Ecological Indicators 125 (2021) 107537

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

How does molecular taxonomy for deriving river health indices correlate with traditional morphological taxonomy?

M.E. Shackleton [a,*], K.A. Dafforn [b], N.P. Murphy [a], P. Greenfield [c], M. Cassidy [d], C.H. Besley [d]

[a] Centre for Freshwater Ecosystems, La Trobe University, Australia
[b] Department of Earth and Environmental Sciences, Macquarie University, Sydney, Australia
[c] CSIRO, Energy, Sydney, Australia
[d] Sydney Water, Parramatta, Australia

ARTICLE INFO

Keywords:
Metabarcoding
Macroinvertebrate
DNA
Biological assessment

ABSTRACT

Macroinvertebrate surveys are commonly used for assessing the health of freshwater systems around the world. Traditionally, surveying involves morphologically identifying the families, and sometimes genera, present in samples. Biological indices, derived from taxonomic lists, provide convenient ways to summarise community data and may be fairly insensitive to species-level changes in community compositions. In recent years, molecular techniques for identifying taxa have become increasingly popular and metabarcoding approaches that offer the ability to identify species from mixtures of whole animals (bulk-samples) or from environmental samples have gained much attention. However, generating accurate species lists from metabarcode data is challenging and can be impacted by sample type, choice of primers, community composition within samples, and the availability of reference sequences. This study compares the performance of molecular data extracted from bulk-samples against morphological data in calculating two biological indices (the Stream Invertebrate Grade Number Average Level 2 (SIGNAL2), which is calculated from family-level data, and a genus-level equivalent of this index, SIGNAL_SG) and one biological metric (taxon richness). Further, molecular indices and metrics derived from global, local or mixed reference DNA libraries and with varying degrees of filtering processes applied to them, are compared with respect to the strength of their relationships with morphological indices and metrics. Molecularly derived SIGNAL2 and SIGNAL_SG scores correlated strongly with morphologically derived scores, and were strongest when using a reference library containing a mix of local and global data. Molecularly derived richness metrics were moderately correlated with morphological taxa richness; however, the strongest correlations were observed when taxa that could not be assigned SIGNAL grades were omitted from analyses. This study highlights the utility of using molecular data as an objective and sensitive alternative to traditional freshwater biological assessment using macroinvertebrates.

1. Introduction

Biological assessment of freshwater systems is used globally for deriving information on stream conditions and aids in tracking the impact of management actions (Buss et al., 2015; Carew et al., 2017). Measures of river health are typically evaluated through deriving indices or metrics based on the presences, or absences, of macroinvertebrate taxa within freshwater systems. Various metrics and indices have been developed and tailored to specific regions or countries. Examples include the Australian River Assessment System (AusRivAS) (Smith et al., 1999) and Stream Invertebrate Grade Number – Average Level (SIGNAL) (Chessman, 1995) in Australia, the River Invertebrate Prediction and Classification System (RIVPACS, (Wright et al., 1998) in the United Kingdom, the Empirical Biotic Index (EBI) (Chutter, 1972) in South Africa and the Family-level Biotic Index (FBI) (Hilsenhoff, 1988) in the United States of America. In most cases, these involve identifying macroinvertebrate specimens to the family level. However, in some cases genus-level abundance data may be used (Chessman et al., 2007; Besley and Chessman, 2008). The collection and identification of macroinvertebrates can be costly, with costs increasing as the resolution of taxonomic identification increases (Marshall et al., 2006).

DNA barcodes are fragments of DNA, typically unique to individual species, that can be used to delineate and identify taxa (Hebert et al., 2003; Goldstein and DeSalle, 2011; Carew et al., 2017). There is now a

* Corresponding author.
E-mail address: m.shackleton@latrobe.edu.au (M.E. Shackleton).

https://doi.org/10.1016/j.ecolind.2021.107537
Received 5 August 2020; Received in revised form 30 January 2021; Accepted 15 February 2021
Available online 5 March 2021

large body of evidence that shows the utility of DNA barcodes for species identification. Coupled with High-Throughput Sequencing (HTS) technologies, DNA barcodes can be extracted from multiple target organisms simultaneously; an approach termed DNA metabarcoding (Taberlet et al., 2012; Ruppert et al., 2019). The presence of species can be determined from sample types ranging from "bulk" mixtures of whole animals (bulk-samples) to environmental samples, such as water, sediments or air (Ruppert et al., 2019). Metabarcoding, thus, offers an opportunity to collect species-level presence/absence data, for a range of organisms simultaneously and possibly at lower cost than traditional methods.

Metabarcoding for generating accurate species lists offers some challenges and can be impacted by sample type, choice of primers, compositions of the communities within samples, and the availability of reference sequences. High detectability rates have been achieved from samples containing whole macroinvertebrate communities (Carew et al., 2018), whereas many target taxa are missed when analysing water-based environmental DNA (eDNA) samples using standard Cytochrome Oxidase I (COI) primers (Hajibabaei et al., 2019a); however, see Leese et al. (2020) for recent improvements in this area. So-called universal primers do not capture all taxa, for instance commonly used invertebrate primers fail to detect flatworms, and the affinity of primers can vary among taxa, leading to primer bias (Kanagawa, 2003; Elbrecht and Leese, 2015). For this reason, multiple primers may need to be used to comprehensively survey fauna (Hajibabaei et al., 2019b). Ultimately, assigning correct species names to sequence data requires a comprehensive reference database with barcodes that represent species from the sampling area (Weigand et al., 2019). Where this is lacking, taxonomy will be either assigned at higher taxonomic levels or to the next best matching species, which can often be a species not from the sampled region (Shackleton and Rees, 2016).

While the above-mentioned issues may have effects on the detectability of taxa or the ability to compile comprehensive species lists, it is probable that they will have a lesser effect on being able to derive indices commonly used for inferring river health. Many such indices require identification to genus or, more commonly, family level only. Moreover, for indices that use systems to grade macroinvertebrate taxa, such as The Stream Invertebrate Grade Number – Average Level (SIGNAL, Chessman (1995)), identifications incorrectly made to closely related taxa are not likely to have a significant impact, as closely related taxa are more likely to have similar grades due to phylogenetic niche conservatism. SIGNAL is the index commonly used in Australia to characterise river health from the presence of macroinvertebrate taxa. The SIGNAL method applies a grade from 1 to 10 to each taxon, with 1 indicating taxa that are highly tolerant and 10 highly sensitive to pollution (Chessman, 2003). SIGNAL has been used extensively for biological assessment of freshwater systems and for investigating a variety of environmental impacts (Chessman et al., 2006; Lester et al., 2007; Besley and Chessman, 2008; Rose et al., 2008; Davies et al., 2010; Nichols et al., 2010; Tippler et al., 2012). The mean, or weighted mean, of all taxa within a sample provide a SIGNAL score for a reach, with scores indicating levels of impact as severe (<4), moderate (4–5), mild (5–6) and healthy (>6). The most commonly used form, SIGNAL2 (Chessman, 2003), uses a system that places grades on families. However, the SIGNAL_SG system, developed specifically for the region around Sydney, Australia, derives scores based on genus-level grades (Chessman et al., 2007). SIGNAL-SG was developed in response to suggestions that region-specific models are more suitable than those derived for the broad scale as was the case for the original version of SIGNAL (Bunn, 1995; Bunn and Davies, 2000) and later Australia-wide objectively derived SIGNAL2 (Chessman, 2003).

Taxonomic richness is also commonly used in routine biological monitoring programs. However, it often provides a poorer indicator of possible impacts to a system than SIGNAL. Growns et al. (1997) found taxa (family) richness to be a weaker measure of the effects of pollution by municipal sewage effluent and urban stormwater than SIGNAL for 12 streams of outer suburban Sydney and the lower Blue Mountains, Australia. Walsh (2006) in a study of 16 streams subject to urban disturbance in eastern Melbourne, Australia, found SIGNAL to be a more sensitive indicator than taxa (family) richness.

Very few studies have investigated the use of metabarcoding for deriving water quality indices. Carew et al. (2018) dissected tissues from macroinvertebrates in river bioassessment samples, metabarcoded COI fragments, and compared molecularly and morphologically derived indices and metrics. They found little difference between the two methods for SIGNAL2 scores, the Number of Families, Key Families and Australian Rivers Assessment System (AusRivAS) bands. Marshall and Stepien (2020) found that eDNA metabarcoding of 16S fragments revealed similar trends in multiple alpha and beta diversity metrics to those seen in morphological data.

The present study used a short, 313 base-pair, fragment of the COI barcoding region to determine whether SIGNAL2 and SIGNAL_SG biotic indices derived from DNA data are comparable to those derived from traditional, morphological data. However, to differentiate this study from other studies in this area key differences were made, some of which aimed to reduce the cost and time involved with sample preparation. Firstly, DNA was extracted from whole bulk-samples rather than dissecting tissue from individual animals as was done in Carew et al. (2018). Secondly, only a single set of primers was used in this study, compared to three sets in Carew et al. (2018) and Marshall and Stepien (2020). While this may reduce taxonomic coverage it also reduces the sample preparation time and increases the sequencing read depth available per sample. Thirdly, past studies have investigated family level metrics, whereas the present study includes a genus level metric (SIGNAL_SG). Because taxonomic assignment of Operational Taxonomic Units (OTUs) can be affected by the taxonomic composition of reference DNA databases, metrics were derived using three molecular datasets containing the same OTUs but with differing taxonomic identifications applied from different reference databases in order to investigate how incomplete barcode libraries effect metric outcomes. The effect that filtering OTUs, based on their percent contribution to samples, has on index and metric outcomes was investigated at varying thresholds. Lastly, comparisons were made between morphological and molecular taxonomic richness, including for molecular analyses richness of OTUs as well as taxa richness (i.e. genus, family), taxa lists were compared, and the detectability of taxa investigated.

2. Methods

2.1. Site selection and macroinvertebrate sampling

Macroinvertebrates were collected from 7 sites in 3 freshwater creeks (Tributary of Devlins, Lalor and Vineyard) in Sydney, Australia (Fig. 1) on multiple occasions between December 2016 and July 2017 (see Supplementary Table S1). On each sampling occasion, three replicate samples were taken from the edges of pools. Edges were sampled with hand-held nets (320 × 250-mm opening; 250-μm mesh) and sweep sampling over transects of approximately 10 m. Macroinvertebrates were subsampled selectively by the unaided eye for 30 to 60 min, with the goals of picking approximately 100 specimens per sample and a wide range of species rather than large numbers of the same species. Further details are given by Chessman (1995).

Samples were originally collected into 70% ethanol. However, this was removed for transport between laboratories (approx. 2 days) and topped up with 100% ethanol and placed in a freezer at −20 °C until DNA extraction. Macroinvertebrates were morphologically identified to genus level where possible, with the aid of microscopes and published keys and identification guides. Details of published keys and identification guides for identification of Australian invertebrates has been consolidated by the Centre for Freshwater Ecosystems (Hawking et al., 2013; https://www.mdfrc.org.au/bugguide/index.htm). Keys and guides used in morphological identification included Arachnida (Cook,

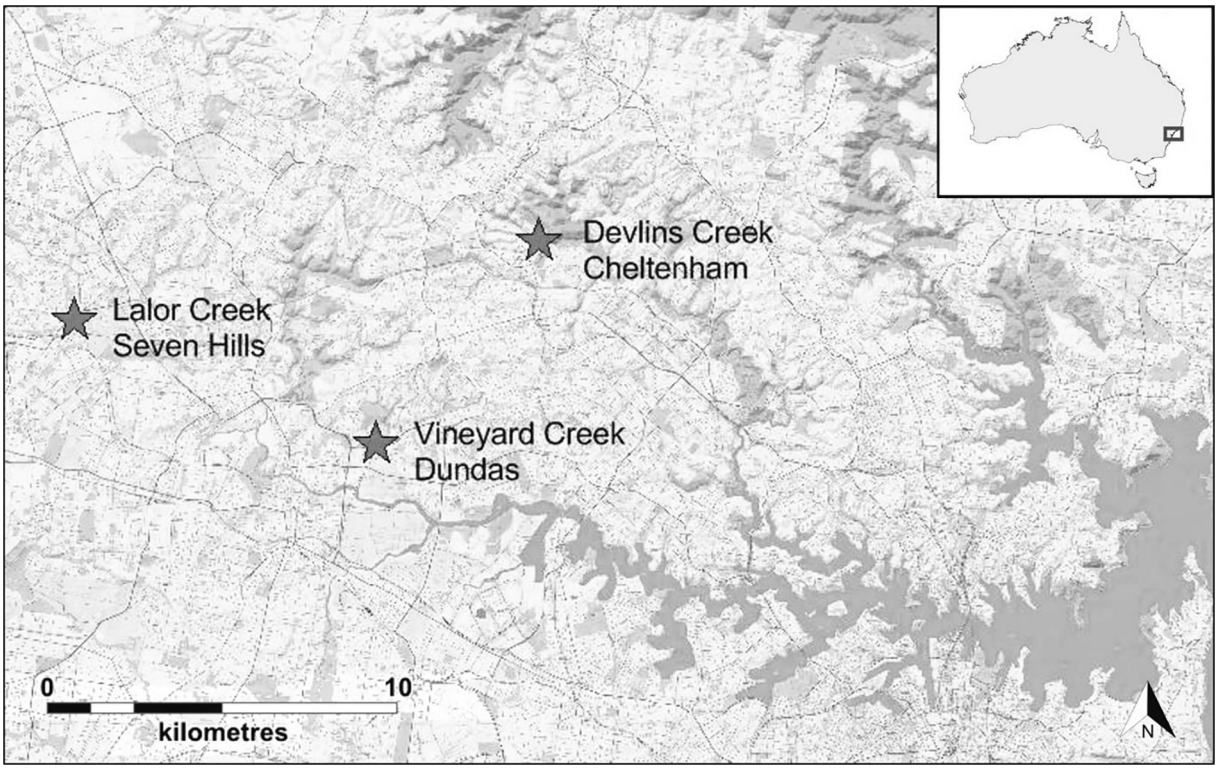

Fig. 1. Map of sample locations.

