1 **Article Type:** Original Article

2 **Implications of taxonomic and numerical resolution on DNA metabarcoding-based**

3 **inference of benthic macroinvertebrate responses to river restoration**

4 Running Title: Taxonomic and numerical resolution on metabarcoding

5 **Authors and Affiliations**

6 Joeselle M. Serrana[1,2], Bin Li[1], Tetsuya Sumi[3], Yasuhiro Takemon[3], and Kozo Watanabe[1,2]

7 [1]Center for Marine Environmental Studies, Ehime University, Bunkyo-cho 3, Matsuyama, Ehime 790-

8   8577, Japan

9 [2]Graduate School of Science and Engineering, Ehime University, Bunkyo-cho 3, Matsuyama, Ehime

10   790-8577, Japan

11 [3]Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan

12 **Correspondence**

13 Prof. Kozo Watanabe, Ph.D.; Molecular Ecology and Health (MEcoH) Laboratory, Engineering Building

14 No. 2, Ehime University, Bunkyo-cho 3, Matsuyama, Ehime 790-8577, Japan; Email:

15 watanabe.kozo.mj@ehime-u.ac.jp

16 **Abstract**

17 Exploring and clearly defining the level of taxonomic identification and quantification approaches for

18 diversity and biomonitoring studies are essential, given its potential influence on the assessment and

19 interpretation of ecological outcomes. In this study, we evaluated the response of benthic

20 macroinvertebrate communities to the restoration or construction of gravel bars conducted in the dam-

21 impacted Trinity River, with the non-dam influenced tributaries serving as the reference sites. We aim to

22 evaluate the performance of different taxonomic and numerical (i.e., abundance vs. presence/absence

23 data) resolutions of DNA metabarcoding with consequent comparison to morphology-based

24 identification and how it affects assessment outcomes. DNA metabarcoding detected 93% of the

25 morphologically identified individuals and provided finer taxonomic resolution. We also detected

26 significant correlations between morphological sample abundance, biomass, and DNA metabarcoding

27 read abundance. We observed a relatively high and significant congruence in macroinvertebrate

28 community structure and composition between different taxonomic and numerical resolutions of both

29 methods, indicating a satisfactory surrogacy between the two approaches and their varying identification

30 levels and data transformation. Additionally, the community-environmental association were significant

31 for all datasets but showed varying significant associations against the physicochemical parameters.

32 Furthermore, both methods identified *Simulium* spp. as significant indicators of the dam-impacted gravel

33 bars. Still, only DNA metabarcoding showed significant false discovery rate proving the method's

34 robustness compared to morphology-based identification. Our observations imply that coarser

35 taxonomic resolution could be highly advantageous to DNA metabarcoding-based applications in

36 situations where the lack of taxonomic information, e.g., poor reference database, might severely affect

37 the quality of biological assessments.

38 **Keywords**

INTRODUCTION

Benthic macroinvertebrates are the most commonly used focal groups for the ecological survey of freshwater ecosystems since they serve as valuable indicators of ecosystem health due to their high diversity and different sensitivity to a range of natural and anthropogenic disturbances (Menezes et al., 2010), which have been used to develop biotic indices for extensive monitoring programs (Bush et al., 2019). The accuracy of identification is a key point for macroinvertebrate community-based biomonitoring and assessment. Baird and Hajibabaei (2012) pointed out several constraints that have severely limited the utility of morphology-based macroinvertebrate metrics in broad-scale biomonitoring programs over the last fifty years, e.g., the extensive time required for field sample processing, the identification difficulties for finer taxonomic resolution, the potential identification bias amongst experts, and the general lack of verification of morphology-based identifications.

DNA metabarcoding is a transformative approach to biomonitoring, biodiversity discovery, and ecosystem health assessments in freshwater ecosystems (Cordier et al. 2017; Elbrecht et al. 2017; Serrana et al. 2019), providing solutions to various morphology-based constraints and has been tested and used to study different freshwater habitats and taxonomic groups (e.g., Serrana et al., 2018; Bailet et al., 2019; Ficetola et al., 2020). However, although the method has been reported to identify taxa with finer taxonomic resolution, some argue that this advantage is not necessarily valuable when taxa cannot be linked to a binomial taxonomic name (Bush et al., 2019), which may emerge from incomplete reference DNA libraries (Curry et al., 2018), and mixed template PCR-based challenges, e.g., inadequacy of high performance primers to amplify all relevant taxa (Taberlet et al., 2018).

Taxonomic resolution has substantial effects on ecological study outcomes (Cordier et al., 2017; Pawlowski et al., 2018), but taxonomic details are often set without explicit justification and are usually based on the subjective criteria of sample-processing costs or times (Jones, 2008). Increased taxonomic resolution may increase information and observed variation among communities. This implies that depending on the specificity of the community response to environmental stress, finer taxonomic resolution may either enable better or worse detection of the environmental stress (Bailey, Norris & Reynoldson, 2001). Exploring and clearly defining the community representatives' taxonomic resolution

68   from diversity and biomonitoring studies is important for understanding the trade-offs associated with

69   different taxonomic levels (Bailey, Norris & Reynoldson, 2001) that is vital for improving the comparability

70   between biogeographically separate programs (Bailet et al., 2019). In particular, the difference in

71   ecological surveys and their employed levels of taxonomic resolution creates potential discrepancies in

72   the results and in the conclusions drawn when comparing the performance of separate programs and

73   procedures (Herbst & Silldorff, 2006). Taxonomic sufficiency – the pragmatic concept of balancing the

74   trade-off between the level of identification against feasibility in ecological studies (Ferrero & Cole, 1992)

75   needs revisiting and discussion, given that species-level identification, which is considered the gold

76   standard in environmental biomonitoring and assessment is not always achievable (Pawlowski et al.,

77   2018), even for DNA metabarcoding-based identifications (Staats et al., 2016; Hleap et al., 2021).

78   Although previous studies have assessed the influence of taxonomic resolution on DNA metabarcoding-

79   based assessments (e.g., Laini et al., 2020), the evaluation of different taxonomic level for identifying

80   taxa without losing significant ecological detail is still largely unexplored.

81   Furthermore, the numerical resolution of data, i.e., qualitative - presence/absence vs.

82   quantitative – relative or absolute abundance, may influence the outcome of ecological analyses

83   (Mueller, Pander & Geist, 2013; Heino, 2014; Pires et al., 2021). Previous DNA metabarcoding-based

84   macroinvertebrate biomonitoring studies had contrasting reports on numerical resolution. Some suggest

85   the use of presence/absence data (Buchner et al., 2019; Zizka, Geiger & Leese, 2020) since it leads to

86   similar assessment results compared to abundance-based data, while some presented that the

87   quantification of relative species abundance based on read depth information provides better

88   assessment efficiency instead of presence/absence data (Aylagas et al., 2018; Serrana et al., 2019;

89   Meyer et al., 2020).

90   Sediment deficit downstream reaches due to sediment loading in reservoirs (Petts & Gurnell,

91   2005) leads to reduced habitat complexity and decreased biodiversity in downstream ecosystems (Graf,

92   2006). As a resolve, some restoration programs on dam-fragmented rivers employ the construction of

93   gravel bars by gravel augmentation and channel rehabilitation (Ock et al., 2015). Gravel bars provide

94   areas of increased biogeochemical activities due to the enforced hydrodynamic exchange (Sackett et

95   al., 2019), retaining organic matter filtered from surface waters in the hyporheic zone (i.e., the interface

between surface and groundwater) for the utilization of river biota, and enhances nutrient cycling with consequent benefits to ecosystem metabolism (Mendoza-Lera & Datry, 2017). Additionally, the wet and dry processes in gravel bars due to water level changes alternately offers terrestrial and aquatic habitats, increasing environmental heterogeneity to provide diverse habitats for different organisms, e.g., fishes (Beechie et al., 2005), microorganisms (Boano et al., 2014), and macroinvertebrates (Merz et al., 2005). Macroinvertebrates have been used as indicators of restoration or enhancement of habitat complexity from gravel bar structures (Lepori et al., 2005; Dunbar et al., 2010) and gravel-bed rivers (Rice & Greenwood, 2001). However, knowledge on the influence of gravel bars on benthic macroinvertebrate communities with specific consideration of the river (i.e., dam fragmented vs. non-dam fragmented river) and reach (gravel bars vs. non-gravel bars) scales are less explored and needs further assessment.

In this study, we aim to evaluate the performance of different taxonomic and numerical resolutions in DNA metabarcoding in comparison to the traditional morphology-based identification and how this would affect the inference of benthic macroinvertebrate responses to river restoration, specifically the monitoring and assessment outcomes on the influence of gravel bar construction or rehabilitation in the dam-impacted Trinity River in California. We hypothesize that different taxonomic levels (i.e., family, genus, species, or haplotype level) and numerical resolution (i.e., absolute abundance or presence/absence data) from both morphology-based and DNA metabarcoding identifications affect the outcome of ecological assessments, i.e., multivariate community pattern, diversity measures, and environmental associations at different spatial scales (i.e., river and reach). Particularly, we expect that the increased information from genus, species, or haplotype-level data relative to a coarser taxonomic level, i.e., family, provides better discrimination in community composition and structure to differentiate test and reference sites at the river (i.e., dam fragmented vs. non-dam fragmented river) and reach (i.e., gravel bars vs. non-gravel bars) scales. We also performed indicator taxa analysis on the datasets of both methods to identify specific genera or species that can be used as indicators of restoration in the gravel bars of the dam-impacted Trinity River in comparison to the reference sites.

121   **MATERIALS AND METHODS**

122   *Study design and sampling*

123   The Trinity River is a large gravel-bed river located in northwest California, impounded by the Trinity

124   Dam (164 m a.b.l and 3,020 million m^3 storage) and the smaller Lewiston Dam (28 m a.b.l and 18 million

125   m^3 storage) located downstream of the Trinity Dam since 1964. The river is under current dam operating

126   guidelines with a mean annual flood of approximately 180 m^3/s (Gaeuman et al., 2017). The Trinity River

127   Restoration Program, a multi-agency partnership, manages and implements these releases alongside

128   gravel augmentations and mechanical rehabilitations downstream of the Lewiston Dam with the aims of

129   restoring salmonid habitat and dynamic channel processes in the river (USDOI, 2000 as cited in Ock et

130   al., 2015).

131   The field survey was conducted in the Trinity River along a 60-km river length downstream of

132   the Lewiston Dam in August 2017 to assess the influence of dam-impoundment and gravel bars on

133   benthic macroinvertebrate communities. We performed a paired sampling scheme with two spatial

134   hierarchical scales: i.e., river scale [2 groups: dam-influenced (Trinity River, DAM+) vs. non-dam

135   influenced (tributaries, DAM-)], and reach scale [(2 groups: with gravel bar (GB+) vs. non-gravel bar

136   (GB-)], and their combinations river × reach (4 groups: DAM+GB+, DAM+GB-, DAM-GB+, and DAM-

137   GB-). Six gravel bars and two non-gravel bar reaches of the Trinity River were selected as test sites,

138   while two gravel bars and two non-gravel bar reaches from two non-dam impounded tributaries, i.e., the

139   Rush Creek and the Canyon Creek were sampled as reference sites (**Figure 1**). We sampled each

140   gravel bar site at the bar head (down-welling zone) and bar tail (up-welling zone), while the non-gravel

141   bar segments were sampled at the up- and downstream points approximately 15-m in distance (relative

142   to the average length of the gravel bars). In total, we have 24 sampling points in this study.

