamoto, T. & Doi, H. Using environmental DNA to estimate the
446 distribution of an invasive fish species in ponds. *PLoS ONE*, **8**, e56584 (2013).
447 doi: 10.1371/journal.pone.0056584.
448 7. Takahara, T., Minamoto, T. & Doi, H. Using environmental DNA to estimate the
449 distribution of an invasive fish species in ponds. *PLoS ONE,* **8**, e56584 (2013).
450 8. Goldberg, C. S., Strickler, K. M. & Pilliod, D. S. Moving environmental DNA
451 methods from concept to practice for monitoring aquatic macroorganisms. *Biol*
452 *Conserv.*, **183**, 1-3 (2015). doi: 10.1016/j.biocon.2014.11.040.
453 9. Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. & Gough, K. C.
454 Review: the detection of aquatic animal species using environmental DNA — a
455 review of eDNA as a survey tool in ecology. *J. Appl. Ecol.*, **51**, 1450-1459
456 (2014). doi: 10.1111/1365-2664.12306.
457 10. Thomsen, P. F., Willerslev, E. Environmental DNA–an emerging tool in
458 conservation for monitoring past and present biodiversity. *Biol Conserv.*, **183**, 4-
459 18 (2015). doi: 10.1016/j.biocon.2014.11.019.

460 11. Miya, M., et al. MiFish, a set of universal PCR primers for metabarcoding
461 environmental DNA from fishes: detection of more than 230 subtropical marine
462 species. *R. Soc. Open Sci.*, **2**, 150088 (2015).
463 12. Doi, H., et al. Environmental DNA analysis for estimating the abundance and
464 biomass of stream fish. *Freshw. Biol.*, **6**, 30-39 (2017).
465 13. Fujii, K., Doi, H., Matsuoka, S., Nagano, S., Sato, H. & Yamanaka, H.
466 Environmental DNA metabarcoding for fish community analysis in backwater
467 lakes: A comparison of capture methods. *PLoS ONE*, **14**, e0210357 (2019).
468 14. Doi, H., et al. Evaluation of detection probabilities at the water-filtering and initial
469 PCR steps in environmental DNA metabarcoding using a multispecies site
470 occupancy model. *Sci. Rep.*, **9**, 3581(2019).
471 15. Bista, I., et al. Annual time-series analysis of aqueous eDNA reveals ecologically
472 relevant dynamics of lake ecosystem biodiversity. *Nat. Commun.*, **8**, 14087
473 (2017). doi: 10.1038/ncomms14087.
474 16.Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J. C. & Altermatt, F.
475 Environmental DNA reveals that rivers are conveyer belts of biodiversity
476 information. *Nat. Commun.*, **7**, 12544 (2016).
477 17.Deiner K, et al. Environmental DNA metabarcoding: Transforming how we survey
478 animal and plant communities. *Mol. Ecol.*, **26**, 5872-5895 (2017). doi:
479 10.1111/mec.14350.
480 18.Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards
481 next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.*,
482 **21**(8), 2045-2050 (2012).

483 19.Sato, H., Sogo, Y., Doi, H. & Yamanaka, H. Usefulness and limitations of sample
484 pooling for environmental DNA metabarcoding of freshwater fish communities. *Sci.*
485 *Rep.*, **7**, 14860 (2017). doi: 10.1038/s41598-017-14978-6.
486 20. Bylemans, J., et al. Monitoring riverine fish communities through eDNA
487 metabarcoding: determining optimal sampling strategies along an altitudinal and
488 biodiversity gradient. *Metabarcoding Metagenom.*, **2**, e30457 (2018).
489 21. Nakagawa, H., et al. Comparing local- and regional-scale estimations of the
490 diversity of stream fish using eDNA metabarcoding and conventional
491 observation methods. *Freshw. Biol.*, **63**, 569-580 (2018).
492 22. Yamamoto, S., et al. Environmental DNA metabarcoding reveals local fish
493 communities in a species-rich coastal sea. *Sci. Rep.*, **7,** 40368 (2017).
494 23. Chase, J. M. & Ryberg, W. A. Connectivity, scale-dependence, and the
495 productivity–diversity relationship. *Ecol. Lett.*, **7**(8), 676-683 (2004).
496 24. Chase, J.M. Stochastic community assembly causes higher biodiversity in more
497 productive environments. *Science*, **328**, 1388–1391 (2010).
498 25. McCune, J. L. & Vellend, M. Gains in native species promote biotic
499 homogenization over four decades in a human-dominated landscape. *J. Ecol.*,
500 **101**, 1542-1551 (2013).
501 26. Olden, J. D. & Poff, N. L. Toward a mechanistic understanding and pre-diction of
502 biotic homogenization. *Am. Nat.*, **162**, 442–460 (2003).
503 27. Van Calster, H., et al. Management driven changes (1967–2005) in soil acidity and
504 the understorey plant community following conversion of a coppic-with-
505 standards forest. *For. Ecol. Manage.*, **241**, 258–271 (2007).