1974, 1986; Harvey, 1996; Harvey and Growns, 1998), Annelida (Pinder, 2010), Diptera (Cranston, 2019; Debenham, 1987; Elson-Harris, 1990; Madden, 2009), Coleoptera (Davis, 1998; Glaister, 1999; Porch and Perkins, 2010; Watts, 2002), Gastropoda (Ponder, 2013), Hemiptera (Andersen and Weir, 2004; Porch and Perkins, 2010), Hirudinida (Govedich, 2001), Odonata (Theischinger and Hawking, 1999; Theischinger, 2000, 2001; Theischinger and Endersby, 2009), Trichoptera (Dean et al., 2004).

2.2. DNA extraction and amplification

Extraction of DNA was undertaken on bulk samples (i.e. all macroinvertebrates collected in a sample were processed together). DNA extraction was undertaken in a laminar flow cabinet, with UV serialisation, to avoid contamination issues. Prior to extraction, ethanol was drained from and then evaporated off the samples by placing the open samples on a heating block at 70 °C. Genetic material was extracted using a DNeasy blood and tissue kit (Qiagen, USA) with the following modifications to the manufacturer's guidelines: samples were placed in a solution of 20 µl Proteinase K (20 mg/mL) and 200–400 µl of buffer and bead beaten for 5 min using a Mini-bead beater (Daintree scientific), followed by manual crushing using a pestle. Tubes and beads used for bead beating were from the MoBio Power Water DNA extraction kit. From each sample, 200 µl of material was taken and used as the sample from which genetic material would be extracted and the remaining steps in the DNeasy blood and tissue protocol were adhered to. A blank sample was prepared, alongside the samples, as a control and underwent all steps that non-blank samples underwent.

A 313 base-pair (bp) internal region of the COI barcode region was amplified using the primer pair mlCOIintF (GGWACWGGWTGAACWGTWTAYCCYCC) and jgHCO2198 (TAIACYTCIGGRTGICCRAARAAYCA) (Leray et al., 2013) with 8 bp index barcodes attached. Amplification and sequencing were done by the sequence provider Mr DNA (www.mrdnalab.com, Shallowater, TX, USA), during June 2018, using the provider's standard protocols. Duplicate, one-step PCRs were undertaken using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) with a protocol of 94 °C for 3 min; 30 cycles of 94 °C for 30 s, 53 °C for 40 s and 72 °C for 1 min; and a final elongation step at 72 °C for 5 min. Successful amplification was assessed through checking on a 2% agarose gel. Multiple samples were pooled in equal proportions, based on molecular weight, and purified using Ampure XP beads. An Illumina DNA library was prepared from the pooled samples. Sequencing was performed on an Illumina MiSeq sequencer using V2 300 cycle kit.

2.3. Taxonomic assignment

Sequence data were demultiplexed using a custom built script provided by Mr DNA; FASTq Processor (http://www.mrdnafreesoftware.com/). Processing and cleaning of data, and creation of OTU tables was performed using the Greenfield Hybrid Analysis Pipeline (GHAP) (Greenfield, 2017). In summary, sequences with a minimum overlap of 25 bp and homology of at least 80% were merged. Sequences were quality filtered using a Maximum Expected Error (max_EE) threshold of 1. Only sequences that were between 304 and 350 bp long were retained for further analyses. Sequences were clustered into operational taxonomic units (OTUs) using a 97% clustering threshold and OTUs that occurred over less than three samples or consisted of fewer than three reads were filtered out. To test the effect that filtering OTUs based on read numbers has on downstream analyses, OTUs were filtered at 6 thresholds where read abundance was greater or equal to 0%, 0.025%, 0.075%, 0.01%, 0.05% and 0.1%.

Taxa were assigned to OTUs using three reference libraries of COI barcodes, resulting in three sequence datasets. The first was a library of curated barcodes obtained from GenBank (Benson et al., 2012) (https://www.ncbi.nlm.nih.gov/genbank/ accessed May 2018), which contains data from species across the world. The second was the Aquatic Invertebrates of Australia reference library (AIA), housed on the Barcode of Life Database (BOLD) (Ratnasingham and Hebert, 2007) (http://www.boldsystems.org/ accessed July 2018), which contains only data from Australian macroinvertebrate species and many

sequences from species that were not within the GenBank library. The third was a combination of the two libraries, herein referred to as the Best of Both Worlds (BoBW) library. Taxonomic assignment for BoBW was achieved by taking the highest percent identity match from either the GenBank or AIA datasets. Taxonomic assignment was further filtered using the arbitrary default thresholds in the GHAP pipeline which assign OTUs at various taxonomic levels depending on percent homology: 97% or greater for species, 95 - <97% for genus, 90- <95% for family, 85- <90% for order and <85% are unassigned. For analyses involving genus or family level data, OTUs with identical taxonomic assignments were merged and their read numbers summed. Non-target taxa, such as fish, fungi and microinvertebrates were removed from the data.

2.4. Analyses

SIGNAL grades provide an indication of how tolerant taxa are to pollution and SIGNAL scores, calculated as the un-weighted (presence absence) or weighted (with square root transformed abundance data) average of grades within a sample, are often used to determine the health of river systems. Two methods were used to assign SIGNAL grades to taxa. The first applied a genus-level grade (SIGNAL_SG) based on Chessman et al. (2007) regional version of SIGNAL for Sydney, Australia. For the genus-level analyses OTUs that could not be identified to genus were removed. The second method applied SIGNAL2 grades as per Chessman (2003), which were developed from Australia wide sampling and as such are the most commonly used in Australia. Each family was assigned a family-level SIGNAL2 grade except organisms belonging to Oligochaeta and Acarina were assigned a single grade, respectively, and members of the Chironomidae were assigned grades at the sub-family level. OTUs that could not be identified to the level required for their respective SIGNAL2 grade were removed. Both indices were created with the same approach of setting sensitivity grades of the taxa objectively (Chessman, 2003; Chessman et al., 2007). For each reference library, the total number of taxa that could be assigned SIGNAL grades and the average SIGNAL grades of those taxa were calculated to investigate possible biases due to reference library composition.

For each sample an un-weighted SIGNAL score was calculated from both the morphological and molecular datasets. An un-weighted SIGNAL score was chosen because treating abundances of DNA reads as abundances of organisms suffers from numerous problems that have not yet been adequately resolved, such as primer biases (Elbrecht and Leese, 2015) and differential mitochondrial or cell numbers (Elbrecht et al., 2017a). Pawlowski et al. (2018) state that there is no simple solution to address the abundance issue and advocated the most conservative approach is to use only presence-absence data. Moreover, weighted SIGNAL2 scores are generally calculated using predefined bins of taxa counts (e.g. 1–2, 3–5 organisms) and it is not clear how these would correspond with read numbers. Moreover, other studies suggest that when abundance data are swapped for presence/absence data differences in biotic indices are generally low (Beentjes et al., 2018; Buchner et al., 2019). Correlations between morphologically and genetically determined SIGNAL scores were assessed using Pearson's Correlation Coefficient (PCC) tests. SIGNAL scores are often used to classify river reaches in terms of severity of pollution. Classifications were applied using those provided in Chessman (1995). Confusion matrices of paired morphological and molecular classifications were created at the generic and family level.

Numbers of taxa were similarly treated, with comparisons made between the numbers of unique taxa in the morphological data and the numbers of OTUs and numbers of unique taxa in the molecular data. The set of unique OTUs within the molecular data contained taxa not traditionally targeted in macroinvertebrate monitoring (e.g. fish, microcrustacea and fungi). A comparison using all OTUs was undertaken to investigate whether richness metrics can be reasonably used without identifying OTUs to taxa. Two further comparisons of the number of taxa between morphological and molecular datasets were undertaken using 1) all unique taxa with taxonomic identification taken to genus level where possible, and 2) only those taxa to which SIGNAL scores could be applied. The reasoning for the latter being that this better represents what would be collected in a traditional survey.

Differences in taxonomic assignment between the two methods were assessed by examining taxa lists. Metrics were calculated to investigate the detectability of taxa, including accuracy, precision, prevalence, and true positive, true negative, false positive and misidentification rates. An F_1 score, which provides a harmonic mean between the precision and true positive rate, was also calculated for each family and genus to further aid in assessing how well the molecular methods performed at detecting the presence or absence of taxa. How these metrics change in response to filtering thresholds was also investigated using a single well-performing dataset. It should be noted here that the assumption is that the morphological identification is correct. However, in reality, a false positive does not necessarily mean a taxon was not present in the sample, as it is possible taxa may have been missed or mis-identified in the original morphological assessment. For instance, Chessman et al. (2007) measured genus-level taxonomic disagreement of 4.2% and difference in enumeration of 0.05%, for 94 samples selected randomly and reprocessed by a person who had not done the original identification and counting. These 94 samples were drawn from 2740 samples that were the basis of the Sydney regional version of SIGNAL_SG (Chessman et al., 2007). Moreover, during the time at which the current samples were identified the identification error for the laboratory processing the samples was 3.9%, following similar methods as those of Chessman et al. (2007).

3. Results

3.1. Data processing

An r-markdown file of the R script and details of analyses is provided as a supplementary html file (S2) and r-markdown script (S3) along with the raw data used in the script (S4.1-S4.8). DNA sequencing resulted in over 3 million reads. The minimum number of reads within samples was 42,072 and the maximum was 141,085. Filtering samples across read number thresholds and by removing non-target taxa, had little effect on reducing the number of reads per sample across the three datasets (Fig. 2). Analyses at taxonomic levels (i.e. genus and family) required filtering out OTUs that could not be identified to the required level. For the AIA dataset there were no reads removed between filtering on OTU read number and filtering non-target taxa, as the reference library contained only target taxa, thus OTUs that represent non-target taxa remained in the dataset until genus or family level filtering was applied. On average, this filtering resulted in a greater loss in sequences for the GenBank dataset than either the AIA or BoBW datasets (Fig. 2). An exception occurred in the AIA dataset for one sample that, in the generic level analyses, ranged from 5025–5318 depending on the read threshold applied, which was similar to samples with lowest read numbers in the GenBank database. Mean read numbers for family level analyses ranged from 75,774–76,393 for the BoBW dataset, 67,089–67,577 in the AIA dataset and 42,415–42,756 in the GenBank dataset. Mean read numbers for generic level analyses ranged from 67,823–68,213 for the BoBW dataset, 60,413–60,701 in the AIA dataset and 37,856–38,067 in the GenBank dataset.