143   Physico-chemical characteristics were measured from all sampling points (**Table S1**). Water

144   samples for water quality analysis were collected following the procedures instructed by the United

145   States Environmental Protection Agency (USEPA), kept in an icebox, and sent to the PHYSIS

146   Environmental Laboratories, Inc. (Anaheim, CA, USA) for water quality analysis, i.e., ammonium

147   nitrogen (NH_4-N), nitrate nitrogen (NO_3-N), ash-free dry mass (AFDM), and total suspended solids (TSS).

6

148   pH and electric conductivity (EC) was measured via the LAQUAtwin COND (Model Y071L, Horiba) and

149   pH meters (Model Y017, Horiba). Water temperature and dissolved oxygen were measured with a DO

150   meter (Model OM-51, Horiba, Ltd., Kyoto, Japan).

151   Benthic macroinvertebrates were collected using a Surber net (30 cm × 30 cm, mesh size 500

152   μm) at three random locations to cover most habitats (e.g., riffles and pools) at each sampling point. The

153   Surber replicates were pooled together and directly preserved in 99% ethanol in the field. In the

154   laboratory, the collected macroinvertebrates were sorted from debris, e.g., substrate and non-target

155   organic matters, and morphologically identified at the lowest taxonomic level when possible (usually

156   genus or family) under a stereomicroscope using the taxonomic keys for the identification of aquatic

157   insects in North America (Merritt, Cummins & Berg, 2008). Sorted and morphologically identified

158   samples were then kept in new ethanol and stored at 4°C before processing for molecular analysis.

159   *DNA extraction, amplicon library construction, and sequencing*

160   The morphologically identified and sorted samples from each point were dried overnight at room

161   temperature in sterile Petri dishes to remove the remaining ethanol. The dry biomass of each taxon was

162   measured on a Sartorius RC 210D semi-microbalance. Individuals from each taxon were then pooled

163   as a bulk sample and homogenized by grinding in several 2-mL tubes. Genomic DNA was extracted

164   using the DNeasy Blood & Tissue Kit (Qiagen, Inc.) following manufacturer instructions. DNA

165   concentrations were quantified with the QuantiFluor dsDNA system (Promega, Madison, WI, USA) on

166   the Quantus Fluorometer (Promega, Madison, WI, USA) and normalized to 50 ng/μL DNA for PCR.

167   A two-step PCR protocol following the procedures of Elbrecht and Steinke (2019) was applied

168   for PCR amplification and tagging. For the first PCR step, we amplified the target fragment using the

169   BF2+BR2 primer set designed explicitly for benthic macroinvertebrates and evaluated using both mock

170   and kick-net samples (Elbrecht & Leese, 2017). The PCR master mix consists of 0.25 μl Phusion, 0.75

171   μl Dimethyl sulfoxide (DMSO), 1 μl dNTPs, 1.25 μl each of the forward and reverse primers (10 μM),

172   five μl HF Buffer (New England Biolabs), and 15.5 μl of PCR-grade water. PCR cycling conditions were

173   30 seconds of initial denaturation at 98°C, followed by 25 cycles of 10 seconds denaturation at 98°C, 30

174   seconds annealing at 55°C, 30 seconds extension at 72 °C, and a final extension step of 5 minutes at

175 72°C. For the second PCR, we used fusion primers, which include inline tags and Illumina sequencing

176 tails. One µl of the first PCR was used as the template. PCR master mix and conditions are the same

177 as the first PCR step, but the cycles were adjusted to 15×. The second PCR products were purified by

178 QIAquick PCR Purification Kit (QIAGEN, Inc., Valencia, CA), and the DNA concentration was quantified

179 with qPCR via the KAPPA Illumina Library qPCR Quantification kit (Kappa Biosystems, Wilmington, MA,

180 USA) and qualified with the High-Sensitivity DNA chip (Agilent BioAnalyzer, Palo Alto, CA, USA) for

181 amplicon library assessment. PCR water was used as a negative control to monitor contamination from

182 DNA extraction and library construction to post-amplification library quantity and quality verification, and

183 no quantifiable amplicon was detected for further analysis. The 72 amplicon libraries of the 24 samples

184 with triplicates were then normalized to 4 nM and pooled. Finally, 600 µl of a 6 pM denatured pooled

185 library with PhiX (final concentration 15%; Illumina, Inc.) was prepared, and a 300-bp paired-end

186 sequencing was performed using the MiSeq Reagent Kit v3 (Illumina, Inc.).

187 The generated raw sequence data were deposited into the National Center for Biotechnology

188 Information (NCBI) Sequence Read Archive (SRA) under the accession number SRR12620160.

189 *Bioinformatics and data processing*

190 The raw Illumina paired-end reads were demultiplexed according to sample tags via the R package

191 JAMP v.0.67 (http://github.com/VascoElbrecht/JAMP) (Elbrecht et al., 2018) and were quality-checked

192 with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The primer sequences were

193 trimmed using Cutadapt v.2.1 (Martin, 2011), and the subsequent read processing steps, i.e., quality

194 filtering, merging, and inference of amplicon sequence variants (ASVs) were performed following the

195 denoising pipeline of the DADA2 v.1.12 package (Callahan et al., 2016) in R v.3.6.2 (R Core Team,

196 2019). The forward and reverse reads were trimmed at a minimum length of 220 and quality filtered

197 using a maximum expected error (-maxee) of 3 and 5. The remaining sequences were denoised, and

198 the forward and reverse sequences were merged. ASVs were inferred from the sequence data while

199 subsequently removing chimeric sequences and singletons.

200 For the taxonomic assignment, the ASV sequences were matched to the Barcode of Life

201 Database (BOLD, Barcode of Life Data System, http://www.boldsystems.org/; accessed at September

8

202   1, 2020) using the python program BOLDigger (https://github.com/DominikBuchner/BOLDigger)

203   (Buchner & Leese, 2020). The best-fitting hit option from the JAMP pipeline was performed, wherein

204   different thresholds, i.e., 98% for species level, 95% for the genus, 90% for family, 85% for order level,

205   and <85% for class level, were used as a best-fitting hit parameter. An interactive Krona chart was used

206   to visualize the total individual count and absolute read abundances within the complex hierarchies of

207   the taxonomic assignments (Ondov et al., 2011).

208   *Statistical analysis*

209   Pearson's pairwise correlation test was performed to assess the relationship between morphologically-

210   identified sample abundance and biomass per taxa against the number of reads in the metabarcoding

211   data. The shared and unique taxonomic assignment and ASVs at the river × reach scale was visualized

212   with UpSetR plots (Lex et al., 2014). The read counts per ASV and sample data, along with the

213   taxonomic identifications and sample descriptors were merged into phyloseq objects using the phyloseq

214   v.1.32.0 package (McMurdie and Holmes 2011). For the subsequent analysis, the DNA metabarcoding

215   reads and the morphologically-identified individual counts in each sample were normalized using median

216   depth. Median normalization was performed as recommended in Pereira et al. (2018) since it provides

217   a robust alternative to total count that is less affected by highly abundant samples. Four datasets with

218   different taxonomic resolutions were created from the DNA metabarcoding dataset, i.e., family, genus,

219   species, and ASV-level, while there were the family and genus-level for the morphological dataset.

220   Absolute abundance and presence/absence datasets were also created from these four DNA

221   metabarcoding and two morphological datasets.

222   Alpha diversity metrics, i.e., Chao1 richness, Shannon diversity, Pielou's J evenness, Berger-

223   Parker dominance, and the rare abundance index were calculated and visualized to identify the changes

224   in community structure between each scale using the *plot_alpha_diversities* function

225   (microbiomeutilities; Shetty & Lahti, 2018). Statistical differences of the alpha diversity metrics among

226   each scale were tested using ANOVA and pairwise comparisons via multiple t-tests. Pearson correlation

227   coefficients between the alpha diversity metrics of each taxonomic level and methods were calculated

228   using the *rcorr* function in the Hmisc package. Additionally, a meta-regression analysis was performed

229 to explore the simultaneous effects of all seven environmental variables (log-transformed) with the

230 richness estimate (i.e., Chao1) as the dependent variable using the *lm* function of the stats package.

231 The *plot_models* function of the sjPlot package was then used to plot and compare regression

232 coefficients with confidence intervals of multiple regression models in one plot.

233 To observe the spatial differences between the communities, beta diversity was estimated

234 based on Bray-Curtis distances and assayed by non-metric multidimensional scaling (NMDS) using the

235 *plot_ordination* function from the phyloseq package. Permutational multivariate analysis of variance

236 (PERMANOVA) (vegan; Oksanen et al., 2013) was performed among the scales to test the statistical

237 significance of the between-group distances based on the NMDS ordination via the *adonis2* function in

238 the vegan package. Procrustes tests were employed via the *protest* function with 9,999 permutations to

239 test if different taxonomic or numeric resolutions provided similar community structure. A canonical

240 correspondence analysis (CCA) was performed to visualize and determine the environmental variables

241 responsible for explaining community composition variation amongst sampling points with significance.

242 Before the analysis, the seven environmental variables were evaluated for multi-collinearity via variance

243 inflation factors (VIF) for constraining parameters (VIF <5) using the usdm package (Naimi 2015).

244 Additionally, a homogeneity of multivariate dispersion (PERMDISP) analysis was employed to assess if

245 differences in heterogeneity in environmental parameters existed among the samples based on the river,

246 reach, and river × reach scales.

247 Indicator taxa analysis was performed to the genus-level morphology and metabarcoding

248 datasets to identify and compare the indicator taxa between the two methods based on the river × reach

249 category using the *multipatt* function implemented in the indicspecies package in R with 9,999

250 permutations (De Caceres et al., 2016). The associations were further assessed using a false discovery

251 rate estimation by adjusting the *p*-values for multiple testing using the Benjamini-Hochberg procedure

252 (*p.adjust* function in R) (Strimmer, 2008). Furthermore, the linear discriminant analysis effect size

253 (LEfSe) test was conducted using the Galaxy implementation of LEfSe (http://huttenhower.org/galaxy)

254 (Segata et al., 2011) [parameters: $p < 0.05$, $q < 0.05$, LDA > 2.0, multi-class analysis strategy set to one-

255 against-all (less strict)] to identify which indicator taxa significantly explained differences in community

256   composition between the river × reach category. All the data visualization and statistical analyses were

257   performed in R v.3.6.2 (R Core Team, 2019).