506 28. Keith, S. A., Newton, A. C., Morecroft, M. D., Bealey, C. E. & Bullock, J. M.
507 Taxonomic homogenization of woodland plant communities over70 years. *Proc.*
508 *R. Soc. B. Biol. Sci.*, **276**, 3539–3544 (2009).
509 29. Harrison, S. et al. Beta-diversity on geographic gradients in Britain. *J. Anim. Ecol.*,
510 **61**, 151-158 (1992).
511 30. Baselga, A., Jimenez-Valverde, A. & Niccolini, G. A multiple-site similarity
512 measure independent of richness. *Biol. Lett.*, **3**, 642–645 (2007).
513 31. Baselga, A. Multiplicative partition of true diversity yields independent alpha and
514 beta components; additive partition does not. *Ecology*, **91**(7), 1974-1981 (2010).
515 32. Wright, D. H. & Reeves, J. H. On the meaning and measurement of nestedness of
516 species assemblages. *Oecologia*, **92**(3), 416-428 (1992).
517 33. Ulrich, W. & Gotelli, N. J. Null model analysis of species nestedness patterns.
518 *Ecology*, **88**(7), 1824-1831 (2007).
519 34. Gaston, K. J. & T. M. Blackburn. Pattern and process in macroecology. Blackwell
520 Scientific, Oxford (2000).
521 35. Baselga, A., Gómez-Rodríguez, C. & Lobo, J. M. Historical legacies in world
522 amphibian diversity revealed by the turnover and nestedness components of beta
523 diversity. *PloS ONE*, **7**(2), e32341 (2012).
524 36. Kawanabe, H., Mizuno, N. & K. Hosoya (eds.) Freshwater Fishes of Japan Yama-
525 Kei, Tokyo (2001) (in Japanese).
526 37. Nakabo, T. Fishes of Japan with pictorial keys to the species, 3rd ed. Tokai
527 University Press (2013) (in Japanese).
528 38. Itakura H, et al. Environmental DNA analysis reveals the spatial distribution,
529 abundance, and biomass of Japanese eels at the river-basin scale. *Aquat.*
530 *Conserv.*, 1–13 (2019).

531 39. Barnes, M. A. & Turner, C. R. The ecology of environmental DNA and implications
532 for conservation genetics. *Conserv. Genet.*, **17**(1), 1–17 (2016).
533 40. Seymour, M., et al. Acidity promotes degradation of multi-species environmental
534 DNA in lotic mesocosms. *Commun. Biol.*, **1**(1), 4 (2018).
535 41. Yamanaka, H., et al. On-site filtration of water samples for environmental DNA
536 analysis to avoid DNA degradation during transportation. *Ecol. Res.,* **31**(6), 963-
537 967 (2016).
538 42. Uchii, K., Doi, H. & Minamoto, T. A novel environmental DNA approach to
539 quantify the cryptic invasion of non-native genotypes. *Mol. Ecol. Res*, **16**, 415-422
540 (2016). doi: 10.1111/1755-0998.12460.
541 43.Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J. & Knight, R. Error-correcting
542 barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat.*
543 *Methods*, **5**, 235-237 (2008).
544 44.Cox, M. P., Peterson, D. A., Biggs, P. J. SolexaQA: at-a-glance quality assessment
545 of Illumina second-generation sequencing data. BMC *Bioinform*., **11**, 485 (2010).
546 doi: 10.1186/1471-2105-11-485.
547 45.Schmieder, R., Lim, Y. W., Rohwer, F. & Edwards, R. TagCleaner: Identification
548 and removal of tag sequences from genomic and metagenomic datasets. *BMC*
549 *Bioinform*., **11**, 341 (2010). doi: 10.1186/1471-2105-11-341.
550 46. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
551 *Bioinformatics*, **26**, 2460-2461 (2010). doi: 10.1093/bioinformatics/btq461.
552 47. R Development Core Team. R: A language and environment for statistical
553 computing. R Foundation for Statistical Computing, Vienna, Austria (2018).
554 https://www.R-project.org/.