3.2. SIGNAL grades and unique taxa

The numbers of OTUs that could be assigned SIGNAL_SG and SIGNAL2 grades and that contributed to the total number of unique taxa (at family or generic levels) varied between datasets (Fig. 3). The average SIGNAL grades of taxa within the morphological dataset best matched the average grades of taxa in the BoBW dataset. At the family level, the average SIGNAL2 grades of taxa within all datasets fell within

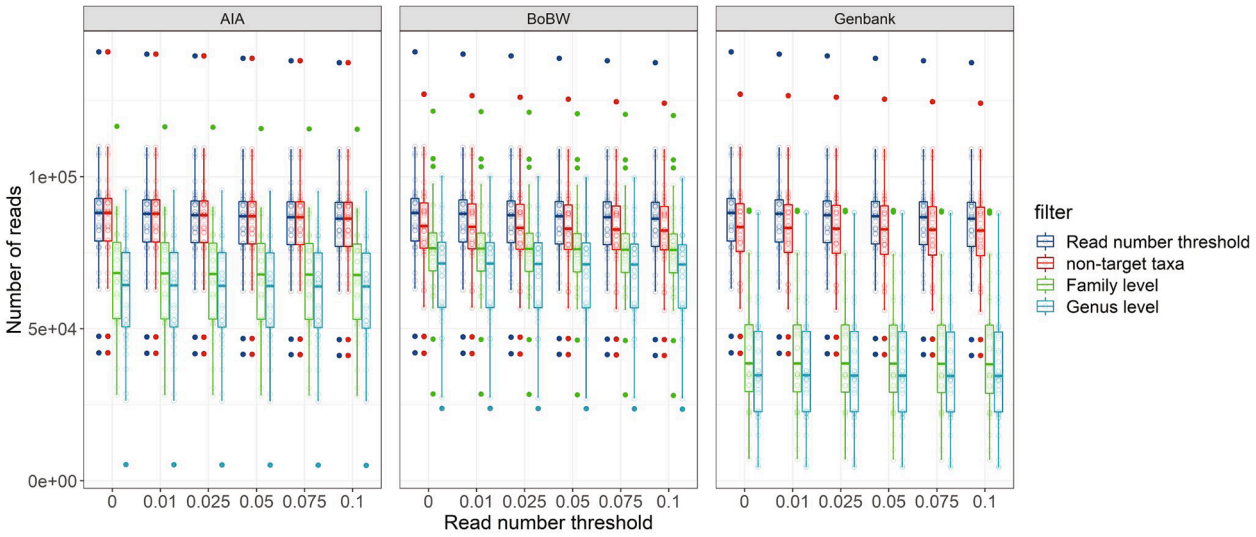

Fig. 2. Effect on total read number within samples from filtering out OTUs below a read number threshold (dark blue), non-target taxa (red), OTUs that could not be identified to the appropriate level for family (green) and generic (light blue) level analyses. Note the read number threshold of 0 represents no filtering having been applied and is thus the original total number of reads. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

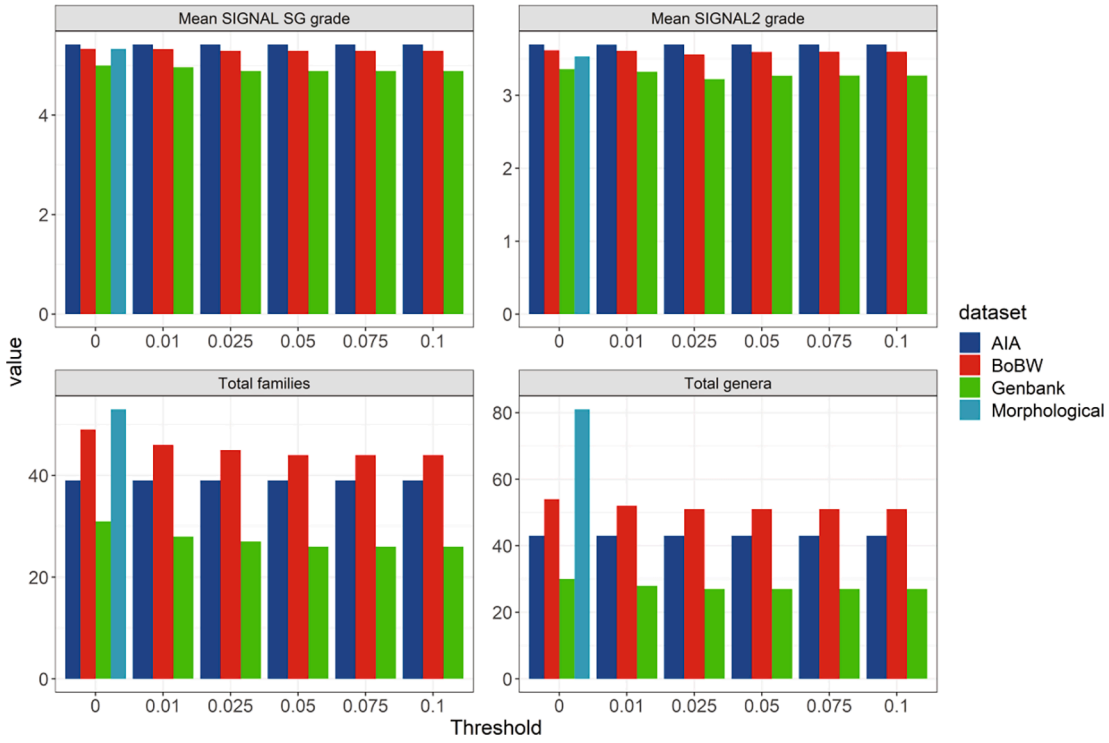

Fig. 3. Four metrics of mean SIGNAL_SG grade, mean SIGNAL2 grade, total number of unique taxa at family or generic levels in three eDNA datasets and a morphological dataset (colours), and the effect of applying read number threshold filters (x axis) on these four metrics.

the 3–4 band. At the genus level, the average SIGNAL_SG grade of taxa in both the BoBW and AIA datasets fell within the same band as the morphological dataset (between 5 and 6), with AIA average grades marginally higher than BoBW. In contrast, taxa in the GenBank dataset had an average SIGNAL_SG grade in the 4–5 band.

While the total numbers of families or genera that could be assigned SIGNAL grades was lower in the molecular datasets than the morphological dataset, the BoBW had the greatest number, followed by AIA and then GenBank. The morphological dataset had around 30 more genera that the BoBW dataset. Applying increasingly more stringent read number threshold filters had little effect on average SIGNAL grades or numbers of unique taxa.

.

3.3. Correlations of molecular and morphological SIGNAL_SG and SIGNAL2 scores

Molecularly derived SIGNAL_SG and SIGNAL2 scores of site samples across collection events were generally significantly and strongly correlated with morphologically derived scores as long as some read

number filtering had been applied; although, generic level GenBank scores were only moderately correlated (Fig. 4). Correlations in unfiltered datasets were weak and in one case non-significant. At the generic level the AIA dataset marginally outperformed the BoBW, with the strongest correlations seen in the comparisons that used 0.05percent threshold for the AIA dataset (PCC = 78.5 p-value < 0.001). At the family level, correlations with the GenBank database improved substantially from the generic level analyses and also with increased threshold filtering. While at the family level the BoBW dataset performed best across most thresholds, except at 0.1 percent threshold the GenBank database outperformed the BoBW and provided the highest correlation. In contrast the correlations for the AIA dataset grew markedly weaker with increasing threshold filtering when analysed at the family level.

Comparisons between morphological and molecular SIGNAL_SG and SIGNAL2 scores at each threshold are graphically illustrated in the supplementary r-markdown file, Figs. 2.3 and 2.11, respectively. When threshold filtering was applied, molecular-based scores mostly fell within the same unit of morphologically-based scores. Scores that fell outside the same unit tended to be only 1 unit either side of the morphological score, with GenBank tending to skew towards lower scores at both the generic and family level and BoBW and AIA tending to skew towards higher scores in the family level analyses (see supplementary r-markdown Figs. 2.4 and 2.10).

Confusion matrices of the classifications into which each dataset places river reaches showed a greater spread of discrepancy between morphological and molecular classifications at the generic level than the family level (see supplementary r-markdown file Figs. 2.5 and 2.13 respectively). At the generic level, the GenBank dataset was more likely to overestimate the severity of pollution, for instance 15 cases assessed by the morphological analyses as moderately polluted were classified as severely polluted by the GenBank analyses at 0.1% read number threshold. In contrast, the AIA and BoBW most often ascribed the same classification as in the morphological analyses, with few cases in disagreement. For instance, the AIA analysis with a 0.05% read number filter had only 6 samples that disagreed with the morphological classification and four of these either ascribing mild pollution to moderately polluted reaches or vice versa. The percentage of cases where the molecular classification agreed with the morphological classification at the generic level, when read number filtering greater than 0 was applied, was between 67.6 and 86.5% in the AIA analyses, 67.6–78.4 in the BoBW analysis and 35.1–48.6% in the GenBank analysis (Table 1). The family level analyses were in much greater agreement. However, all but one sample was morphologically categorised as severely polluted. The percentage of cases where the molecular classification agreed with the morphological classification at the family level was between 89.2 and 97.3% in the AIA analyses, 94.6–100 in the BoBW analysis and 97.3% across all read number thresholds in the GenBank analysis (Table 1).

Table 1
Percent agreement between morphological and molecular water classifications.

Dataset	Threshold	Generic level	Family level
AIA	0	37.84	97.30
	0.01	67.57	94.59
	0.025	70.27	97.30
	0.05	83.78	89.19
	0.075	81.08	89.19
	0.1	86.49	89.19
BoBW	0	45.95	97.30
	0.01	67.57	97.30
	0.025	75.68	100
	0.05	72.97	97.30
	0.075	78.38	94.60
	0.1	78.38	91.90
GenBank	0	59.46	97.30
	0.01	48.65	97.30
	0.025	43.24	97.30
	0.05	45.95	97.30
	0.075	40.54	97.30
	0.1	35.14	97.30

3.4. Correlations of molecular and morphological taxonomic richness

Correlations between taxa richness of morphological data and OTU or taxa richness of molecular data were investigated for each dataset and across all read number filters. Analyses were conducted using three levels of the molecular data 1) all OTUs, 2) all unique taxa with genus as the lowest taxonomic level, and 3) only those taxa for which SIGNAL grades could be applied.

In general, morphological taxa richness was weak to moderately correlated with molecular OTU and unique taxa richness measures. OTU richness returned significant PCCs between 0.34 and 0.44 when some level of read number thresholding was applied (Fig. 5). Molecular unique taxa richness correlated best with morphological taxa richness when using the AIA dataset, with PCCs over 0.5 when read numbers that contributed at least <0.025% were removed. At the 0.025 read number threshold the PCC reached 0.59 (p-value < 0.01). The BoBW and GenBank datasets correlated relatively poorly with morphological taxa richness, with PCCs increasing as read number thresholds increased. However, at the 0.1% read number threshold PCCs were 0.51 and 0.49 respectively, with p-values < 0.01. Including only those taxa for which SIGNAL grades could be applied (i.e. SIGNAL taxa richness) increased the performance of the BoBW dataset, and this dataset returned comparable although slightly lower PCC values than the AIA dataset (Fig. 5). In contrast, the GenBank PCC values increased and became significant at the 0% read number threshold, but became non-significant for thresholds between 0.01 and 0.075%. The greatest correlation over all analyses was between the morphological taxa richness and the AIA SIGNAL taxa richness at 0.01% read number threshold (PCC = 0.63, p-value <

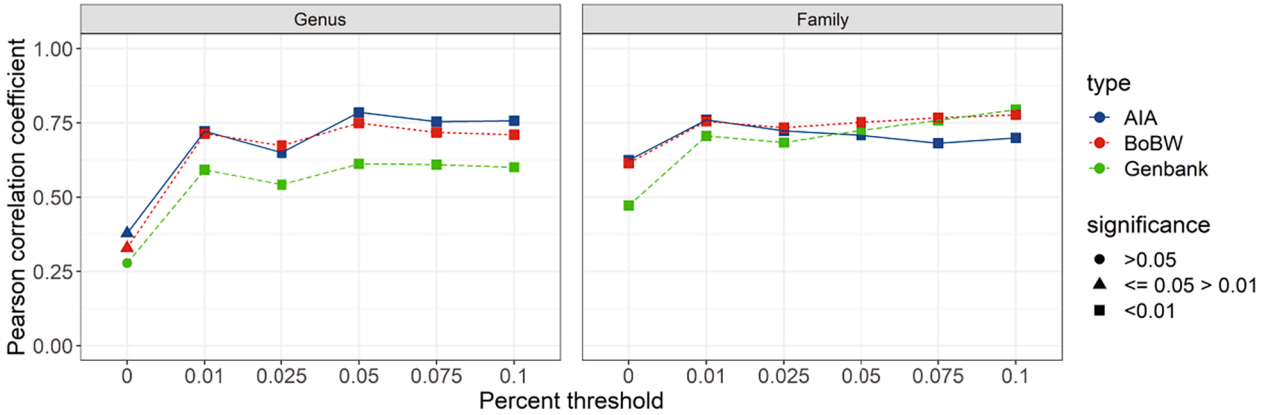

Fig. 4. Pearson correlation coefficient scores among three datasets (colour) and thresholds (x axis). Shapes of points indicate the degree of significance.