258   **RESULTS**

259   *Taxonomic identification based on morphology and DNA metabarcoding*

260   We collected a total of 4,053 macroinvertebrates and morphologically identified 39 multi-level taxa from

261   2 phyla, 3 classes, 9 orders, 31 families, 25 genera, and 2 species. Almost all samples had family level

262   assignments (31 families), except for the individuals identified as order Haplotaxida (Oligochaeta). Only

263   68% (2,737 individuals) were assigned at the genus level, from which 99 and 37 individuals had species-

264   level assignments, i.e., *Calineuria californica* and *Pteronarcys californica*. A total of 7,399,885 paired-

265   end reads were demultiplexed from the raw sequence data. After read processing, i.e., quality filtering,

266   denoising, and paired-end merging, 4,098,620 non-chimeric reads (~421-bp) were retained and

267   assigned to 1,210 ASVs. From this, 4,083,001 reads from 1,112 ASVs were assigned to

268   macroinvertebrate taxa. ASVs assigned as bacteria, fungi, and plants (17 ASVs; 11,703 reads) and

269   ASVs without taxonomic identifications (79 ASVs; 3,916 reads) were discarded from the dataset. Out of

270   the total taxonomically assigned reads, 91% (3,701,486 reads) had a family level assignment, while 87%

271   (3,561,696 reads) and 65% (2,640,658) had genera and species level assignment. The sample, read

272   abundance, and ASV count were summarized in detail in **Table 1**. Interactive Krona charts were

273   provided as an additional file to visualize the total individual count and absolute read abundances of the

274   taxonomic assignments from the two methods.

275   From the morphology-based identification, the most abundant order was Diptera (35%), followed

276   by the EPT orders: Ephemeroptera (33%), Plecoptera (17%), and Trichoptera (11%). However, DNA

277   metabarcoding detected Plecoptera (39%) as the order with the most reads, followed by Ephemeroptera

278   (33%), Trichoptera (11%), and Diptera (8%). See **Figure S1** for the relative abundance of order-level

279   assignment of the two identification methods per sampling site. DNA metabarcoding detected 93%

280   (3,767 individuals) of the morphologically identified taxa, accounting for 83% of the reads (**Figure 2a** to

281   **2c**). Only 7% (286 individuals) were false negatives, representing three genera, i.e., *Mataeopsephus*

282 (Coleoptera), *Nemoura* (Plecoptera), and *Rhitrogena* (Ephemeroptera). The reads related to these

283 genera were possibly assigned to coarser taxonomic levels, given that some ASVs have unclassified

284 class, order, and family identifications. The remaining 17% (696,709 reads) consists of 44 taxa, i.e., 13

285 species, 12 genera, and 18 unclassified ranks unique to the DNA metabarcoding dataset which are

286 considered false positive detections. We found significant positive linear correlations between sample

287 abundance, sample biomass, and read abundance among the 36 taxa commonly found in the

288 morphological and DNA metabarcoding datasets presented in **Figure 2d** to **2f**.

289 *Richness and diversity measures*

290 The shared and unique taxa at the river × reach category for the different taxonomic resolution datasets

291 of both methods were presented in **Figure 3**. Thirteen morphologically identified families and fourteen

292 DNA metabarcoding families were shared by all the river × reach groups, i.e., DAM+GB+, DAM+GB-,

293 DAM-GB+, and DAM-GB-. Both methods identified fourteen genera present in all groups, while 20

294 species were shared between the groups from the DNA metabarcoding dataset. The pairwise correlation

295 of the alpha diversity estimates of the morphological and DNA metabarcoding datasets at different

296 taxonomic resolutions was presented in **Figure 4**. The pairwise correlations between the datasets with

297 the different taxonomic and numerical resolutions (absolute abundance or presence/absence) were

298 significantly positive in 47 pairs (taxonomic: 20; numerical: 27) for Chao1 richness, 42 pairs (taxonomic:

299 18; numerical: 24) for Shannon diversity, 18 pairs (taxonomic: 12; numerical: 6) for Berger-Parker

300 dominance index, and 25 pairs (taxonomic: 11; numerical: 14) for rare taxa abundance index out of 66

301 pairs. We observed varying statistical differences of the alpha diversity metrics among the scales (i.e.,

302 river, reach, or river × reach) at different taxonomic and numerical resolutions presented in **Table S1**.

303 Remarkably, we observed no significant difference in all the datasets for the mean alpha diversity values

304 at the reach scale. Another notable observation was that only the family and species-level absolute

305 abundance DNA metabarcoding datasets showed a significant difference between the river × reach

306 scale.

307   *Community structure and composition, and environmental relationship*

308   To assess whether different taxonomic and numerical resolution interferes in the homogeneity of

309   multivariate dispersion within each scale, the community structure and composition of the identified

310   benthic macroinvertebrates were assessed with NMDS based on Bray-Curtis distance. Ordination plots

311   were relatively similar for all the taxonomic levels, and numerical resolution on both methods, wherein

312   the communities of the samples from the dam-impacted sites (from the Trinity River) with and without

313   gravel bars were closely clustered in the ordination space, and the non-dam impacted samples clustered

314   together (**Figure 4**). The visual similarity between the NMDS ordinations between both methods'

315   different taxonomic and numerical datasets was further supported by the Procrustes analyses, which

316   revealed significant positive correlations (**Figure S2**). However, relatively few comparisons had high

317   correlation coefficient values (from 0.90 to 0.98).

318   The PERMANOVA analyses revealed that community structure and composition at the river

319   scale, dam-influenced rivers, significantly differed against the non-dam influenced rivers (PERMANOVA,

320   p = 0.001) for all the different taxonomic and numerical resolution datasets of both methods (**Table 2**).

321   Almost all datasets showed a significant difference between the communities at the reach scale

322   (PERMANOVA, $p < 0.05$), except for the ASV-level datasets and the family-level morphology-based

323   presence/absence dataset. These results indicate the strong influence of dams and gravel bars on the

324   benthic macroinvertebrates' community structure. However, no significant difference between the

325   community composition at the river × reach scale was found.

326   Environmental and water quality parameters were assessed for all twenty-four sampling sites

327   (**Table S2**). **Figure S3a** to **S3c** presents the distribution of each environmental variable per scale.

328   However, a PERMDISP2 analysis (**Table S3**) only revealed two of the seven environmental variables

329   significantly different amongst the scales, i.e., electric conductivity (EC) significantly different for both

330   river and river × reach, and total suspended solids (TSS) significant different for reach and river × reach.

331   The variation in the correlations of the different datasets in relation to the seven environmental variables

332   is also shown in **Figures S3d** and **S3e**. The CCA analysis models to test the influence of environmental

333   variables on community composition were significant for all the different taxonomic and numerical

334   datasets of both methods, with R^2 values ranging from 0.32 to 0.46. **Figure S4** presents the CCA biplots.

335   However, the different datasets showed a difference in significant associations against the seven

336   physicochemical parameters tested (**Table 3**). For example, pH significantly influenced the benthic

337   macroinvertebrate community composition for the metabarcoding family (presence/absence only),

338   species, and ASV-level datasets. EC was significant for the morphology-based family and

339   metabarcoding family and genus-level datasets. Notably, total suspended solids (TSS) were only

340   significant at the morphology-based family-level absolute abundance dataset.

341   *Indicator taxa analysis*

342   We compared the indicator genera identified between the morphology-based and DNA metabarcoding

343   methods based on the genus-level datasets. Three genera, i.e., *Rhyacophila*, *Simulium,* and *Nemoura*

344   were identified as indicators with a *p*-value significant at 0.05 from the morphology-based dataset, while

345   thirteen were identified from the DNA metabarcoding dataset, with *Wormaldia*, *Hydropsyche*, *Antocha*,

346   *Attenella*, *Haemaphylis*, *Ordobrevia,* and *Simulium* having significant *p*-values at 0.05, and false

347   discovery rate adjusted *p*-values less than 0.10. See **Table S4** for the group associations of the

348   aforementioned indicator taxa at the river × reach scale. Only *Simulium* (Diptera) was reported for both

349   methods, notably an indicator of the dam-impacted Trinity River's gravel bar sites (DAM+GB+).

350   Moreover, an additional LEfSe analysis was performed at the multi-taxa dataset (i.e., all ASVs

351   with varying taxonomic assignment) of both methods. Thirteen and 59 significantly discriminative

352   features (i.e., taxa) with an LDA score of > 2.0 out of the 13 and 105 before internal Wilcoxon were

353   selected from the morphology-based and DNA metabarcoding datasets. See Supplementary **Figures**

354   **S5** and **S6** for the taxa identified that significantly explained differences in community compositions

355   between the river × reach scales. Notably, LEfSe analyses identified genus *Simulium* as an indicator of

356   the dam-impacted Trinity River's gravel bar sites (DAM+GB+) for both methods. However, DNA

357   metabarcoding provided species-level assignments and identified *S. vittatum* alongside *Dicosmoecus*

358   *gilvipes, Hesperoperla pacifica,* and *Sialis californica* as indicator species of the dam-impacted rivers

359   with gravel bars (DAM+GB+).

DISCUSSION

The main objective of this study is to evaluate the influence of the construction or rehabilitation of gravel bars in the dam-impacted Trinity River at the river (i.e., dam fragmented vs. non-dam fragmented river) and reach (i.e., gravel bars vs. non-gravel bars) scales by performing DNA metabarcoding and morphology-based identifications of benthic macroinvertebrate communities of different taxonomic levels, i.e., family, genus, species, and haplotype (ASV-level), and numerical resolutions, i.e., absolute abundance or presence/absence data.

Congruence between DNA metabarcoding and morphology-based identifications

We confirm the effectiveness of DNA metabarcoding as an alternative method to overcome certain limitations of the conventional morphology-based identification. DNA metabarcoding detected most of the morphologically identified taxa and provided finer taxonomic resolution, up to species and haplotype-level information. This advantage was previously demonstrated in macroinvertebrate taxa studies (e.g., Baird & Hajibabaei, 2012; Elbrecht et al., 2017; Emilson et al., 2017; Serrana et al., 2019). The false positive detection between the two methods could be due to the lack of reference sequences and misclassifications in the database or primer bias effects (Carew et al., 2017; Laini et al., 2020). Regardless, we report a significant correlation between morphological abundance and sequence read abundance, between morphological abundance and biomass, and between sequence read abundance and biomass. These results were consistent with previous DNA metabarcoding studies, which reported a positive correlation between species abundance and biomass (Elbrecht & Leese, 2015) and between species abundance and relative or absolute sequence read abundance (Elbrecht et al., 2017). Our results validate and reinforce the use of sequence abundance information for diversity analyses and interpretation of our DNA metabarcoding data to be used with greater confidence.

Different taxonomic and numerical resolutions on diversity, community structure and composition

In contrast to our hypothesis that different taxonomic levels and numerical resolution from both methods lead to varying multivariate community patterns, we report similar spatial patterns of community structure

385 and composition based on the outcomes of ordination techniques (NMDS ordinations and Procrustes

386 tests). Moreover, high Procrustes coefficients were found for comparisons between different numerical

387 resolutions rather than the different taxonomic levels. This contrasts with a previous report by Pires et

388 al. (2021), where comparisons between different taxonomic resolutions have higher coefficients than

389 numerical resolution from invertebrate communities. Still, we detected relatively high and significant

390 congruence between the ordinations performed for all the taxonomic and numerical resolutions of both

391 methods and indicate a satisfactory surrogacy between the two methods and their varying identification

392 levels. Our results are similar to previous studies that reported strong congruence between different

393 taxonomic levels from freshwater macroinvertebrate samples (e.g., Brito et al., 2018; Godoy et al., 2019;

394 Caradima, Reichert & Schuwirth, 2020). Similar community distribution patterns may be due to parallel

395 responses to environmental gradients and similar dispersal abilities among taxonomically or

396 evolutionary close taxa, leading to a strong community congruence at different taxonomic resolutions

397 (Heino & Soininen, 2007; Landeiro et al., 2012).