48. Cáceres, M. D. & Legendre, P. Associations between species and groups of sites: indices and statistical inference. *Ecology*, **90**(12), 3566-3574 (2009).

49. Baselga, A. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Glob. Ecol. Biogeogr.*, **21**, 1223-1232 (2012).

50. Gotelli, N. J. Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606-2621 (2000).

51. Almeida-Neto, M., Guimaraes, P., Guimaraes, P. R., Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, **117**, 1227-1239 (2008).

**Acknowledgement**

MiSeq sequencing was conducted in the Department of Environmental Solution Technology, Faculty of Science and Technology, Ryukoku University, and we thank to H. Yamanaka and H. Sato for supporting the experiments involving MiSeq sequencing. This study was supported by the Environment Research and Technology Development Fund (4-1602) of Environmental Restoration and Conservation Agency, Japan and JST-CREST (JPMJCR13A2).

**Author contributions**. HD, RI, and YA designed the study, RI, MG, TK, and YA contributed to field survey and sampling. SM, HD, RI, and MG contributed to molecular experiments. SM and HD analyzed the data and interpreted the results. HD, SM, and RI wrote the initial draft of the manuscript. All other authors critically reviewed the manuscript.

**Competing financial interests.** The authors declare no competing financial or non-financial interests.

**Materials & Correspondence.** H. D. (hideyuki.doi@icloud.com) and R. I. (inuiryutei@gmail.com)

585
586 **Figure legends**

587

588 Figure 1. Venn diagrams for the number of detected taxa. eDNA metabarcoding
589 (eDNA, blue) and visual/capturing methods (V/C, red) in each site of the five study
590 river systems. The numbers mean the number of taxa.

591

592 Figure 2. Nonmetric multidimensional scaling (NMDS) ordination. Fish communities
593 evaluated by eDNA metabarcoding (blue) and visual/capturing methods (red) in each
594 site of the river. MDS stress was 0.158.

595

596 Figure 3. Pairwise effect size (SES, standard effect size). Comparison between
597 community nestedness and spatial turnover between the communities detected by
598 eDNA metabarcoding and visual/capturing methods. Error bars indicate 95%
599 confidence interval. The horizontal dotted line represents SES = 0, indicating non-
600 significant effect.

601

602 Figure 4. Effect size (SES, standard effect size). a) community nestedness and b) spatial
603 turnover SES among the segment communities evaluated by eDNA metabarcoding and
604 visual/capturing methods. The error bars indicate 95% confidence interval. The
605 horizontal dotted line represents SES = 0, indicating non-significant effect.