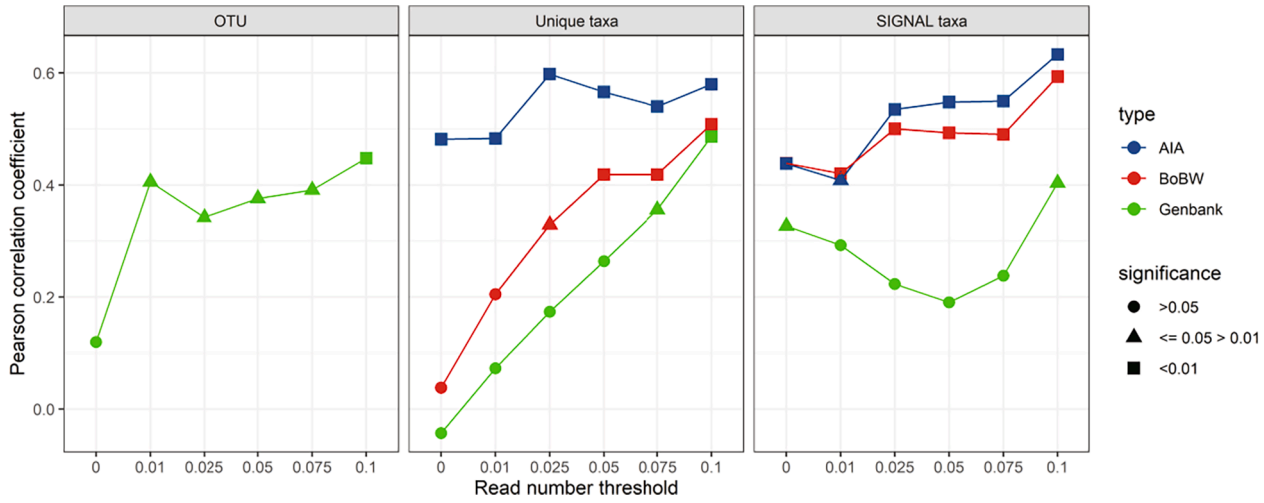

Fig. 5. Pearson's correlation coefficients between morphological taxa richness and richness of molecularly derived OTUs, taxa, and taxa for which SIGNAL grades could be applied. Colour represents datasets and shape of points provide the significance of the correlations. Note that at the OTU level all datasets were the same so here only the GenBank values are supplied.

0.01) (Fig. 5). Fig. 6 shows the scatter of data points between the morphological and AIA dataset at the 0.01% threshold. For brevity only these scatterplots are presented here. However, further scatterplots, including for analyses at all thresholds, are provided in the r-markdown supplementary material Figs. 2.25, 2.26 and 2.27. Similarly, Fig. 7 shows the relationship between morphological and molecular richness at sites over time for the AIA dataset with a read number threshold of 0.1%. Here, OTU richness is generally higher, taxa richness higher or lower and SIGNAL taxa richness generally lower than morphological richness.

3.5. Detectability

The ability of the molecular approach illustrated by the BoBW dataset to detect taxa, varied among genera (Fig. 8) and families (Fig. 9). Increasing the read number threshold reduced the number of taxa that occurred exclusively in the molecular data but also reduced the percentage of genera and families that were shared between the molecular (AIA, BoBW and GenBank) and morphological datasets (Fig. 10). Overall, the morphological dataset shared more taxa with the BoBW dataset than the AIA or GenBank datasets. Whereas the BoBW and AIA datasets had at least some samples with 100% of the families present in the morphological dataset, the GenBank dataset only included 100% of families when a 0% read number threshold was applied and never included 100% of the genera (Fig. 10). The BoBW dataset was the only dataset to have samples that contained 100% of the genera present in the morphological dataset (Fig. 10).

Metrics of detectability and prevalence were calculated across all thresholds and are provided as supplementary material (S5 for genera and S6 for families). These calculations were based on the assumption that the morphological identification was correct. For brevity, only those for the BoBW dataset using a read number threshold of 0.1% are provided here for genera (Fig. 11) and families (Fig. 12). Prevalence was highly variable among families and genera, with the majority of taxa having a prevalence below 50%. Prevalence below 10% occurred in 28 of the 59 families and 64 of 101 genera.

Overall, the accuracy of the molecular method was relatively high, with the notable exception of the flatworm Dugesiidae which had a true positive rate of 0% and very high misclassification rate (81.1%) (Fig. 11). However, the primers used in this project are known to not

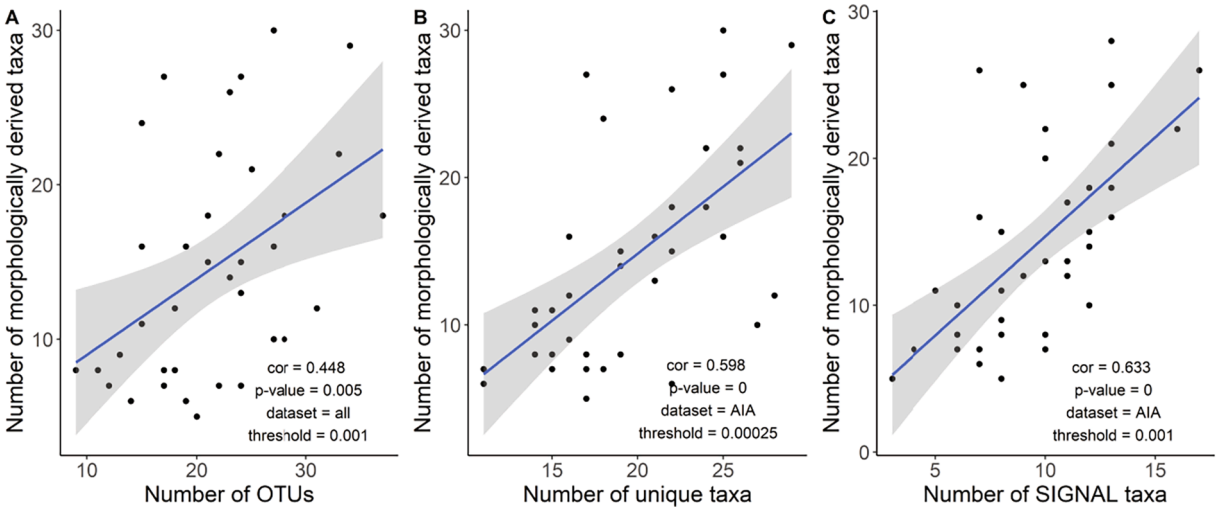

Fig. 6. Scatterplots of correlations between morphological taxa richness and the molecular dataset and threshold with the highest correlated richness measure for A) OTUs, B) unique taxa and C) SIGNAL taxa, at the 0.1% read number threshold. Blue line is a linear model of best fit with 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

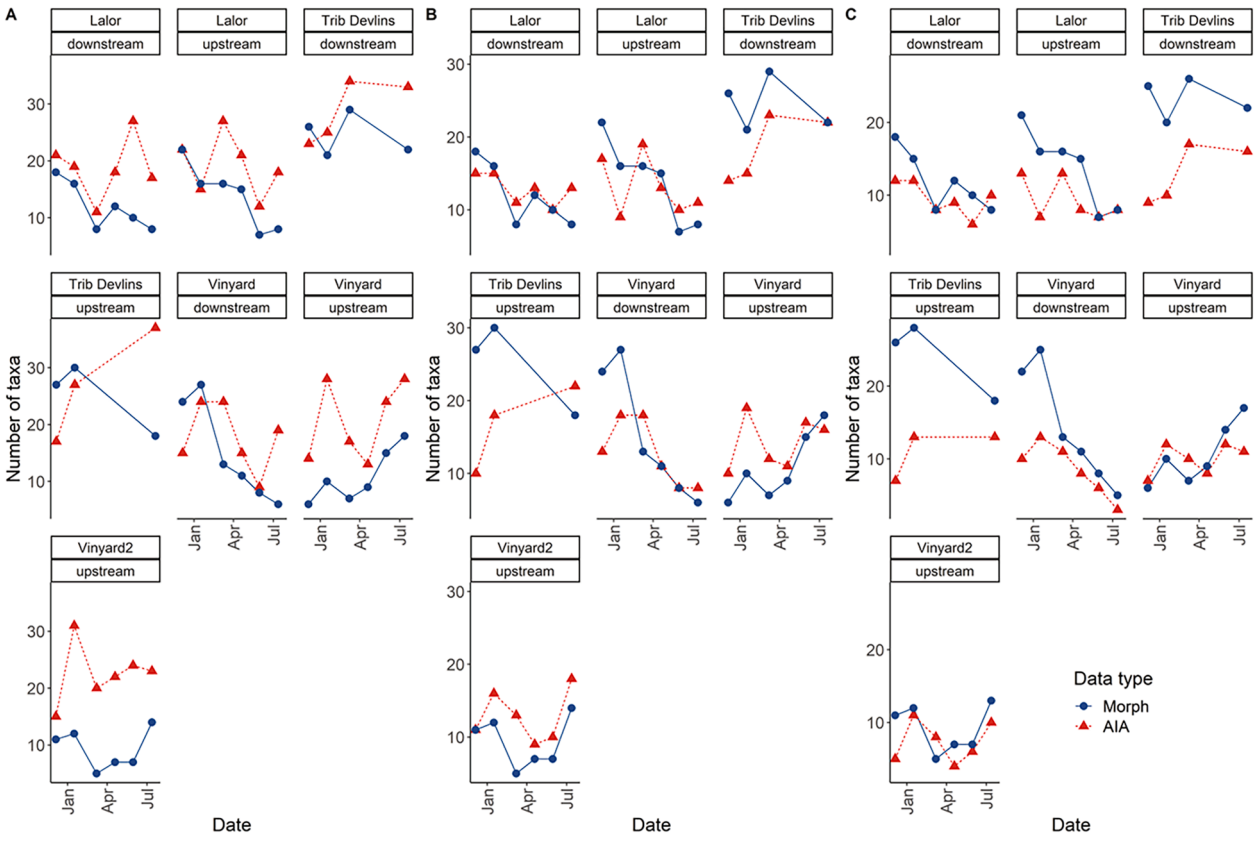

Fig. 7. Per site comparisons between number of taxa in the morphological (Morph) data and in the AIA molecular dataset with a read number threshold of 0.1% for the number of A) OTUs, B) unique taxa and C) taxa for which SIGNAL grades could be placed.

amplify flatworm DNA. The high accuracy values were, in most cases, largely driven by high true negative rates. To gain a better understanding of how the molecular methods performed at determining the presences or absences of taxa an F_1 score was used, which provides a balance between precision and true positive rate. The molecular analyses correctly determined the presences/absences for five families and six genera (i.e. F_1 scores = 1): Pionidae *Piona* (mite), Scirtidae (beetle), Psychodidae *Psychoda* (true fly), Simuliidae *Simulium* (true fly), and *Culicidae culex* (mosquito), and Philopotamidae *Chimarra* (Caddisfly) (Fig. 11). Generally, most families returned F_1 scores greater than 0.5. However, one or two families within most orders returned null F_1 scores. These were often families that occurred at low prevalence within the samples. However, some occurred at a similar prevalence to other families of their order; for instance, Synthemistidae and Gomphidae both had a prevalence of 0.08, but where the former returned a null F_1 score, the latter returned a score of 0.8 (Fig. 12).

Values of the detectability metrics changed with increasing read number threshold (Fig. 13). As read number thresholds are increased, there is a trade-off between decreasing the mean and increasing the spread of values in the true positive rate and increasing the mean and reducing the range of values in the true negative rate. However, all metrics improve with at least some degree of read number filtering while filtering beyond the 0.05% threshold only marginally improved these metrics.