398 We observed significant differences in the community structure and composition at the river

399 scale between the dam-impacted Trinity River and its pristine tributaries. This observation corroborates

400 with previous studies that comprehensively studied and reported the negative impacts of reservoir and

401 dam construction on macroinvertebrate diversity and community composition (e.g., Monaghan et al.,

402 2005; Serrana et al., 2018; Wang et al., 2020). Interestingly, although all datasets have significant

403 positive correlations based on Procrustes tests, the percentage of the variance explained (R^2) by the

404 river scale groups at the family level was higher than at finer taxonomic resolutions (i.e., genus, species,

405 and ASV-levels) for both absolute abundance and presence/absence data. This implies that even if

406 different taxonomic and numerical resolutions are correlated and commonly showed significant

407 differences at the river scale, family-level datasets provided relatively higher variance explained by the

408 dam-impact, and the percentage of variance explained decreased as the taxonomic resolution became

409 finer, i.e., order from family, to genus, to species, and to the ASV-levels.

410 Likewise, Godoy et al. (2019) reported that family-level identification reduced the unexplained

411 portion of variance compared with genus-level identification of aquatic insect community samples.

412 Family-level identifications have been proposed as surrogates in freshwater macroinvertebrate

413   assemblages for ecological monitoring and conservation management (Chessman, Williams & Besley,

414   2007; Brito et al., 2018). Here, we reported that DNA metabarcoding and morphology-based family-level

415   data of both numerical resolutions are enough to reveal the damming impacts on the differences of the

416   macroinvertebrate community structure and composition.

417   Furthermore, significant differences in the community structure and composition were also

418   reported for most of the morphology and DNA metabarcoding datasets of varying taxonomic and

419   numerical resolutions at the reach scale, except for both the ASV levels and the morphology-based

420   family-level (presence/absence) datasets. The non-significant difference from the ASV dataset for both

421   absolute abundance and presence/absence composition indicates that the increased taxonomic

422   resolution, i.e., intraspecific diversity (haplotypes), was not robust enough for community-based

423   analyses and resulted in the insignificant difference of macroinvertebrate community composition at the

424   reach scale. Moreover, the observed non-significant difference on the morphology-based family level

425   presence/absence dataset can be accounted for by the turnover of rare families at the reach scale. The

426   analysis based on presence/absence tends to increase the influence on the spatial distributions of rare

427   taxa in the dataset (Bailey, Norris & Reynoldson, 2001). However, for this dataset, 25 out of the 31

428   identified families were shared by the gravel bars and non-gravel bar reaches, with only three unique

429   families on each category. Among-reach scale differences in community composition were not

430   significantly observed because most of the families were distributed across these reaches.

431   For each of the alpha diversity metrics, we found significant and strong correlations between

432   most of the numerical resolution datasets of the Chao1 richness and Shannon diversity index, with a

433   majority of correlations exceeding the 0.7 threshold of congruence suggested by Heino (2010). This

434   suggests a relatively congruent estimation of these alpha diversity metrics at varying taxonomic levels,

435   regardless of numerical resolution. On the other hand, the Berger-Parker dominance and rare taxa

436   abundance indices showed mostly non-significant, even negative correlation between most of the DNA

437   metabarcoding and morphology-based datasets because the former method introduces more taxa

438   undetected from the latter approach, specifically rare taxa, thus, the different patterns in dominance and

439   rarity indices (Gibson et al., 2015).

17

440   *Community and environmental relationships at different taxonomic and numerical resolutions*

441   We found the significant influence of environmental variables on community composition for all the

442   different taxonomic and numerical resolution datasets of both methods, but each dataset showed varying

443   sets of associated physicochemical parameters. In particular, the DNA metabarcoding genus, species,

444   and ASV-level datasets showed non-significant associations with some of the environmental variables

445   that were significant for the morphology-based family and genus-level datasets. Some environmental

446   parameters, e.g., electric conductivity, dissolved oxygen and ash-free dry mass, influenced both family-

447   level datasets, while pH only showed significant associations with the DNA metabarcoding species and

448   ASV-level datasets. Finer resolution in taxonomic information introduces more unique or rare taxa which

449   could be attributed to more sensitive response to certain environmental gradients, while coarser

450   taxonomic resolutions detect a lower number of rare taxa (Emilson et al., 2017) which may lead to biased,

451   and a more limited range of environmental factors detected for their influence on the macroinvertebrate

452   communities.

453   Previous DNA metabarcoding-based macroinvertebrate biomonitoring studies documented

454   stronger discriminatory power due to the finer taxonomic resolution of the method (e.g., Gibson et al.,

455   2015). However, our environmental association analyses presented coarser taxonomic resolution of

456   DNA metabarcoding was more congruent with the morphology-based datasets. Our observation

457   corresponds with former studies. For example, Emilson et al. (2017) reported that increased statistical

458   power from finer resolution to detect the influence of environmental gradients could be study-specific.

459   Bailey, Norris, and Reynoldson's (2001) reported that additional information from finer taxonomic

460   resolutions may introduce undesirable ecological noise (i.e., a more comprehensive description of the

461   benthic macroinvertebrate community) unless the impacts of environmental disturbances are more

462   evident with finer-level identification than coarser taxonomic resolution. Other studies also reported that

463   using coarser taxonomic resolution showed small or inconspicuous loss of information compared to

464   species-level identifications for the estimation of the relationships between communities and

465   environmental conditions (e.g., Heino & Soininen, 2007; de Oliveira et al., 2020), with some even

466   reporting lower efficiency of finer taxonomic resolution for the environmental associations due to the

467   ecological noise (e.g., Heino, 2014).

468   For the relationship between richness and the environmental variables, we present that the

469   explanatory power of most of the environmental variables was relatively higher for the morphology-

470   based datasets, but both showed a decrease in explanatory power from finer to coarser taxonomic

471   resolution. Alpha diversity measures such as richness can potentially be strongly underestimated at

472   coarser taxonomic resolution (Maurer, 2000; Mueller, Pander & Geist, 2013), specifically for

473   morphology-based identifications (Emilson et al., 2017), which could be the reason for this decrease in

474   explanatory power. Heino (2014) observed similar patterns for richness-environment relationships

475   where the explanatory power of environmental variables decreased from fine to coarser taxonomic

476   resolution.

477   *Implications for DNA metabarcoding-based biomonitoring and restoration assessment*

478   The primary goal of ecological assessments or biomonitoring is to precisely detect biological response

479   to environmental change and impact based on community structure and composition rather than a

480   comprehensive and detailed description of the community (Bailey, Norris & Reynoldson, 2001). The

481   identification of a coarser taxonomic resolution could be advantageous for DNA metabarcoding-based

482   applications in situations where the lack of taxonomic information, e.g., poor reference database, might

483   severely affect the quality of biological assessments for all major taxonomic groups. Nonetheless, it

484   should be noted that the convenience of using coarser taxonomic levels should be based on the

485   redundant and in parallel environmental responses of the multiple finer taxa, e.g., species or populations,

486   in the same coarser taxonomic group (Balmford, Lyon & Lang, 2000).

487   Different ecological patterns and processes are scale-dependent (Bracken et al., 2017; de

488   Oliveira et al., 2020). In this study, most of the datasets revealed similar patterns of community

489   composition and structure to differentiate both at river and reach scales but failed to differentiate at river

490   × reach scale. We expected to observe difference in the community composition of the restored gravel

491   bars along the dam-influenced river against the gravel bars in its tributaries since most of the gravel bars

492   in the dam-influenced river were constructed via fluvial deposition of locally added sediments or by

493   mechanical construction of gravel islands and bars, whereas the tributaries assessed have naturally

19

494 created gravel bars (Gaeuman, 2014). However, the different datasets did not present significant

495 differentiation of the community structure and composition.

496 To differentiate the gravel bar communities in the dam-influenced river against the gravel bars

497 in the tributaries, and the communities from non-gravel bar sites, we performed indicator taxa analysis

498 and successfully determined genera and species-level taxa in distinguishing the four groups within the

499 river × reach scale (i.e., DAM+GB+, DAM+GB-, DAM-GB+, and DAM-GB-). Both methods identified

500 *Simulium* as a significant indicator genera of the dam-impacted Trinity River's gravel bar sites. However,

501 only the DNA metabarcoding dataset showed significant adjusted *p*-values after a false discovery rate

502 (FDR) estimation proving the method's robustness compared to the morphology-based identification.

503 *Simulium* is a genus of blackflies from the family Simuliidae (Diptera). DNA metabarcoding

504 identified nine *Simulium* species, i.e., *S. vittatum*, *S. vittatum* complex, *S. saxoxum*, *S. defoliarti*, *S.*

505 *tuberosum* complex, *S. canadense*, *S. bracteatum*, *S. arcticum* complex, and *S. annulitarse*, whereas

506 morphological identifications were only made up to the genus-level. Additionally, the LEfSe analysis on

507 the DNA metabarcoding dataset identified *S. vittatum* alongside *Dicosmoecus gilvipes, Hesperoperla*

508 *pacifica,* and *Sialis californica* as indicator species of the dam-impacted reaches with gravel bars.

509 Blackfly larvae are dominant suspension feeders in most lotic ecosystems (Zhang et al., 1998) and are

510 considered efficient and opportunistic colonizers and filterers with an important association with

511 suspended particles and predators. They inhabit fast-flowing environments that assure high amounts of

512 transported materials (Lock, Adriaens, & Goethals, 2014). Simuliid communities reflect environmental

513 conditions and have been used as indicators of ecological restoration or degradation, given that their

514 composition is closely related to physicochemical variables and morphological characteristics of running

515 water environments (Kazanci, 2006). The recovery of *Simulium* spp. and these other taxa indicate the

516 restoration effect of gravel bar constructions in the dam-fragmented Trinity river.

517 **CONCLUSION**

518 Information on the influence of different taxonomic and numerical resolutions of benthic

519 macroinvertebrate community composition and structure on their responses to environmental

20

520 disturbance or management, e.g., river restoration, may relieve certain limitations or challenges in the

521 aim for finer taxonomic resolution, i.e., species or haplotype-level assignment from DNA metabarcoding-

522 based assessments. Here we compared the performance of different taxonomic and numerical

523 resolutions of DNA metabarcoding with consequent comparison to traditional morphology-based

524 identification and how it affects assessment outcomes on the response of benthic macroinvertebrate

525 communities to the restoration or construction of gravel bars conducted in the dam-impacted Trinity

526 River, with the non-dam influenced tributaries serving as the reference sites. DNA metabarcoding

527 detected 93% of the taxa identified with morphological identification and provided finer taxonomic

528 resolution to the species and haplotype (ASV) levels. We also reported significant correlations between

529 morphological sample abundance, biomass, and DNA metabarcoding read abundance, validating and

530 reinforcing the reliability of using sequence abundance for downstream diversity analyses. This also

531 supports the potential of DNA metabarcoding's application for quantitative analyses. Moreover, we

532 observed a relatively high and significant congruence in macroinvertebrate community structure and

533 composition of different taxonomic and numerical resolutions of both methods, indicating a satisfactory

534 surrogacy between the two approaches and their varying identification levels and data transformation.