606

607

608

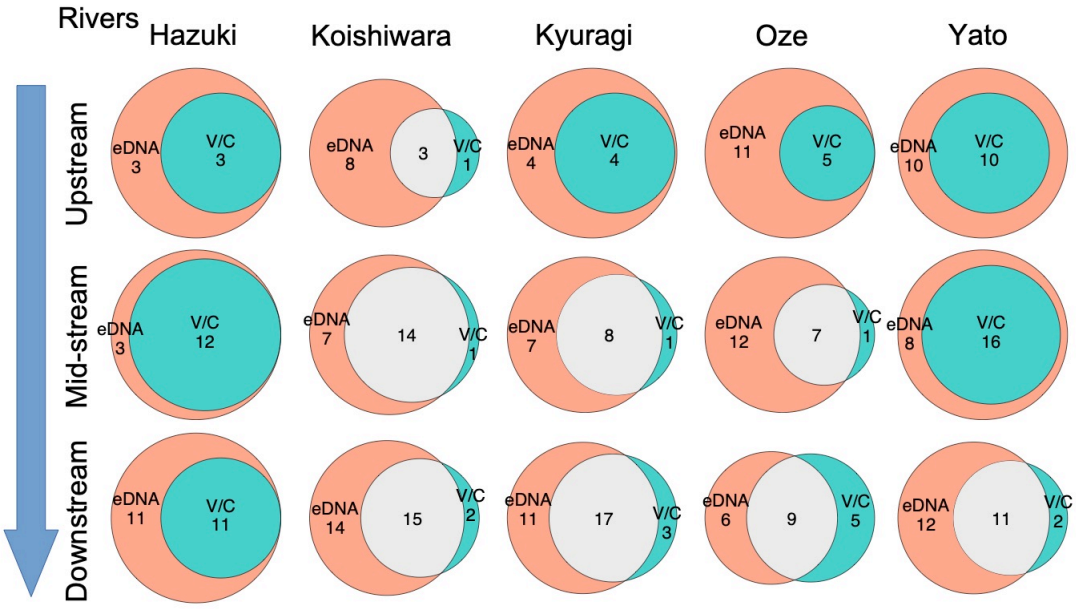

Rivers
Hazuki
Koishiwara
Kyuragi
Oze
Yato
Upstream
Mid-stream
Downstream
eDNA 3
V/C 3
eDNA 8
3
V/C 1
eDNA 4
V/C 4
eDNA 11
V/C 5
eDNA 10
V/C 10
eDNA 3
V/C 12
eDNA 7
14
V/C 1
eDNA 7
8
V/C 1
eDNA 12
7
V/C 1
eDNA 8
V/C 16
eDNA 11
V/C 11
eDNA 14
15
V/C 2
eDNA 11
17
V/C 3
eDNA 6
9
V/C 5
eDNA 12
11
V/C 2

Figure 1

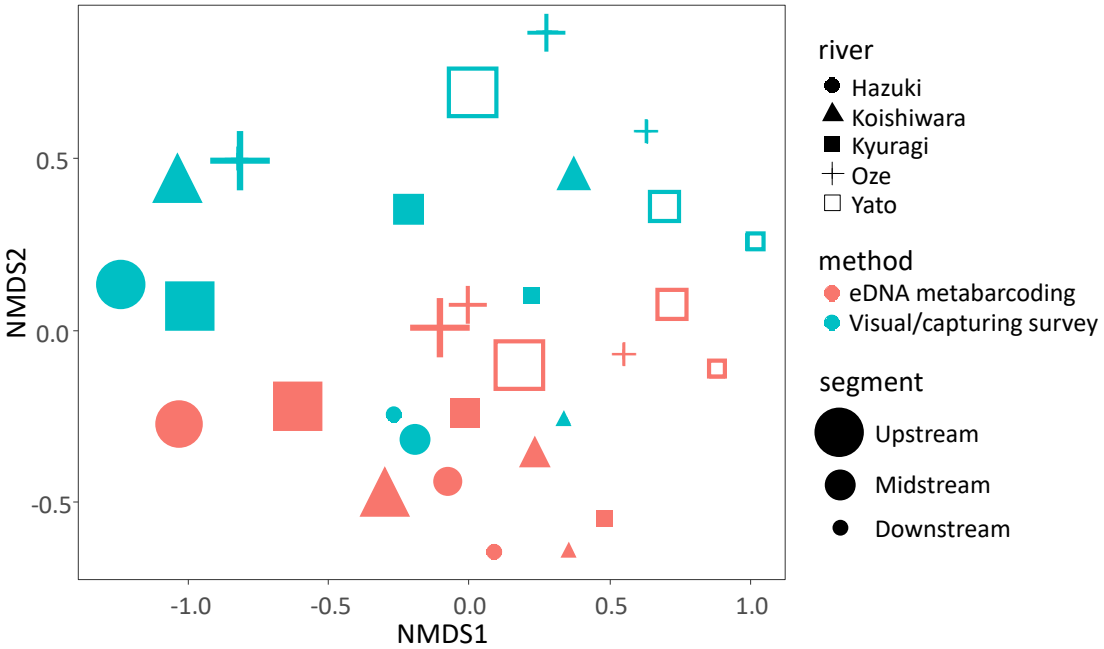

NMDS2
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
1.0
NMDS1
river
Hazuki
Koishiwara
Kyuragi
Oze
Yato
method
eDNA metabarcoding
Visual/capturing survey
segment
Upstream
Midstream
Downstream