4. Discussion

This study demonstrates that the river health SIGNAL biotic indices can be derived from bulk sample DNA data with results that are comparable to those derived through traditional morphological analyses. Strong and significant correlations between morphologically and molecularly derived SIGNAL scores were observed for both family (SIGNAL2) and genus-level (SIGNAL_SG) analyses (Fig. 4). However, the choice of DNA reference library and data pre-processing influences the significance and strength of correlations. In both generic and family level analyses, correlations greatly improved when at least some filtering of low contribution OTUs was performed (e.g. filtering those that contributed<0.01%). At the generic level, datasets with taxonomic identifications made using a reference database of local taxa (i.e. AIA) performed better than using the GenBank reference library, with the AIA dataset returning the highest Pearson Coefficient Correlation (PCC) of 0.785 (p-value < 0.001) when using a 0.05% read number threshold. At the family level, the correlation of the GenBank with the morphological SIGNAL2 scores was greatly improved and when applying a read number filter at the 0.1% threshold the GenBank dataset had the highest correlation of the molecular datasets (PCC = 0.795, p-value = <0.001). However, at lower thresholds the BoBW dataset performed marginally better. In contrast the performance of the AIA dataset decreased.

In practice, SIGNAL scores are interpreted as bands (water quality status classes) indicating gradients of pollution. Chessman (1995), who introduced the first signal score classified bands as greater than 6 = clean water, 5–6 possible mild pollution, 4–5 moderate pollution and <4 severe pollution. In the present study, when applying at least some degree of read number filtering, most molecularly derived scores were in agreement with the morphologically derived scores in terms of water quality classification. When molecular classifications deviated from morphological classifications they predominately classified to the next lower or higher classification; the notable exception being in the genus level analyses using the GenBank dataset which classified a few mildly polluted samples as severely polluted.

While application of the SIGNAL biotic indices can be used to assign water quality status classes, Besley and Chessman (2008) demonstrated graphical and statistical assessment of SIGNAL_SG scores based on morphological data collected from paired sites situated upstream and

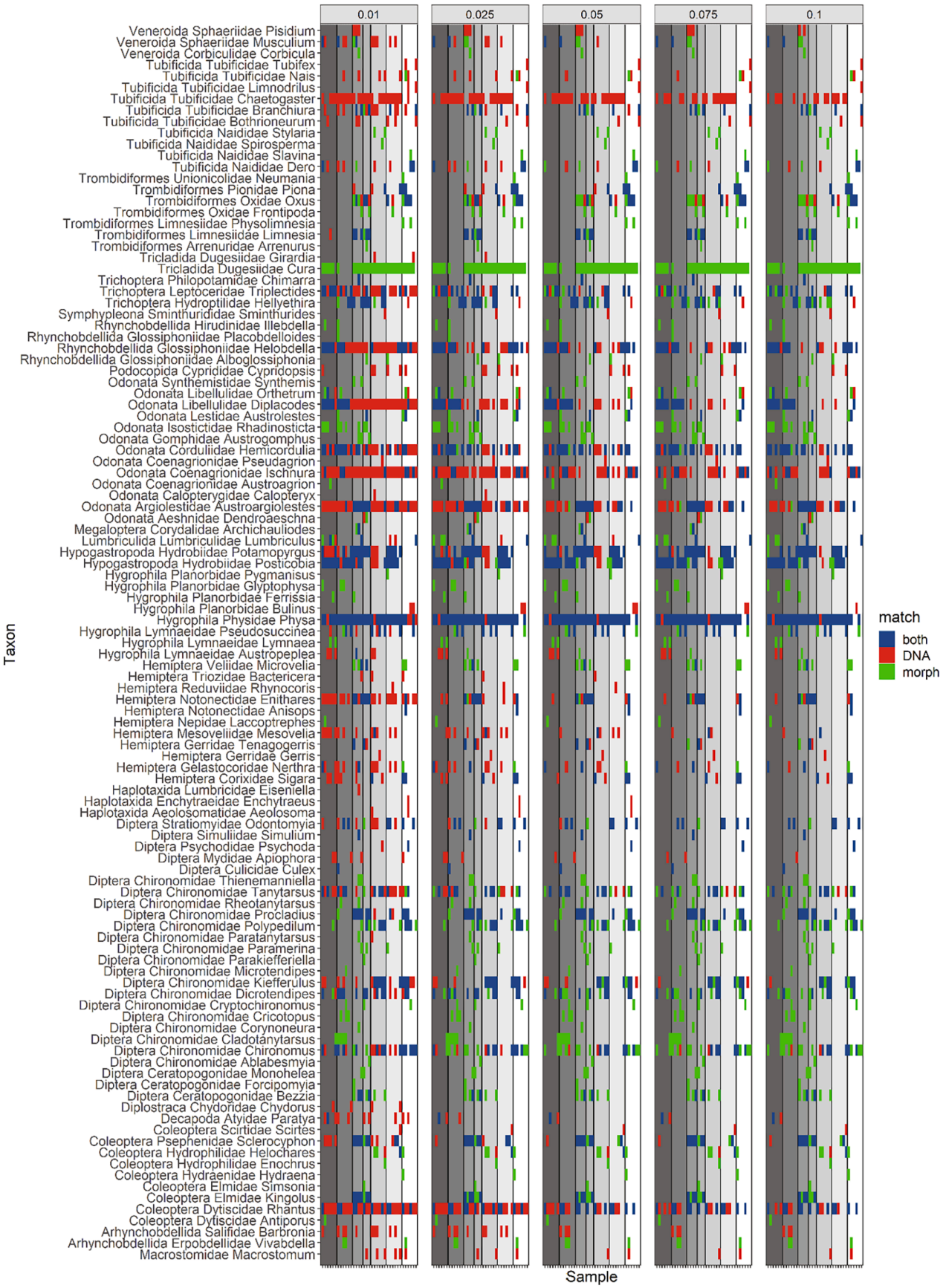

Fig. 8. Comparison of genera detected in samples within the morphological (morph) and BoBW datasets across read number thresholds (panels). Sample names have been removed for ease of plotting; however, grey divisions (x-axis) indicate different samples, which are arranged alphabetically and by date.

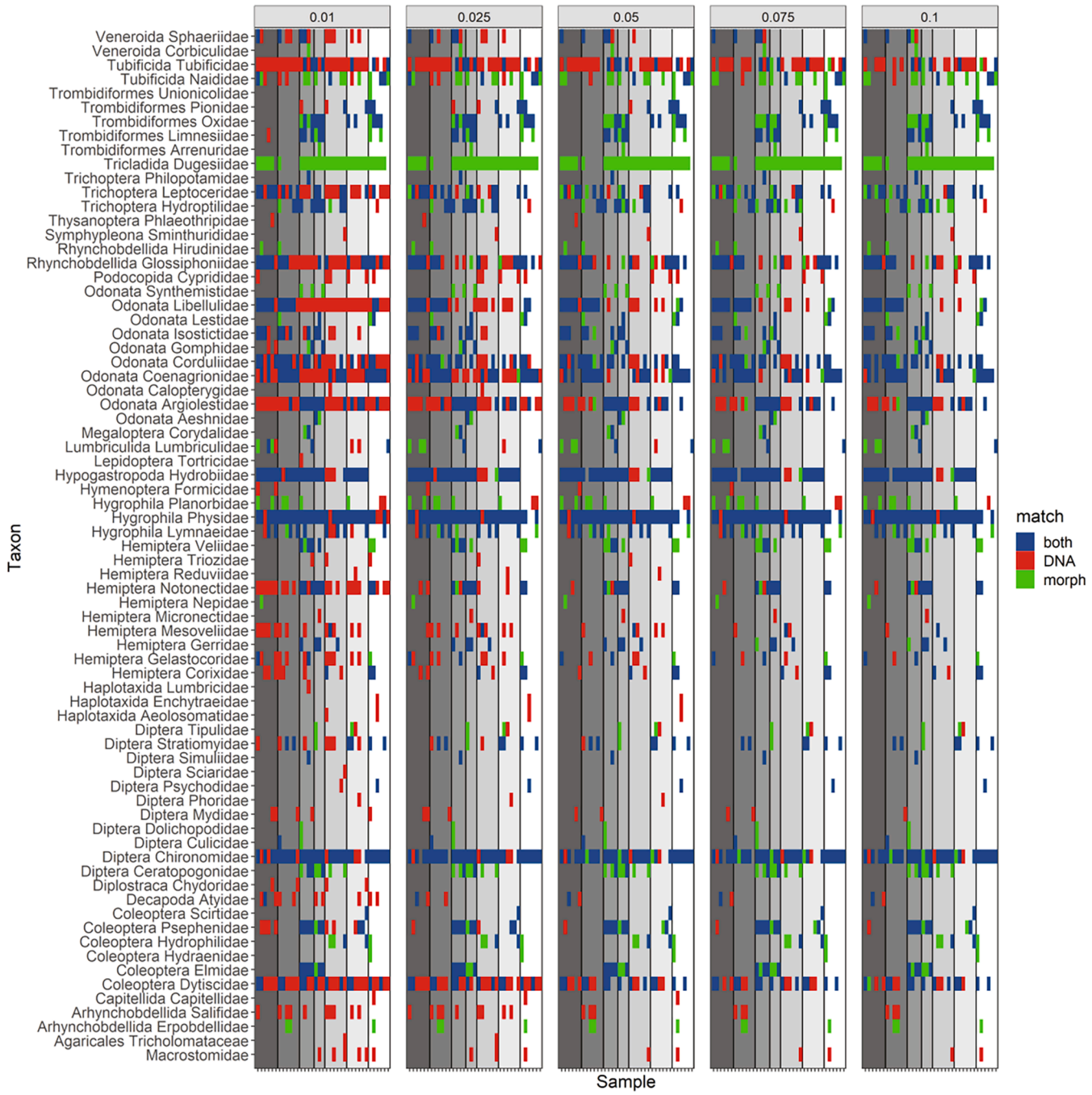

Fig. 9. Comparison of families detected in samples within the morphological (morph) and BoBW datasets across read number thresholds (panels). Sample names have been removed for ease of plotting; however, grey divisions (x-axis) indicate different samples, which are arranged alphabetically and by date.

downstream of point source discharge of treated sewage wastewater. That graphical assessment illustrated SIGNAL_SG scores do not neatly fall into a band, and often occur across two water quality status classes (bands). Addition of an overall upstream mean of SIGNAL_SG scores with error bars of ± one standard deviation for a temporal period allows presentation in a process control chart for ecological monitoring as advocated by Burgman et al. (2012). An example of this control chart approach is provided by the 25-year long-term study (1995 to 2020) of the Nepean River near the West Camden sewage treatment plant in the Sydney region, Australia, which illustrates the SIGNAL_SG range of morphologically derived scores of about a unit fluctuation as typical variation (see Fig. S1 in Supplementary material). In adopting metabarcoding data as the basis for assessment with biotic indices such as SIGNAL, our study suggests the underlying barcode library will influence slight differences in SIGNAL scores and a period where both morphological and metabarcoding data are obtained would provide an understanding of the potential site specific ranges. This conservative approach would consider Buchner et al. (2019) advocation of the importance of properly evaluating the potential to link metabarcoding data to established indices and relating them to existing data. This approach seems prudent as management decisions can be expensive, for example the Blue Mountains Sewage Transfer Scheme in the Sydney region, Australia, was established to upgrade the sewerage system at a cost of $AU 360 million, by progressively closing small, local plants and diverting sewage to a larger, more efficient plant (Besley and Chessman, 2008).

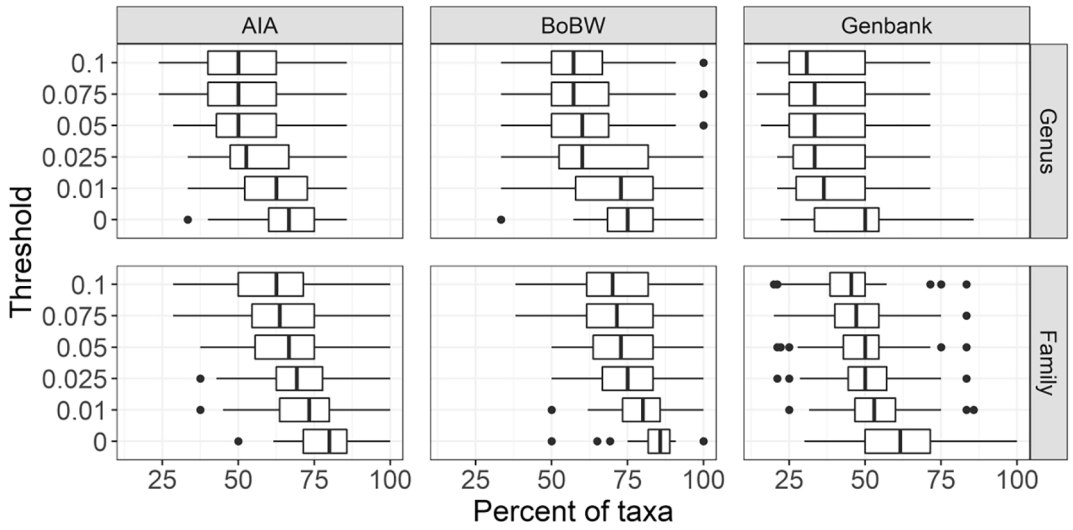

Fig. 10. Boxplots of the percent of genera and families shared between the morphological and molecular datasets and how these percentages change with read number threshold.