535 Although the community-environmental association were significant for all datasets but showed varying

536 significant associations against the physicochemical parameters, our observations still imply that

537 coarser taxonomic resolution could be advantageous for DNA metabarcoding-based applications in

538 situations where the lack of taxonomic information, e.g., poor reference database, might severely affect

539 the quality of biological assessments.

540 **Acknowledgments**

References

547

548 Aylagas, E., Borja, Á., Muxika, I., & Rodríguez-Ezpeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into
549   routine marine ecological status assessment networks. *Ecological Indicators*, 95, 194-202.

550 Bailet, B., Bouchez, A., Franc, A., Frigerio, J. M., Keck, F., Karjalainen, S. M., ... & Kahlert, M. (2019). Molecular versus
551   morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status.
552   *Metabarcoding and Metagenomics*, 3: e34002.

553 Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2001). Taxonomic resolution of benthic macroinvertebrate communities in
554   bioassessments. *Journal of the North American Benthological Society*, 20(2), 280-286.

555 Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-
556   generation DNA sequencing. *Molecular Ecology*, 21(8), 2039-2044.

557 Balmford, A., Lyon, A. J. E., & Lang, R. M. (2000). Testing the higher-taxon approach to conservation planning in a megadiverse
558   group: the macrofungi. *Biological Conservation*, 93(2), 209-217.

559 Beechie, T.J., Sear, D.A., Olden, J.D., Pess, G.R., Buffington, J.M., Moir, H., Roni, P., Pollock, M.M., 2010. Process-based
560   principles for restoring river ecosystems. *BioScience*. 60 (3), 209-222.

561 Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport
562   processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, 52(4), 603-679.

563 Bracken, M. E., Douglass, J. G., Perini, V., & Trussell, G. C. (2017). Spatial scale mediates the effects of biodiversity on marine
564   primary producers. *Ecology*, 98(5), 1434-1443.

565 Brito, J. G., Martins, R. T., Oliveira, V. C., Hamada, N., Nessimian, J. L., Hughes, R. M., ... & de Paula, F. R. (2018). Biological
566   indicators of diversity in tropical streams: Congruence in the similarity of invertebrate assemblages. *Ecological Indicators*, 85,
567   85-92.

568 Buchner, D., Beermann, A. J., Laini, A., Rolauffs, P., Vitecek, S., Hering, D., & Leese, F. (2019). Analysis of 13,312 benthic
569   invertebrate samples from German streams reveals minor deviations in ecological status class between abundance and
570   presence/absence data. *PloS one*, 14(12), e0226547.

571 Buchner, D., & Leese, F. (2020). BOLDigger–a Python package to identify and organise sequences with the Barcode of Life Data
572   systems. *Metabarcoding and Metagenomics*, 4, e53535.

573 Bush, A., Compson, Z. G., Monk, W., Porter, T. M., Steeves, R., Emilson, E. J., ... & Baird, D. (2019). Studying ecosystems with
574   DNA metabarcoding: Lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution*, 7, 434.

575 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution
576   sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581.

577 Chessman, B., Williams, S., & Besley, C. (2007). Bioassessment of streams with macroinvertebrates: effect of sampled habitat
578   and taxonomic resolution. *Journal of the North American Benthological Society*, 26(3), 546-565.

579 Cordier, T., Esling, P., Lejzerowicz, F., Visco, J., Ouadahi, A., Martins, C., ... & Pawlowski, J. (2017). Predicting the ecological
580   quality status of marine environments from eDNA metabarcoding data using supervised machine learning. *Environmental
581   Science & Technology*, 51(16), 9118-9126.

582 Caradima, B., Reichert, P., & Schuwirth, N. (2020). Effects of site selection and taxonomic resolution on the inference of stream
583   invertebrate responses to environmental conditions. *Freshwater Science*, 39(3), 415-432.

584 Carew, M. E., Coleman, R. A., & Hoffmann, A. A. (2018). Can non-destructive DNA extraction of bulk invertebrate samples be
585   used for metabarcoding? *PeerJ*, 6, e4980.

586 Curry, C. J., Gibson, J. F., Shokralla, S., Hajibabaei, M., & Baird, D. J. (2018). Identifying North American freshwater invertebrates
587   using DNA barcodes: are existing COI sequence libraries fit for purpose? *Freshwater Science*, 37(1), 178-189.

588 De Caceres, M., Jansen, F., & De Caceres, M. M. (2016). Package 'indicspecies'. indicators, 8, 1.

589 de Oliveira Jr, S. S., Ortega, J. C., dos Santos Ribas, L. G., Lopes, V. G., & Bini, L. M. (2020). Higher taxa are sufficient to
590   represent biodiversity patterns. *Ecological Indicators*, 111, 105994.

591 Dunbar, M. J., Warren, M., Extence, C., Baker, L., Cadman, D., Mould, D. J., ... & Chadd, R. (2010). Interaction between
592   macroinvertebrates, discharge and physical habitat in upland rivers. *Aquatic Conservation: Marine and Freshwater
593   Ecosystems*, 20(S1), S31-S44.

594   Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA
595   metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*,
596   8(10), 1265-1275.

597   Elbrecht, V., & Leese, F. (2017). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate
598   bioassessment. *Frontiers in Environmental Science*, 5, 11.

599   Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA
600   metabarcoding data. *PeerJ*, 6, e4644.

601   Elbrecht, V., & Steinke, D. (2019). Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. *Freshwater
602   Biology*, 64(2), 380-387.

603   Emilson, C. E., Thompson, D. G., Venier, L. A., Porter, T. M., Swystun, T., Chartrand, D., ... & Hajibabaei, M. (2017). DNA
604   metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an
605   environmental gradient. *Scientific Reports*, 7(1), 1-11.

606   Ficetola, G. F., Boyer, F., Valentini, A., Bonin, A., Meyer, A., Dejean, T., ... & Taberlet, P. (2020). Comparison of markers for the
607   monitoring of freshwater benthic biodiversity through DNA metabarcoding. *Molecular Ecology*, DOI:
608   https://doi.org/10.1111/mec.15632.

609   Gaeuman, D. (2014). High-flow gravel injection for constructing designed in-channel features. *River Research and Applications*,
610   30(6), 685-706.

611   Gaeuman, D., Stewart, R., Schmandt, B., & Pryor, C. (2017). Geomorphic response to gravel augmentation and high-flow dam
612   release in the Trinity River, California. *Earth Surface Processes and Landforms*, 42(15), 2523-2540.

613   Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-scale biomonitoring of
614   remote and threatened ecosystems via high-throughput sequencing. *PloS one*, 10(10), e0138432.

615   Godoy, B. S., Faria, A. P. J., Juen, L., Sara, L., & Oliveira, L. G. (2019). Taxonomic sufficiency and effects of environmental and
616   spatial drivers on aquatic insect community. *Ecological Indicators*, 107, 105624.

617   Graf, W. L. (2006). Downstream hydrologic and geomorphic effects of large dams on American rivers. *Geomorphology*, 79(3-4),
618   336-360.

619   Heino, J., & Soininen, J. (2007). Are higher taxa adequate surrogates for species-level assemblage patterns and species richness
620   in stream organisms? *Biological Conservation*, 137(1), 78-89.

621   Heino, J. (2010). Are indicator groups and cross-taxon congruence useful for predicting biodiversity in aquatic ecosystems?
622   *Ecological Indicators*, 10(2), 112-117.

623   Heino, J. (2014). Taxonomic surrogacy, numerical resolution and responses of stream macroinvertebrate communities to
624   ecological gradients: are the inferences transferable among regions? *Ecological Indicators*, 36, 186-194.

625   Herbst, D. B., & Silldorff, E. L. (2006). Comparison of the performance of different bioassessment methods: similar evaluations of
626   biotic integrity from separate programs and procedures. Journal of the north american Benthological society, 25(2), 513-530.

627   Hleap, J. S., Littlefair, J. E., Steinke, D., Hebert, P. D., & Cristescu, M. E. (2021). Assessment of current taxonomic assignment
628   strategies for metabarcoding eukaryotes. *Molecular Ecology Resources*, DOI: https://doi.org/10.1111/1755-0998.13407.

629   Jones, F. C. (2008). Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic
630   macroinvertebrates. *Environmental Reviews*, 16(NA), 45-69.

631   Kazanci, N. İ. L. G. Ü. N. (2006). Ordination of Simuliidae and climate change impact. *Acta Entomologica Serbica Suplement*, 69-
632   76.

633   Laini, A., Beermann, A. J., Bolpagni, R., Burgazzi, G., Elbrecht, V., Zizka, V. M., ... & Viaroli, P. (2020). Exploring the potential of
634   metabarcoding to disentangle macroinvertebrate community dynamics in intermittent streams. *Metabarcoding and
635   Metagenomics*, 4, e51433.

636   Landeiro, V. L., Bini, L. M., Costa, F. R., Franklin, E., Nogueira, A., de Souza, J. L., ... & Magnusson, W. E. (2012). How far can
637   we go in simplifying biomonitoring assessments? An integrated analysis of taxonomic surrogacy, taxonomic sufficiency and
638   numerical resolution in a megadiverse region. *Ecological Indicators*, 23, 366-373.

639   Lepori, F., Palm, D., Brännäs, E., & Malmqvist, B. (2005). Does restoration of structural heterogeneity in streams enhance fish
640   and macroinvertebrate diversity? *Ecological Applications*, 15(6), 2060-2071.

641   Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE*
642   *transactions on visualization and computer graphics*, 20(12), 1983-1992.

643   Lock, K., Adriaens, T., & Goethals, P. (2014). Effect of water quality on blackflies (Diptera: Simuliidae) in Flanders (Belgium).
644   *Limnologica*, 44, 58-65.

645   Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal, 17(1), 10-12.

646   Maurer, D. (2000). The dark side of taxonomic sufficiency (TS). *Marine Pollution Bulletin*, 40(2), 98-101.

647   McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome
648   census data. *PloS one*, 8(4), e61217.

649   Mendoza-Lera, C., Datry, T., 2017. Relating hydraulic conductivity and hyporheic zone biogeochemical processing to conserve
650   and restore river ecosystem services. *Science of the Total Environment*. 579, 1815-1821.

651   Merritt, R. W., Cummins, K. W., & Berg, M. B. (2008). An introduction to the aquatic insects of North America. Fourth Edtition.
652   KendallHunt Publishing Company. ISBN 978-0-7575-6321-8.

653   Merz, J. E., & Ochikubo Chan, L. K. (2005). Effects of gravel augmentation on macroinvertebrate assemblages in a regulated
654   California river. *River Research and Applications*, 21(1), 61-74.