Figure 2

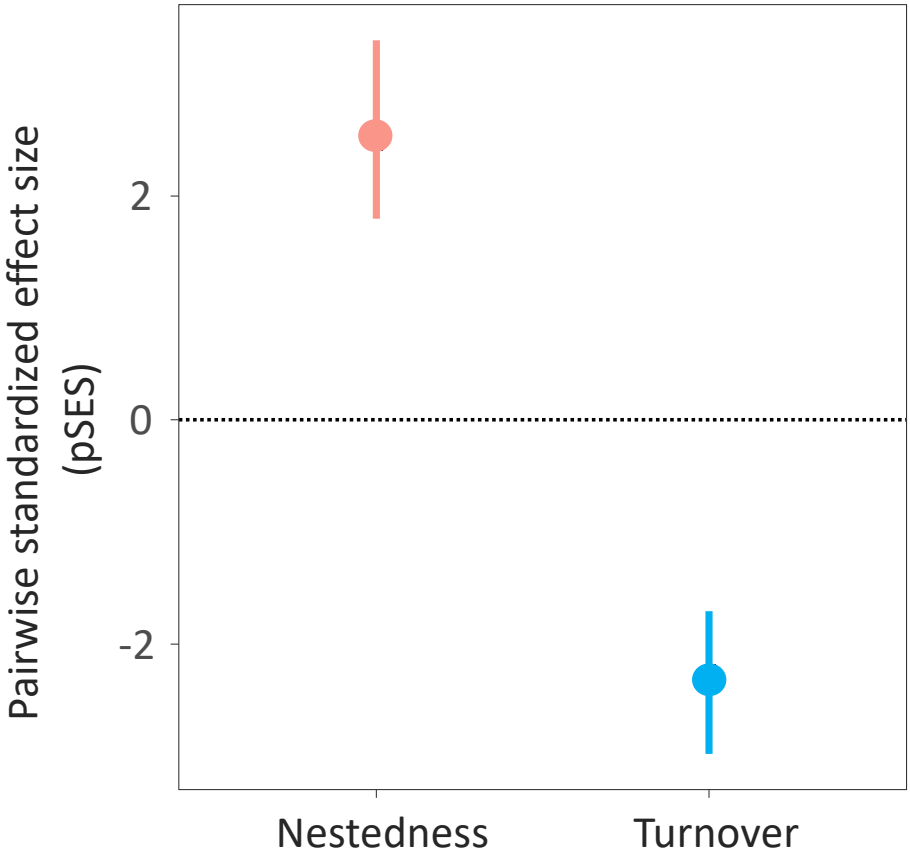
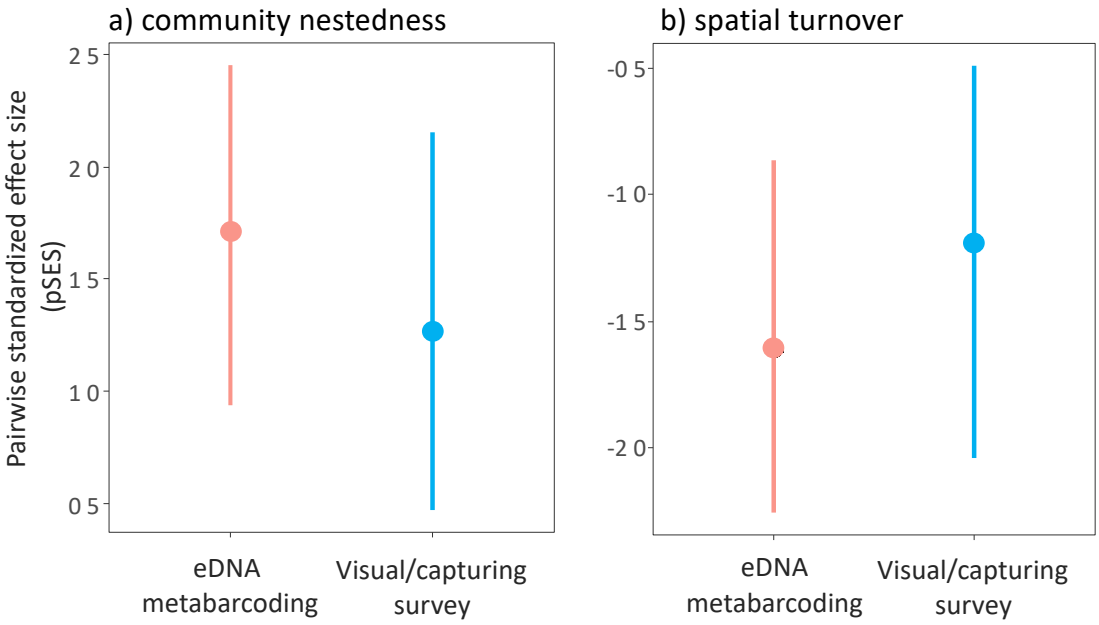

Pairwise standardized effect size
(pSES)
2
0
-2
Nestedness
Turnover

Figure 3

619




620

621 Figure 4

622