Overall, molecular classification was relatively accurate with accuracies over 80% at the generic level and over 95% at the family level obtainable. At these levels, managers of river health could apply the molecular techniques described here with some confidence that their results will be relatively consistent with those of traditional methods. However, a limitation to the current study is that it lacked samples classified as mildly polluted and had only one sample classified as moderately polluted at the family level. Further experimentation on a wider variety of streams with less surrounding urbanisation would provide greater insight into how family level classifications perform outside the category of severely polluted.

The morphologically derived taxa richness metric was generally significantly, but only weak to moderately, correlated with the molecularly derived metrics of richness. At the OTU level, correlations were significant and exceeded PCC values of 0.3 only when some degree of read number thresholding was applied. Correlations improved for measures of unique taxa but only when using the AIA dataset. However, correlations for the GenBank and BoBW datasets increased with increasing read number threshold, with the highest and significant correlations occurring at the 0.1% read number threshold. When only taxa for which SIGNAL grades could be applied were included in analyses, correlations became moderately strong in both the AIA and BoBW datasets, with PCC values ranging from 0.44 at the 0% and 0.63 at the 0.1% read number threshold for the AIA dataset and 0.44 at the 0% and 0.59 at the 0.1% read number threshold for the BoBW dataset. In contrast, the GenBank analyses were only significant at the 0% and 0.1% read number thresholds. The improvement in correlations from the OTU level analyses to the unique taxa and SIGNAL taxa analyses, is likely partly due to the removal of sequences of non-target taxa such as fish, microinvertebrates or terrestrial organisms, thus better representing the suite of taxa collected in a traditional sample. Moreover, the OTU clustering method used a 97% threshold which perhaps better represents delineations of species rather than genera, as in the morphological data (Hebert et al., 2003). Fig. 7 shows a trend where, compared with morphological richness, OTU richness is generally higher, taxa richness is either higher or lower and SIGNAL taxa richness is generally lower. This trend is driven by the distinct clustering of OTUs into taxa and then the filtering out of taxa to which SIGNAL grades could not be applied. However, it should be noted that SIGNAL taxa richness was generally lower than morphological taxa richness because not all the taxa in the morphological dataset have been barcoded, and thus are missing from the SIGNAL taxa richness analysis.

The performance of the genetic data to determine the presence or absence of genera was assessed using F_1 scores derived from the BoBW dataset with 0.1% read number threshold. Around 63% of families and 43% of genera had F_1 scores above 0.5; and 42% of families and 28% of genera had F_1 scores over 0.7. Within orders there were usually one or two families that returned null F_1 scores, with these predominantly being at low prevalence among samples. The Odonata were the most family diverse taxa and performed relatively well in terms of F_1 scores, with scores ranging from 0.67 to 0.92. The notable exception in the Odonata was the family Synthemistidae, which returned a null F_1 score despite being as prevalent as Gomphidae. The family Planorbidae was a similar notable exception among the snails, having a prevalence of 0.16 but a null F_1 score. For the Dugesiidae, the null F_1 score can be explained by the choice of primers used for this study, as they are known not to amplify this taxon (Vanhove et al., 2013; Carew et al., 2018). For other families with related taxa that did return F_1 scores it is less clear as to why they did not. It is possible that genetic variation could have been inadequate to distinguish between some taxa, and that taxa were misidentified to sister taxa, due to the small size of the barcodes used. This occurs usually when large reference libraries are used for identification (Hajibabaei et al., 2006). However, mini-barcodes are frequently used for metabarcoding analyses and lengths of 100 and 250 bp have been shown to distinguish around 90 and 95% of species, respectively (Meusnier et al., 2008; Yeo et al., 2020). The use of references from local species and a barcode length of 313 bp in this study should also have improved performance. Moreover, a cursory investigation into the variation in the gene fragment used among genera suggests that all genera should be reasonably distinct.

In one case, the null F_1 score can be attributed to an error in the reference library. The two mite genera *Oxus* and *Frontipoda* were morphologically recorded among the samples but were never found within the same sample. *Frontipoda* did not occur in the molecular data; however, suspiciously, *Oxus* was recorded in all of the samples that contained *Frontipoda*. A review of the *Oxus* voucher specimens that contributed to the AIA barcode library, revealed, incorrect identifications placed onto some of the *Oxus* specimens, specifically those that genetically matched the misidentified *Frontipoda*, and that the vouchers were in fact *Frontipoda*. These voucher specimens were the same that genetically matched to the misidentified *Frontipoda* in the molecular analyses. This highlights the issue of incorrectly assigned taxonomy in DNA databases and the need for adequately curated reference libraries, as has been emphasized by other authors (Nilsson et al., 2006; Tixier et al., 2012; Shen et al., 2013; Shackleton and Rees, 2016; Carew et al., 2017).

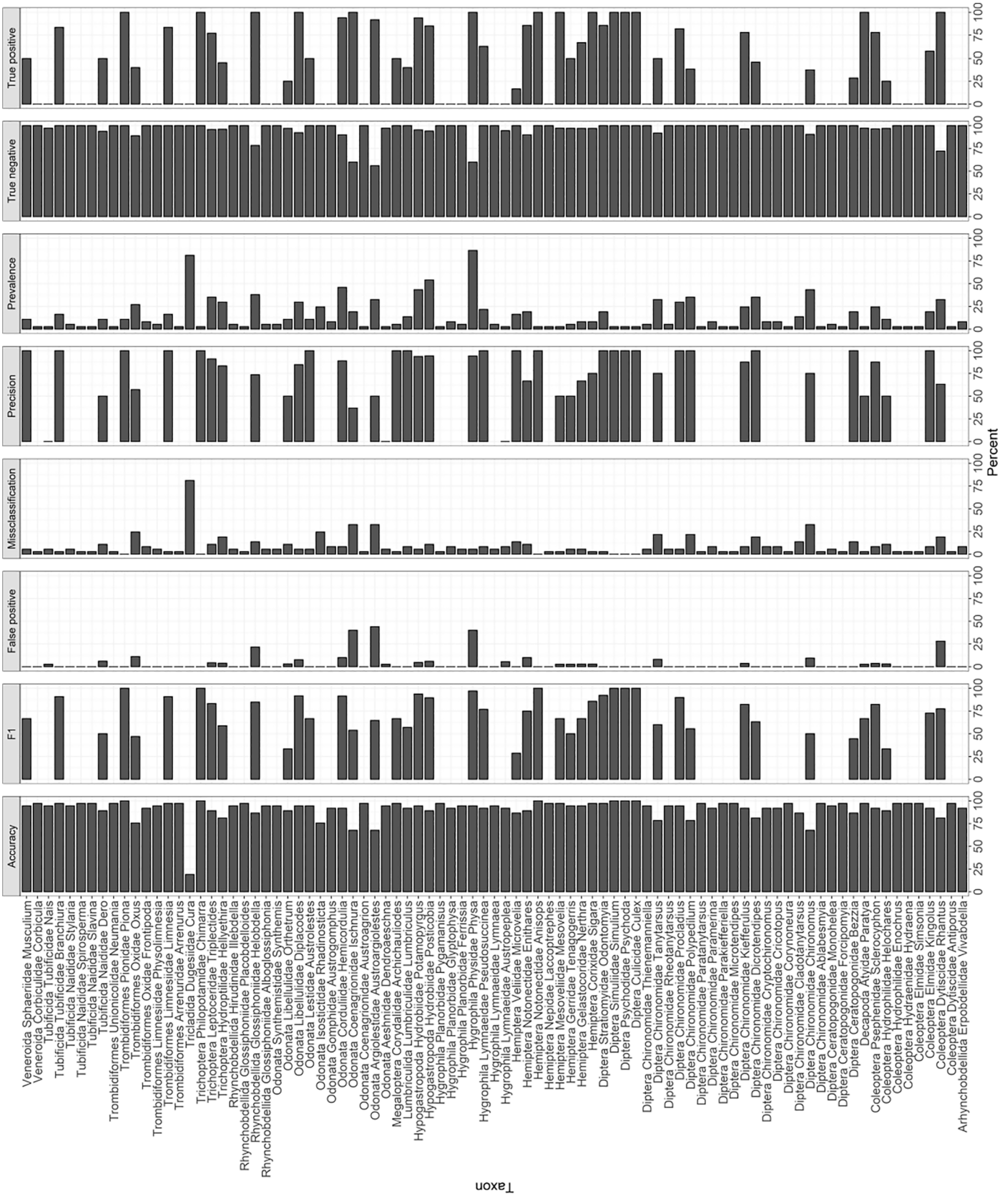

Fig. 11. Metrics of detectability and prevalence for genera in the BoBW dataset with a read number threshold filter of 0.1%.

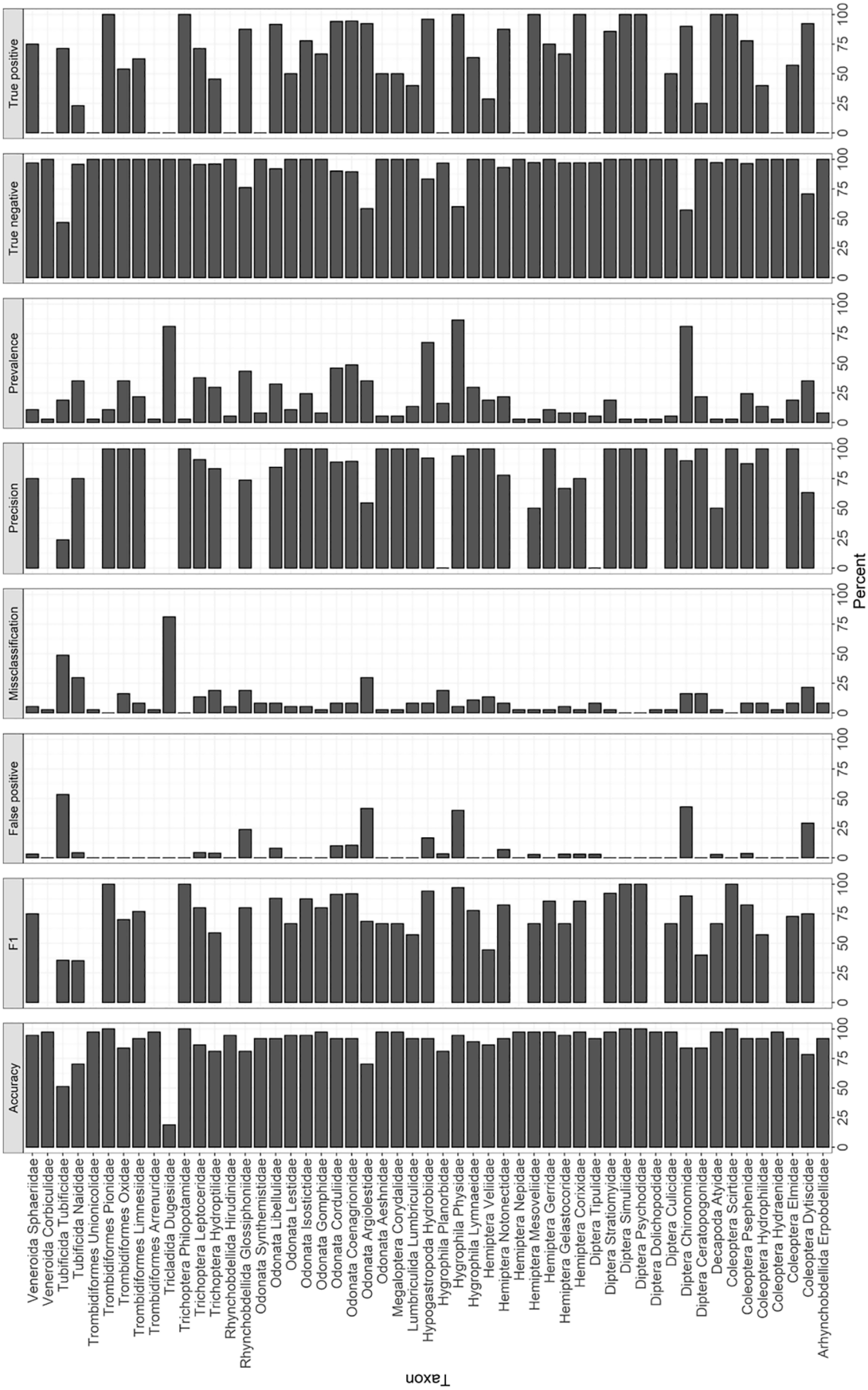

Fig. 12. Metrics of detectability and prevalence for families in the BoBW dataset with a read number threshold filter of 0.1%.