655   Meyer, A., Boyer, F., Valentini, A., Bonin, A., Ficetola, G. F., Beisel, J. N., ... & Usseglio-Polatera, P. (2020). Morphological vs.
656   DNA metabarcoding approaches for the evaluation of stream ecological status with benthic invertebrates: Testing different
657   combinations of markers and strategies of data filtering. *Molecular Ecology*, DOI: https://doi.org/10.1111/mec.15723.

658   Mueller, M., Pander, J., & Geist, J. (2013). Taxonomic sufficiency in freshwater ecosystems: effects of taxonomic resolution,
659   functional traits, and data transformation. *Freshwater Science*, 32(3), 762-778.

660   Monaghan, M. T., Robinson, C. T., Spaak, P., & Ward, J. V. (2005). Macroinvertebrate diversity in fragmented Alpine streams:
661   implications for freshwater conservation. Aquatic Sciences, 67(4), 454-464.

662   Naimi, B., 2015. usdm: Uncertainty analysis for species distribution models. R Package. 1, 1-12.

663   Ock, G., Gaeuman, D., McSloy, J., & Kondolf, G. M. (2015). Ecological functions of restored gravel bars, the Trinity River, California.
664   *Ecological Engineering*, 83, 49-60.

665   Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.,
666   Wagner, H., Oksanen, M.J., 2013. Package 'vegan'. Community ecology package. 2 (9), 1-295.

667   Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC*
668   *Bioinformatics*, 12(1), 1-10.

669   Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., ... & Feio, M. J. (2018). The future
670   of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems.
671   *Science of the Total Environment*, 637, 1295-1310.

672   Pereira, M. B., Wallroth, M., Jonsson, V., & Kristiansson, E. (2018). Comparison of normalization methods for the analysis of
673   metagenomic gene abundance data. *BMC Genomics*, 19(1), 1-17.

674   Petts, G. E., & Gurnell, A. M. (2005). Dams and geomorphology: research progress and future directions. *Geomorphology*, 71(1-
675   2), 27-47.

676   Pires, M. M., Grech, M. G., Stenert, C., Maltchik, L., Epele, L. B., McLean, K. I., ... & Batzer, D. P. (2021). Does taxonomic and
677   numerical resolution affect the assessment of invertebrate community structure in New World freshwater wetlands?
678   *Ecological Indicators*, 125, 107437.

679   R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna,
680   Austria. URL https://www.R-project.org/.

681   Rice, S. P., Greenwood, M. T., & Joyce, C. B. (2001). Tributaries, sediment sources, and the longitudinal organisation of
682   macroinvertebrate fauna along river systems. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(4), 824-840.

683   Sackett, J.D., Shope, C.L., Bruckner, J.C., Wallace, J., Cooper, C.A., Moser, D.P., 2019. Microbial Community Structure and
684   Metabolic Potential of the Hyporheic Zone of a Large Mid-Stream Channel Bar. *Geomicrobiology Journal*. 36 (9), 765-776.

685   Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker
686   discovery and explanation. *Genome Biology*, 12(6), 1-18.

687   Serrana, J. M., Yaegashi, S., Kondoh, S., Li, B., Robinson, C. T., & Watanabe, K. (2018). Ecological influence of sediment bypass
688   tunnels on macroinvertebrates in dam-fragmented rivers by DNA metabarcoding. *Scientific Reports*, 8(1), 1-10.

689   Serrana, J. M., Miyake, Y., Gamboa, M., & Watanabe, K. (2019). Comparison of DNA metabarcoding and morphological
690   identification for stream macroinvertebrate biodiversity assessment and monitoring. *Ecological Indicators*, 101, 963-972.
691   Serrana, J. M. Li, B., Sumi, T., Takemon, Y. & Watanabe, K. (2021). Data from: Implications of taxonomic and numerical resolution
692   on DNA metabarcoding-based inference of benthic macroinvertebrate responses to river restoration. *figshare*,
693   https://doi.org/10.6084/m9.figshare.15035412.v1.
694   Shetty, S. A. & Lahti, L. (2018). microbiomeutilities: An R package for utilities to guide in-depth marker gene amplicon data analysis
695   (Version 0.99.00). *Zenodo*. DOI:10.5281/zenodo.1471685.
696   Staats, M., Arulandhu, A. J., Gravendeel, B., Holst-Jensen, A., Scholtens, I., Peelen, T., ... & Kok, E. (2016). Advances in DNA
697   metabarcoding for food and wildlife forensic species identification. *Analytical and Bioanalytical Chemistry*, 408(17), 4615-
698   4630.
699   Strimmer, K. (2008). A unified approach to false discovery rate estimation. *BMC Bioinformatics*, 9(1), 303.
700   Taberlet, P., Bonin, A., Coissac, E., & Zinger, L. (2018). Environmental DNA: For biodiversity research and monitoring. Oxford
701   University Press.
702   Wang, J., Ding, C., Tao, J., Jiang, X., Heino, J., Ding, L., ... & He, D. (2020). Damming affects riverine macroinvertebrate
703   metacommunity dynamics: Insights from taxonomic and functional beta diversity. *Science of The Total Environment*, 142945.
704   Zhang, Y., Malmqvist, B., & Englund, G. (1998). Ecological processes affecting community structure of blackfly larvae in regulated
705   and unregulated rivers: a regional study. *Journal of Applied Ecology*, 35(5), 673-686.
706   Zizka, V. M., Geiger, M. F., & Leese, F. (2020). DNA metabarcoding of stream invertebrates reveals spatio-temporal variation but
707   consistent status class assessments in a natural and urban river. *Ecological Indicators*, 115, 106383.

Data availability statement

The raw sequence data were deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRR12620160). The interactive Krona charts, ASV sequences, morphological and metabarcoding data matrices, taxonomy and sample metadata, and the scripts for the bioinformatics analyses were deposited in the Figshare data repository at https://doi.org/10.6084/m9.figshare.15035412.v1 (Serrana et al., 2021).

ORCID

Joeselle M. Serrana: https://orcid.org/0000-0002-6967-5407; Tetsuya Sumi: https://orcid.org/0000-0002-1423-7477; Kozo Watanabe: https://orcid.org/0000-0002-7062-595X

Author contributions

K.W., B.L. and J.M.S. conceptualized and designed the study; J.M.S., B.L., Y.T., and T.S. conducted fieldwork; B.L. performed the morphological identification and laboratory work; J.M.S. performed

720   bioinformatics analyses and data interpretation. The manuscript was drafted by J.M.S. with writing inputs

721   and revisions from all coauthors.

722   **Declaration of competing interests**

723   The authors declare that they have no known competing financial interests or personal relationships that

724   could have appeared to influence the work reported in this paper.

725   **Supplementary information**

726   The supplementary data to this article is provided as a supplemental file (DOCX).

727   **Tables**

728   **TABLE 1** Absolute abundance and presence/absence count of morphologically identified samples and
729   DNA metabarcoding reads.

	Sample	Absolute Abundance						Presence/Absence					
		Morphological		Metabarcoding				Morphological		Metabarcoding			
		Family	Genera	Family	Genera	Species	ASV	Family	Genera	Family	Genera	Species	ASV
DAM+GB+	B1DZ	129	89	130,170	121,928	70,793	156,643	9	6	12	22	14	89
	B1UZ	183	116	187,318	180,510	61,294	187,564	10	7	11	14	12	86
	B2DZ	329	237	185,850	185,850	106,759	185,850	11	8	9	15	11	107
	B2UZ	135	84	156,758	153,952	140,890	159,088	15	9	11	12	10	50
	B3DZ	74	56	128,926	127,191	125,937	130,073	14	11	11	16	11	59
	B3UZ	106	102	268,978	268,490	182,477	268,978	10	9	11	12	14	40
	B4DZ	257	180	114,254	111,005	99,196	119,038	18	12	16	19	17	64
	B4UZ	84	71	169,180	168,532	118,521	191,302	13	11	11	12	11	62
	B5DZ	230	145	139,844	136,449	95,521	150,309	15	11	13	23	17	96
	B5UZ	292	220	113,537	111,297	63,714	118,528	16	11	12	17	15	95
	B6DZ	133	103	39,554	38,653	21,468	39,739	19	12	15	18	13	57
	B6UZ	158	104	159,904	154,852	78,772	162,097	20	12	19	24	19	113
DAM-GB+	B7DZ	152	81	207,708	200,217	164,470	219,310	21	14	19	23	21	111
	B7UZ	124	81	356,340	349,293	255,165	397,134	23	15	27	37	30	156
	B8DZ	100	63	135,175	133,045	119,928	144,216	12	8	13	15	12	61
	B8UZ	229	128	243,003	235,338	228,201	265,259	17	12	15	20	19	95
DAM-GB-	F1US	131	80	79,175	61,883	54,355	106,961	15	10	16	24	22	104
	F1DS	177	101	69,670	61,046	52,307	91,510	23	15	21	29	23	101
	F2US	203	78	79,142	67,585	58,874	90,348	15	8	18	18	15	57
	F2DS	148	101	79,634	56,727	53,885	133,737	17	11	13	17	12	54
DAM+GB-	F3US	162	126	343,289	335,230	255,603	350,837	12	7	15	22	14	90
	F3DS	255	173	125,479	121,543	101,993	128,327	16	9	14	19	14	82
	F4US	130	107	116,511	110,710	83,680	159,527	13	10	10	12	12	65
	F4DS	132	111	72,087	70,370	46,855	126,626	17	13	10	10	9	61
Total		**4,053**	**2,737**	**3,701,486**	**3,561,696**	**2,640,658**	**4,083,001**	**31**	**25**	**42**	**80**	**69**	**1,114**

Sampling site abbreviations: "DAM-GB-" stands for Reference Non-gravel bar reach; "DAM-GB+" stands for Reference Gravel Bars; "DAM+GB-" stands for Trinity Non-gravel bar reach; "DAM+GB+" stands for Trinity Gravel Bars; "DZ" represents the down-welling and "UZ" up-welling zones of a gravel bar (B); "US" represents the up-stream and "DS" down-stream points of collection for the non-gravel bar reaches of the rivers (F) (approx. 20-m length).

730

TABLE 2 Permutational analysis of variance (PERMANOVA) results for each scale.

| Taxonomic Rank			River			Reach			River × Reach		
			F.Model	R^2	Pr(>F)	F.Model	R^2	Pr(>F)	F.Model	R^2	Pr(>F)
Absolute Abundance	Morp	Family	6.15	0.21	**0.001**	2.21	0.07	0.038	1.46	0.05	0.171
		Genus	4.44	0.16	**0.001**	2.42	0.09	0.015	1.22	0.04	0.282
	Meta	Family	6.15	0.21	**0.001**	2.25	0.08	0.031	1.32	0.04	0.241
		Genus	5.65	0.20	**0.001**	2.16	0.07	0.027	1.01	0.04	0.398
		Species	3.95	0.15	**0.001**	2.02	0.07	0.022	1.21	0.04	0.284
		ASV	2.51	0.10	**0.001**	1.19	0.05	0.142	1.23	0.05	0.122
Presence/Absence	Morp	Family	7.25	0.24	**0.001**	1.97	0.06	0.074	1.39	0.05	0.239
		Genus	3.15	0.12	**0.008**	2.31	0.09	0.035	1.35	0.05	0.276
	Meta	Family	7.31	0.23	**0.001**	2.87	0.09	0.016	1.03	0.03	0.442
		Genus	6.18	0.21	**0.001**	2.73	0.09	0.011	0.84	0.03	0.579
		Species	4.65	0.16	**0.001**	2.54	0.09	0.014	1.18	0.04	0.319
		ASV	2.47	0.10	**0.001**	1.23	0.05	0.101	1.20	0.05	0.148

Significant values are indicated in highlight (<0.05) and bold print (<0.01). "Morp" indicates morphologically identified; "Meta" for DNA metabarcoding.