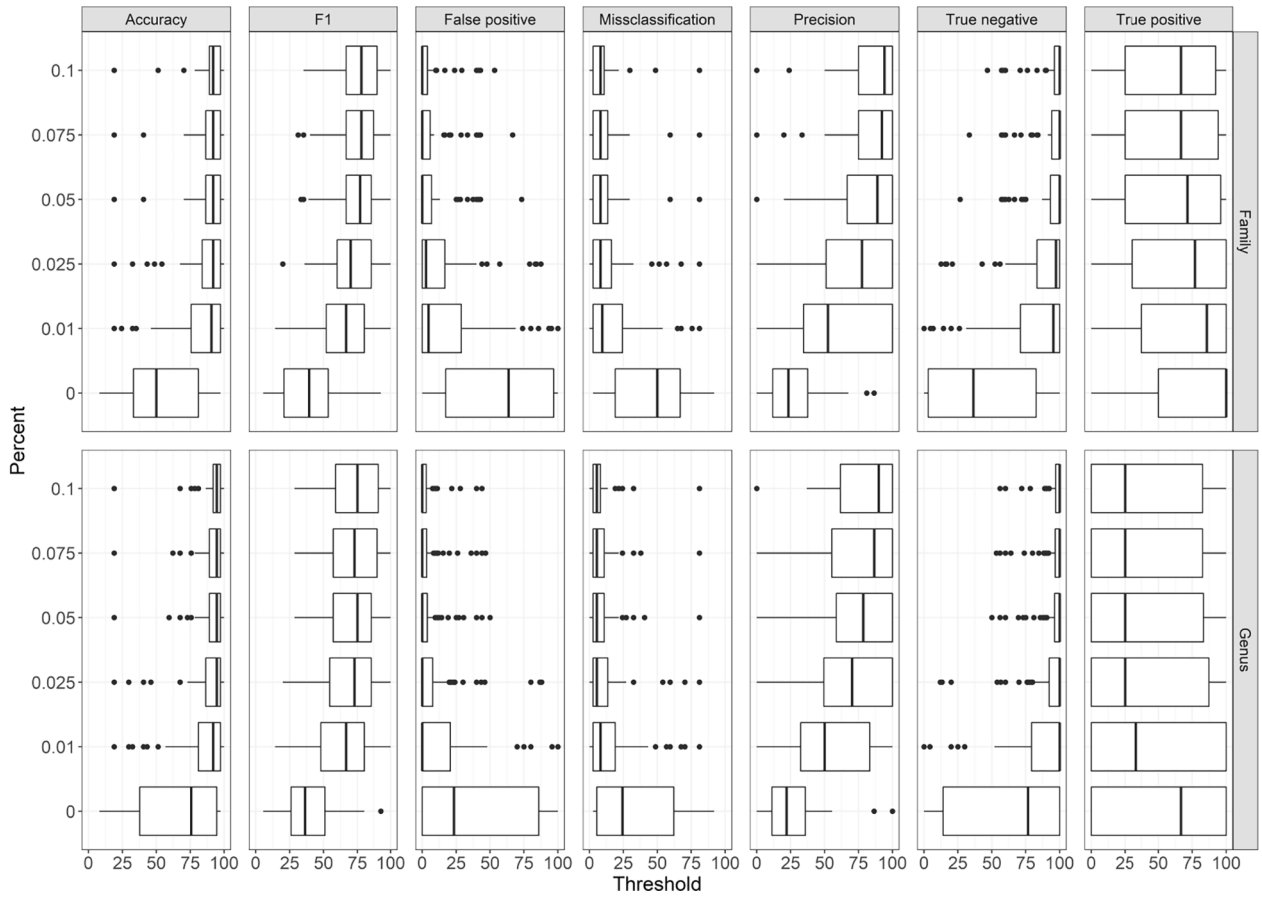

Fig. 13. Changes in detectability metrics over read number thresholds for family and generic level analyses.

In general, this study attributed discrepancies between morphological and molecular taxonomic assignments to be errors in molecular assignment. Because the samples used were destroyed during the genetic extraction process, morphological identifications could not be double checked. However, errors in morphological identification are probable and not unexpected. An error rate of about 4% was documented for the laboratory that performed morphological identification and this corresponds with the rate found by Chessman et al. (2007) for genus level analyses on taxa in the same region.

One possibility for unexpected positive results is that trace or environmental DNA (eDNA) may have been present in the samples (Beermann et al., 2020), including for species that may have been present as dietary components of the collected specimens (Zaidi et al., 1999; Sheppard et al., 2005; Hosseini et al., 2008). A further possibility, as recorded for *Oxus* above, is that specimens have been miss-identified in the reference databases used and thus the best matches for sequence data are to incorrectly assigned species. Pawlowski et al. (2018) suggested the most cited explanation for discrepancies between molecular and morphological datasets is the incompleteness and lack of accuracy of the molecular reference databases that impedes the correct taxonomic assignment of DNA sequences. Elbrecht et al. (2017b) describes databases like BOLD (where AIA is hosted) as containing misidentified taxa or conflicting taxonomic assignments for the same Barcode Index Number. Some of the cases of false negatives were due to taxa missing from the reference databases, but these were limited to *Physolimnesia* (mite), *Illebdella*, *Vivabdella* (leeches), *Synthemis* (Dragonfly) and *Pygmanisus* (snail).

Despite the differences between morphologically and genetically detected taxa, SIGNAL scores (for both index versions tested) were generally comparable between data types. This is likely to be primarily driven by phylogenetic signal to pollution tolerance occurring within taxa, with closely related taxa likely to share similar SIGNAL grades. For instance, Carew et al. (2011) showed phylogenetic signal within the Ephemeroptera and Chironomidae to tolerance to organic pollution and zinc concentration, respectively. In the presence of phylogenetic signal, it follows that if sequences are assigned to incorrect but closely related taxa the effect on SIGNAL grade assignment would be minimal. This suggests that while the molecular data used in this study may not have provided 100% detection rate for certain taxa, reliable SIGNAL scores can still be derived.

Our findings support those of Carew et al. (2018) who found little difference between DNA derived and morphologically derived family level indices: SIGNAL2, AusRivAS (Reynoldson et al., 1997) and a Chironomidae-based pollution index developed as part of their study. However, unlike Carew et al. (2018), the present study used a single primer pair, reducing the time and costs involved in processing samples and increasing the available sequencing depth per sample. This does, however, come at the cost of increasing the number of undetected taxa due to primer bias and primer-template mismatches. While Carew et al. (2018) were able to recover 85% of families known to be in their samples, in the present study the average number of families recovered ranged from 70.3% to 84.3% in the BoBW dataset with a range from 38.1% to 100% of families known to occur in the samples were recovered. It should be noted that many of the species used in the present study have since been added to GenBank, and it is thus possible that these percentages will increase if analyses are performed on updated libraries. However, our work has also shown that databases with local taxa maybe more important, thus GenBank identifications are unlikely to improve in regions where local data is depauperate.

While the full value of using DNA as a way to track river health lies in its ability to monitor species-level changes in communities, species-level metrics are yet to be developed (Nichols et al., 2020). Buchner et al.

(2019) suggest the central incentive for including genetic data in assessment of ecological status should be the fundamental improvement of resolution down to species or even population level that can be obtained in a standardised fashion. For biotic indices, such as SIGNAL, this will require compiling species responses to environmental stressors. Additionally, biotic index grade values could be inferred through machine learning predictive models trained on metabarcoding data linked to associated pressure data (Pawlowski et al., 2018). Buchner et al. (2019) suggests using supervised machine learning based on direct comparisons of metabarcoding data and traditional morphological taxa. Hence, suitable data for this development could be obtained from bulk-sample DNA methods as they are applied to routine biomonitoring, especially as detectability of individual species improves. In the meantime, our results show how bulk-sample DNA derived data can be used as an alternative way to calculate family- and genus-level river health metrics with similar results to current practices. Buchner et al. (2019) indicated DNA metabarcoding provided high-resolution taxonomic data for the data sets required by any of the currently used EU Water Framework Directive assessment methods including those at the genus level. While still in its infancy, DNA metabarcoding shows promise as a cheaper, yet robust, alternative to traditional morphological methods for biological monitoring.

Funding

This work was supported by Sydney Water, Sydney Australia.

CRediT authorship contribution statement

M.E. Shackleton: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **K.A. Dafforn:** Conceptualization, Writing - original draft, Writing - review & editing. **N.P. Murphy:** Resources, Writing - original draft, Writing - review & editing. **P. Greenfield:** Software, Writing - original draft, Writing - review & editing. **M. Cassidy:** Funding acquisition, Writing - original draft, Writing - review & editing. **C.H. Besley:** Conceptualization, Funding acquisition, Resources, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Sydney Water for financial assistance and for support in the collection of macroinvertebrate samples and morphological laboratory work. We also thank Ms. Merran Griffith from Sydney Water for providing valuable feedback on the manuscript. We thank and are very grateful to two anonymous reviewers who provided valuable feedback and enhanced our manuscript greatly.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.107537.

References

Andersen, N.M., Weir, T.A., 2004. Australian Water Bugs. Their Biology and Identification (Hemiptera-Heteroptera, Gerromorpha & Nepomorpha) - Entomograph, vol. 14, 344 pages Apollo Books, CSIRO Publishing. ISBN 87-88757-78-1.

Beentjes, K.K., Speksnijder, A.G., Schilthuizen, M., Schaub, B.E., van der Hoorn, B.B., 2018. The influence of macroinvertebrate abundance on the assessment of freshwater quality in The Netherlands. Metabarcoding Metagenomics 2, 18. https://doi.org/10.3897/mbmg.2.26744.

Beermann, A.J., Werner, M.-T., Elbrecht, V., Zizka, V.M., Leese, F., 2020. DNA metabarcoding improves the detection of multiple stressor responses of stream invertebrates to increased salinity, fine sediment deposition and reduced flow velocity. Sci. Total Environ. 750.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2012. GenBank. Nucleic acids research 41:D36-D42.

Besley, C.H., Chessman, B.C., 2008. Rapid biological assessment charts the recovery of stream macroinvertebrate assemblages after sewage discharges cease. Ecol. Ind. 8, 625–638.

Bunn, S.E., 1995. Biological monitoring of water quality in Australia: workshop summary and future directions. Aust. J. Ecol. 20, 220–227.

Bunn, S.E., Davies, P.M., 2000 Biological processes in running waters and their implications for the assessment of ecological integrity. In: Jungwirth, M., Muhar, S., Schmutz, S. (Eds) Assessing the Ecological Integrity of Running Waters. Developments in Hydrobiology, vol. 149. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-4164-2_5.

Buss, D.F., Carlisle, D.M., Chon, TS. Culp, J. S. Harding, H. E. Keizer-Vlek, W. A. Robinson, S. Strachan, C. Thirion, and R. M. Hughes. 2015. Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environ. Monit. Assess. 187:4132. https://doi.org/10.1007/s10661-014-4132-8.

Buchner, D., Beermann, A.J., Laini, A., Rolauffs, P., Vitecek, S., Hering, D., Leese, F., 2019. Analysis of 13,312 benthic invertebrate samples from German streams reveals minor deviations in ecological status class between abundance and presence/absence data. PLoS ONE 14 (12), 1–18. https://doi.org/10.1371/journal.pone.0226547.

Burgman, M., Lowell, K., Woodgate, P., Jones, S., Richards, G., and Addison, P., 2012. An endpoint hierarchy and process control charts for ecological monitoring in (Eds) Lindenmayer, D., and Gibbons, P. Biodiversity Monitoring in Australia. CSIRO Publishing, Collingwood, Australia.

Carew, M., Nichols, S., Batovska, J., St Clair, R., Murphy, N., Blacket, M., Shackleton, M., 2017. A DNA barcode database of Australia's freshwater macroinvertebrate fauna. Mar. Freshw. Res. 68, 1788–1802.