**TABLE 3** Canonical correspondence analysis (CCA) of the environmental variables.

Taxonomic Rank			CCA Model				Physico-chemical Parameters						
			F	Pr(>F)	R^2	Adj. R^2	pH	EC	DO	NH4	AFDM	TSS	NO3
Absolute Abundance	Morp	Family	1.95	**0.00**	0.46	0.22	0.31	0.03	**0.01**	**0.01**	**0.01**	0.04	**0.01**
		Genus	1.81	**0.00**	0.44	0.20	0.37	0.07	**0.01**	**0.01**	**0.01**	0.12	0.03
	Meta	Family	1.61	**0.00**	0.41	0.16	0.13	**0.01**	0.02	0.09	**0.01**	0.39	0.06
		Genus	1.48	**0.00**	0.39	0.13	0.18	**0.01**	0.08	0.06	**0.01**	0.21	0.03
		Species	1.35	**0.01**	0.37	0.10	0.02	0.30	0.09	0.20	0.04	0.36	0.05
		ASV	1.07	0.02	0.32	0.02	**0.00**	0.13	0.59	0.28	0.06	0.55	0.32
Presence/Absence	Morp	Family	1.86	**0.00**	0.45	0.21	0.18	0.05	**0.01**	**0.01**	**0.00**	0.06	**0.01**
		Genus	1.78	**0.00**	0.44	0.19	0.14	0.10	**0.00**	**0.01**	0.03	0.15	**0.01**
	Meta	Family	1.55	**0.00**	0.40	0.14	0.04	0.03	0.03	0.04	0.05	0.67	0.06
		Genus	1.37	**0.00**	0.38	0.10	0.12	**0.01**	0.13	0.06	0.02	0.37	0.03
		Species	1.31	**0.00**	0.36	0.09	0.02	0.52	0.09	0.16	0.07	0.46	0.04
		ASV	1.06	0.05	0.32	0.02	**0.00**	0.24	0.69	0.29	0.09	0.53	0.36

ANOVA results (R^2- and p-values). Significant values are indicated in highlight (<0.05) and bold print (<0.01). "Morp" indicates morphologically identified; "Meta" for DNA metabarcoding.
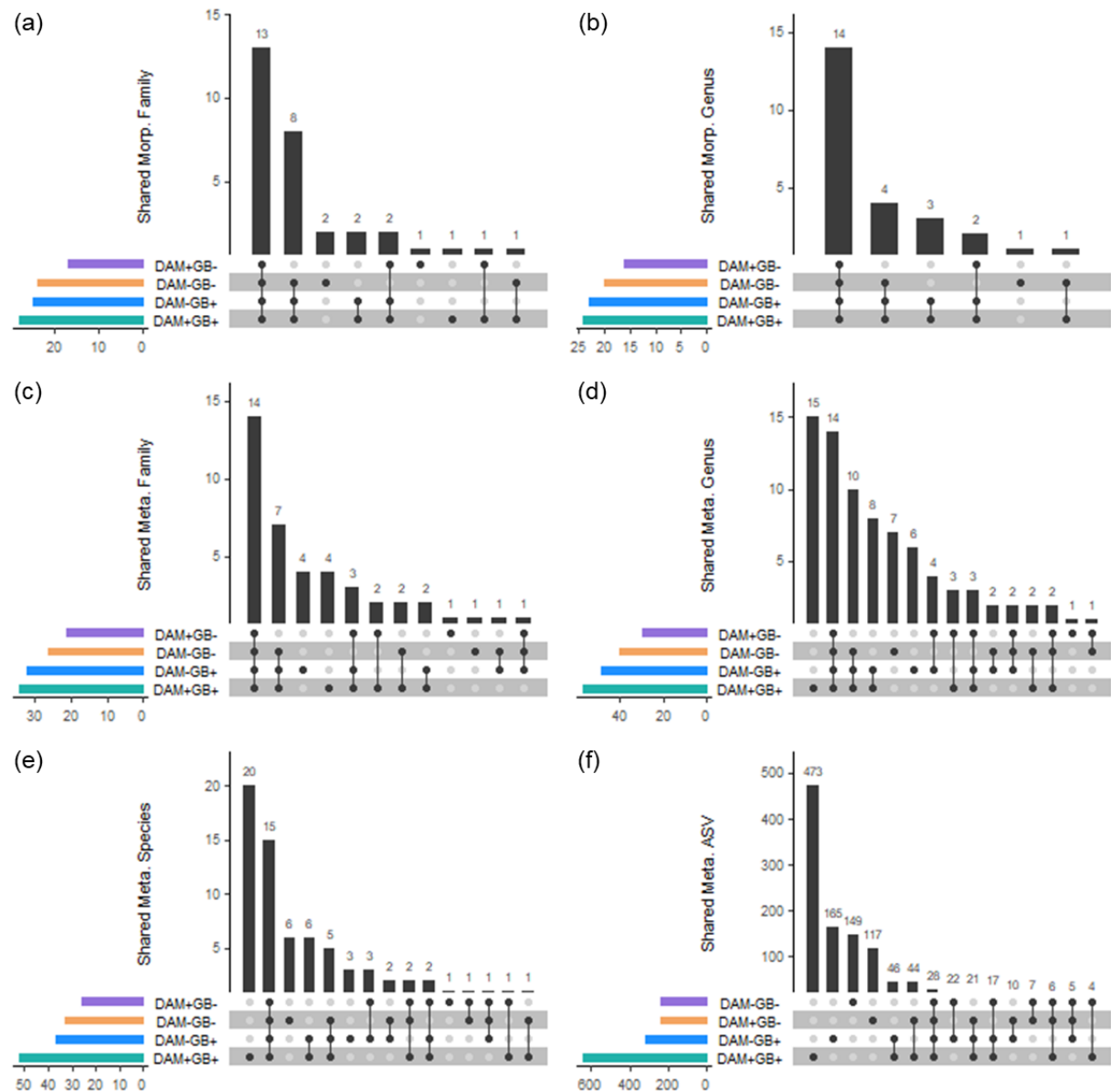
Figures
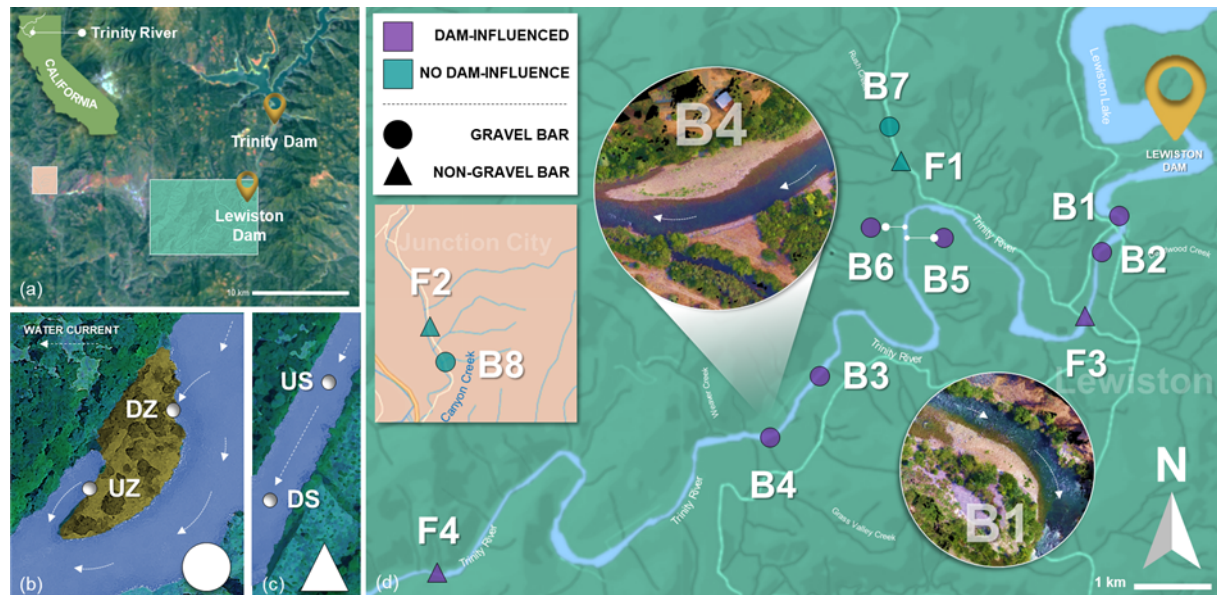
FIGURE 1 Sampling location. The Trinity River and its tributaries impounded by the Trinity Dam and the Lewiston Dam in California, USA (a). Gravel bars were sampled at the head (down-welling zone; DZ) and tail (up-welling zone; UZ) points (b). Non-gravel bar sites were assessed, and samples were collected on up- (US) and downstream (DS) points (c). In total, eight gravel bars and four non-gravel bar sites were assessed in the study; gravel bars B1 and B4 drone photos shown as examples (d). Parts of the map was generated from Google Earth Pro (version 7.3.2.5776; https://www.google.com/earth/).
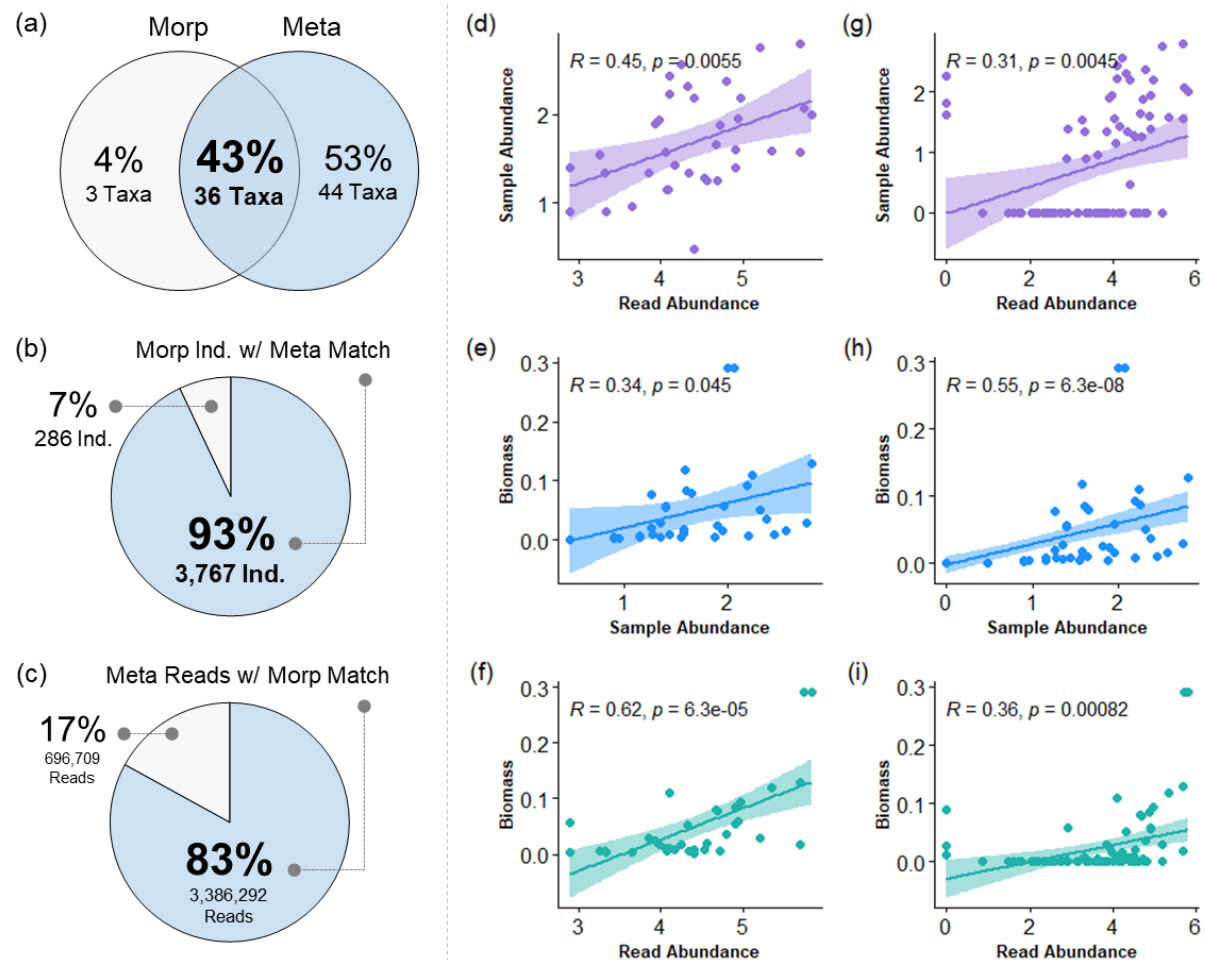
FIGURE 2 Venn diagram showing the shared and unique taxa between the morphological ("Morp") and metabarcoding ("Meta") methods (a). The pie charts present the number of morphologically identified individuals (b) and their corresponding read abundance (c) detected in metabarcoding. Pearson correlations between the sample abundance (i.e., individual count), biomass (i.e., dry biomass in grams), and metabarcoding read abundance (i.e., absolute sequence count). Showing the analysis excluding (d-f), and including (g-i) false positive detections. Data were log-transformed prior to analysis.