Carew, M.E., Kellar, C.R., Pettigrove, V.J., Hoffmann, A.A., 2018. Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river? Ecol. Ind. 85, 440–450.

Carew, M.E., Miller, A.D., Hoffmann, A.A., 2011. Phylogenetic signals and ecotoxicological responses: potential implications for aquatic biomonitoring. Ecotoxicology 20, 595–606.

Chessman, B., Williams, S., Besley, C., 2007. Bioassessment of streams with macroinvertebrates: effect of sampled habitat and taxonomic resolution. J. North Am. Benthol. Soc. 26, 546–565.

Chessman, B.C., 1995. Rapid assessment of rivers using macroinvertebrates: a procedure based on habitat-specific sampling, family level identification and a biotic index. Aust. J. Ecol. 20, 122–129.

Chessman, B.C., 2003. New sensitivity grades for Australian river macroinvertebrates. Mar. Freshw. Res. 54, 95–103.

Chessman, B.C., Thurtell, L.A., Royal, M.J., 2006. Bioassessment in a harsh environment: a comparison of macroinvertebrate assemblages at reference and assessment sites in an Australian inland river system. Environ. Monit. Assess. 119, 303–330.

Chutter, F., 1972. An empirical biotic index of the quality of water in South African streams and rivers. Water Res. 6, 19-30.

Cook, D.R., 1974. Water mite genera and subgenera. Memoirs of the American Entomological Institute, 21: 1-860.

Cook, D.R., 1986. Water mites from Australia. Memoirs of the American Entomological Institute, 40, 1–568.

Cranston, P.S., 2019. Identification guide to genera of aquatic larval Chironomidae (Diptera) of Australia and New Zealand. Zootaxa 4706 (1), 071–102. https://doi.org/10.11646/zootaxa.4706.1.3.

Davies, P., Harris, J., Hillman, T., Walker, K., 2010. The sustainable rivers audit: assessing river ecosystem health in the Murray-Darling Basin, Australia. Mar. Freshw. Res. 61, 764–777.

Davis, J., 1998. A guide to the identification of larval Psephenidae water pennies (Insecta: Coleoptera). Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 17. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.

Dean, J., St Clair, R., Cartwright, D., 2004. Identification keys to Australian families and genera of caddis-fly larvae (Trichoptera). Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 50. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.

Debenham, M.L., 1987. The biting midge genus Forcipomyia (Diptera: Ceratopogonidae) in the Australasian region (exclusive of New Zealand). IV. The subgenera allied to Forcipomyia, s.s., and Lepidohelea, and the interrelationships and biogeography of the subgenera of Forcipomyi. Invertebrate Taxonomy 1 (6), 631–684.

Elbrecht, V., Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. PLoS ONE 10.

Elbrecht, V., Peinert, B., Leese, F., 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecol. Evol. 7, 6918–6926.

Elbrecht, V., Vamos, E.E., Meissner, K., Aroviita, J., Leese, F., 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods Ecol. Evol. 8, 1265–1275.

Elson-Harris, M.M., 1990. Keys to the immature stages of some Australian Ceratopogonidae (Diptera). J. Aust. Entomol. Soc. 29, 267–275. https://doi.org/10.1111/j.1440-6055.1990.tb00361.x.

Glaister, A., 1999. Guide to the identification of Australian Elmidae larvae (Insecta: Coleoptera). Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 21. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Goldstein, P.Z., DeSalle, R., 2011. Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. BioEssays 33, 135–147.
Govedich, F., 2001. A reference guide to the ecology and taxonomy of freshwater and terrestrial leeches (Euhirudinea) of Australasia and Oceania. Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 35. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Greenfield, Paul, 2017. Greenfield Hybrid Analysis Pipeline (GHAP). v1. CSIRO. Software Collection. https://doi.org/10.4225/08/59f98560eba25.
Growns, J.E., Chessman, B.C., Jackson, J.E., Ross, D.G., 1997. Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of 6 methods of sample processing. J. North Am. Benthol. Soc. 16, 682–693.
Hajibabaei, M., Porter, T.M., Robinson, C.V., Baird, D.J., Shokralla, S., Wright, M.T.G., 2019. Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. PLoS ONE 14 (12). https://doi.org/10.1371/journal.pone.0225409.
Hajibabaei, M., Porter, T.M., Wright, M., Rudar, J., 2019. COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS ONE 14.
Hajibabaei, M., Smith, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B., Hebert, P.D., 2006. A minimalist barcode can identify a specimen whose DNA is degraded. Mol. Ecol. Notes 6, 959–964.
Harvey, M.S., 1996. A review of the water mite family Pionidae in Australia (Acarina: Hygrobatoidae). Rec. the Western Aust. Museum 17, 361–393.
Harvey, M., Growns, J., 1998. A guide to the identification of families of Australian water mites (Arachnida: Acarina). Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 18. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Hebert, P.D., Cywinska, A., Ball, S.L., 2003. Biological identifications through DNA barcodes. Proc. Roy. Soc. Lond. Ser. B: Biol. Sci. 270, 313-321.
Hawking, J.H., Smith, L.M., LeBusque, K., Davey, C., 2013. Identification and Ecology of Australian Freshwater Invertebrates. http://www.mdfrc.org.au/bugguide.
Hilsenhoff, W.L., 1988. Rapid field assessment of organic pollution with a family-level biotic index. J. North Am. Benthol. Soc. 7, 65–68.
Hosseini, R., Schmidt, O., Keller, M.A., 2008. Factors affecting detectability of prey DNA in the gut contents of invertebrate predators: a polymerase chain reaction-based method. Entomol. Exp. Appl. 126, 194–202.
Kanagawa, T., 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J. Biosci. Bioeng. 96, 317–323.
Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., Zizka, V.M., 2020. Improved freshwater macroinvertebrate detection from eDNA through minimized non-target amplification. bioRxiv.
Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., Machida, R.J., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front. Zool. 10, 34.
Lester, R.E., Wright, W., Jones-Lennon, M., 2007. Does adding wood to agricultural streams enhance biodiversity? An experimental approach. Mar. Freshw. Res. 58, 687–698.
Madden, C., 2009. Key to genera of larvae of Australian Chironomidae (Diptera). Taxonomy Research and Information Network (TRIN) guide. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Marshall, J.C., Steward, A.L., Harch, B.D., 2006. Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: the benefits and costs of using species abundance data. Hydrobiologia 572, 171–194.
Marshall, N.T., Stepien, C.A., 2020. Macroinvertebrate community diversity and habitat quality relationships along a large river from targeted eDNA metabarcode assays. Environ. DNA 2, 572–586. https://doi.org/10.1002/edn3.90.
Meusnier, I., Singer, G.A., Landry, J.-F., Hickey, D.A., Hebert, P.D., Hajibabaei, M., 2008. A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9, 214. https://doi.org/10.1186/1471-2164-9-214.
Nichols, S.J., Kefford, B.J., Campbell, C., Bylemans, J., Chandler, E., Bray, J.P., Shackleton, M.E., Robinson, K.L., Carew, M.E., Furlan, E.M., 2020. Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: Effects of debris and storage conditions on the recovery of target taxa. Freshw. Biol. 65, 607–620. https://doi.org/10.1111/fwb.13443.
Nichols, S.J., Robinson, W.A., Norris, R.H., 2010. Using the reference condition maintains the integrity of a bioassessment program in a changing climate. J. North Am. Benthol. Soc. 29, 1459–1471.
Nilsson, R.H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.-H., Kõljalg, U., 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS ONE 1.
Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., 2018. The future of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637, 1295-1310.
Pinder, A.M., 2010. Tools for identifying selected Australian aquatic oligochaetes (Clitellata: Annelida). Museum Victoria Science Reports 13: 1–26. ISSN 0 7311-7253 1 (Print) 0 7311-7260 4 (On-line) http://www.museum.vic.gov.au/sciencereports/.
Ponder, W., 2013. Introduction to the Australian Freshwater Gastropods. Taxonomy Research and Information Network (TRIN) guide. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Porch, N., Perkins, P., 2010. Australian Hydraenid Beetles: Diversity, Ecology, Biogeography. Taxonomy Research and Information Network (TRIN) guide. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Ratnasingham, S., Hebert, P.D., 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Mol. Ecol. Notes 7 (355-364).
Reynoldson, T.B., Norris, R., Resh, V.H., Day, K., Rosenberg, D., 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. North Am. Benthol. Soc. 16, 833–852.
Rose, P., Metzeling, L., Catzikiris, S., 2008. Can macroinvertebrate rapid bioassessment methods be used to assess river health during drought in south eastern Australian streams? Freshw. Biol. 53, 2626–2638.
Ruppert, K.M., Kline, R.J., Rahman, M.S., 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecol. Conserv. e00547.
Shackleton, M., Rees, G.N., 2016. DNA barcoding Australian macroinvertebrates for monitoring programs: benefits and current short comings. Mar. Freshw. Res. 67, 380–390.
Shen, Y.-Y., Chen, X., Murphy, R.W., 2013. Assessing DNA Barcoding as a Tool for Species Identification and Data Quality Control. PLoS ONE 8.
Sheppard, S., Bell, J., Sunderland, K., Fenlon, J., Skervin, D., Symondson, W., 2005. Detection of secondary predation by PCR analyses of the gut contents of invertebrate generalist predators. Mol. Ecol. 14, 4461–4468.
Smith, M., Kay, W., Edward, D., Papas, P., Richardson, K.S.J., Simpson, J., Pinder, A., Cale, D., Horwitz, P., Davis, J., 1999. AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. Freshw. Biol. 41, 269–282.
Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21, 2045–2050.
Theischinger, G., Hawking, J., 1999. Dragonfly Larvae (Odonata): A guide to the identification of larvae of Australian families and to the identification and ecology of larvae from New South Wales. Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 24.
Theischinger, G., 2000. Preliminary keys for the identification of larvae of the Australian gomphides (Odonata). Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 28. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Theischinger, G., 2001. Preliminary keys for the identification of larvae of the Australian Synthemistidae, Gomphomacromiidae, Pseudocorduliidae, Macromiidae and Austrocorduliidae (Odonata). Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 34. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Theischinger, G., Endersby, I., 2009. Identification guide to the Australian Odonata, Department of Environment, Climate Change and Water NSW, Sydney, 283 pp. DECCW 2009/730 http://www.environment.nsw.gov.au/resources/publications/09730AustOdonata.pdf.
Tippler, C., Wright, I.A., Hanlon, A., 2012. Is catchment imperviousness a keystone factor degrading urban waterways? A case study from a partly urbanised catchment (Georges River, South-Eastern Australia). Water, Air, Soil Pollut. 223, 5331-5344.
Tixier, M.-S., Hernandes, F.A., Guichou, S., Kreiter, S., 2012. The puzzle of DNA sequences of Phytoseiidae (Acari : Mesostigmata) in the public GenBank database. Invertebrate System. 25, 389–406.
Vanhove, M.P.M., Tessens, B., Schoelinck, C., Jondelius, U., Littlewood, D.T.J., Artois, T., Huyse, T., 2013. Problematic barcoding in flatworms: A case-study on monogeneans and rhabdocoels (Platyhelminthes). ZooKeys 365, 355–379.
Walsh, C.J., 2006. Biological indicators of stream health using macroinvertebrate assemblage composition: a comparison of sensitivity to an urban gradient. Mar. Freshw. Res. 57, 37–47.
Watts, C.H.S., 2002. Checklists and guides to the identification, to genus, of adult and larval Australian water beetles of the families Dytiscidae, Noteridae, Hygrobiidae, Haliplidae, Gyrinidae, Hydraenidae and the superfamily Hydrophiloidea (Insecta: Coleoptera). Cooperative Research Centre for Freshwater Ecology Identification and Ecology Guide No. 43. https://www.mdfrc.org.au/bugguide/resources/taxonomy_guides.html.
Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Sci. Total Environ. 678, 499–524.
Wright, J., Furse, M., Moss, D., 1998. River classification using invertebrates: RIVPACS applications. Aquat. Conserv. Mar. Freshwater Ecosyst. 8, 617–631.
Yeo, D., Srivathsan, A., Meier, R., 2020. Longer is not always better: Optimizing barcode length for large-scale species discovery and identification. System. Biol.
Zaidi, R., Jaal, Z., Hawkes, N., Hemingway, J., Symondson, W., 1999. Can multiple-copy sequences of prey DNA be detected amongst the gut contents of invertebrate predators? Mol. Ecol. 8, 2081–2087.