FIGURE 3 Number of taxa shared between River × Reach scale for the morphology-based (a-b) and metabarcoding (c-f) data.
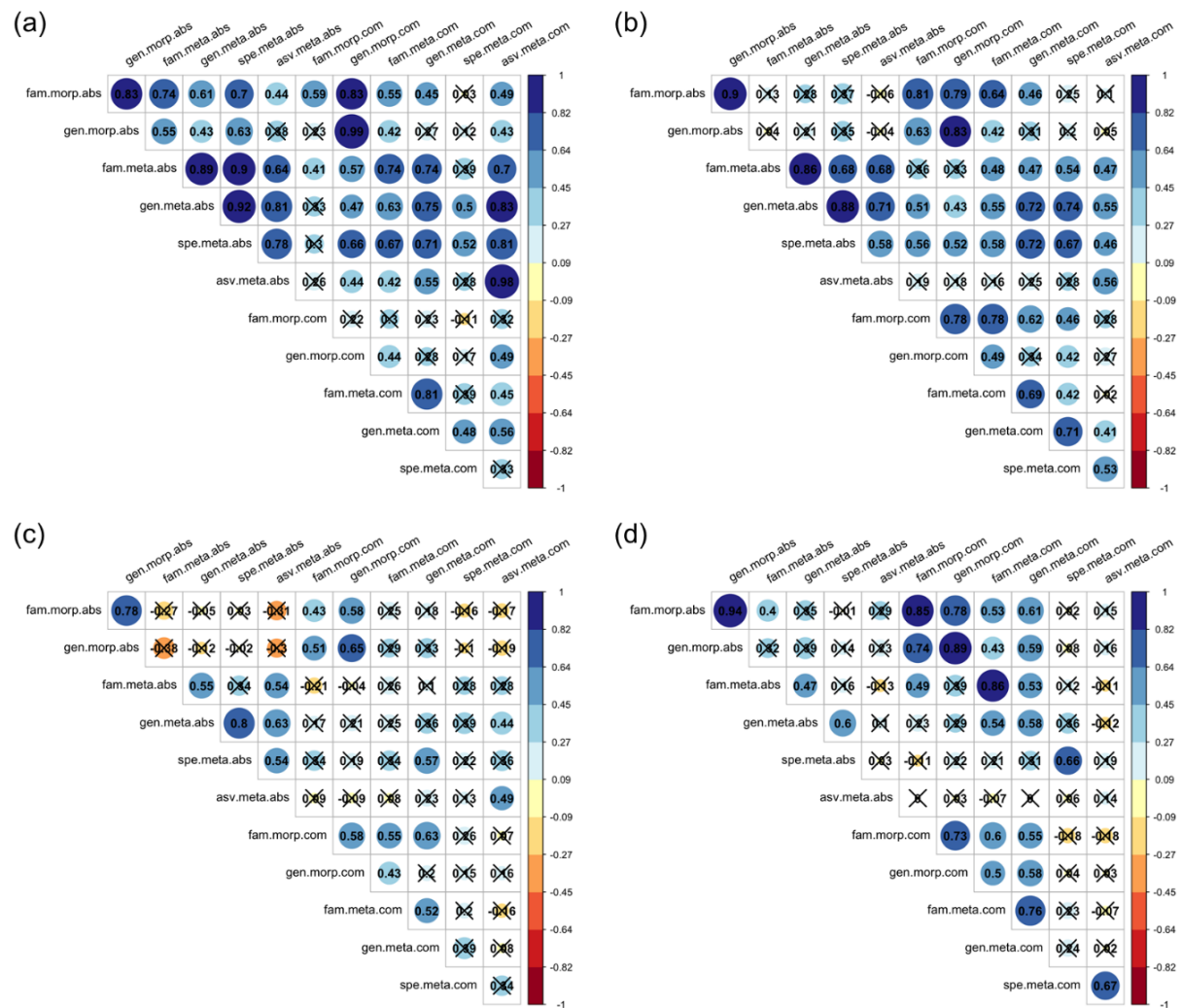
FIGURE 4 Pairwise Pearson's correlation of alpha diversity estimates. Chao1 richness (a), Shannon diversity (b), Berger-Parker dominance index (c), and rare taxa abundance (d). Correlation boxes with Xs are not significant at $p = 0.05$. "morp" indicates morphologically identified; "meta" for DNA metabarcoding; "gen" for genus-level identification; "fam" for family-level; "spe" for species–level; "asv" for the amplicon sequence variant (ASV) level dataset.
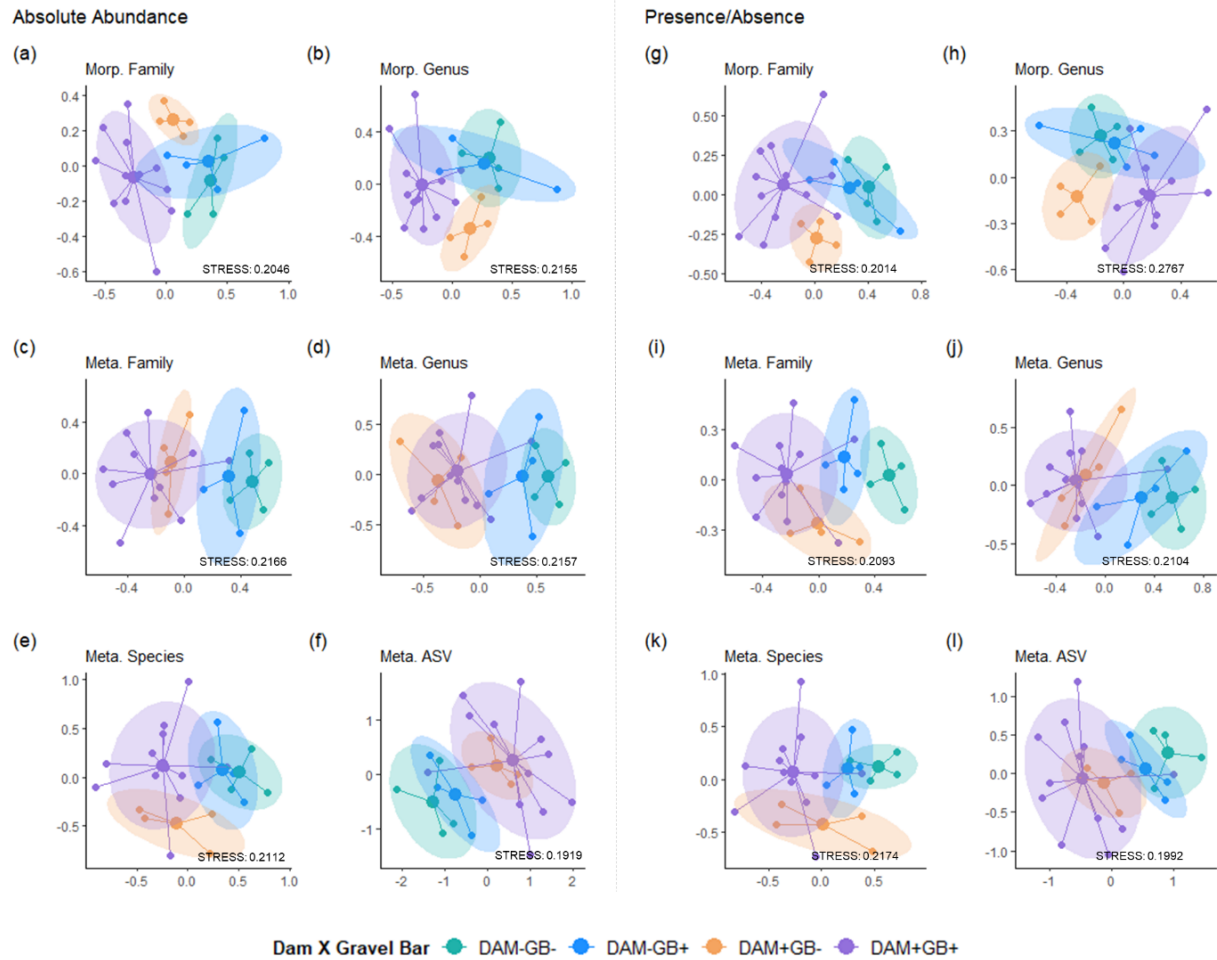
FIGURE 5 Non-metric multidimensional scaling (NMDS) plot on Axes 1 (x-axis) and 2 (y-axis) based on Bray-Curtis dissimilarity for absolute abundance (a-f) and presence/absence (g-l) data of the macroinvertebrates detected by morphological ("Morp") and DNA metabarcoding-based ("Meta") methods.