Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding.

Kelly, M.G.[1,2 *,] Juggins, S.[3,] Mann, D.G.[4,5], Sato, S.[4,6], Glover, R.[7,8], Boonham, N.[7,9,] Sapp, M.[7,10], Lewis, E.[7,11], Hany, U.[7,12], Kille, P.[13], Jones, T.[14] & Walsh, K.[15]

[1] Bowburn Consultancy, 11 Monteigne Drive, Bowburn, Durham DH6 5QB, UK

[2] School of Geography, University of Nottingham, Nottingham NG7 2RD, UK

[3] School of Geography, Politics and Sociology, Newcastle University, Newcastle NE1 7RU, UK

[4] Royal Botanic Garden, Edinburgh EH3 5LR, UK

[5] IRTA-Institute for Food and Agricultural Research and Technology, Marine and Continental Waters Programme. Ctra de Poble Nou Km 5.5, Sant Carles de la Rápita, Catalonia, E43540, Spain

[6]. Present address: Department of Marine Science and Technology, Fukui Prefectural University, Fukui 917-0003, Japan

[7] Food and Environment Research Agency, Sand Hutton, York YO41 1LZ, UK

[8]. Present address: Taxa-genomics Ltd, Unit 11A-12A Village Walk, Onchan , IM3 4EB, UK

[9] Present address: Institute for Agri-Food Research and Innovation, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

[10] Present address: Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich Heine University, Population Genetics, Universitaetstrasse 1, 40225 Duesseldorf, Germany

[11] Present address: University of Leeds, Woodhouse Lane, Leeds LS2 9JT

[12] Present Address: Food Standards Agency, Foss House, Kings Pool 1-2 Peasholme Green, York YO1 7PR

[13] School of Biosciences, Cardiff University, Cardiff CF10 3AT, UK.

[14] Environment Agency, Sunrise Business Park, Higher Shaftesbury Road, Blandford Forum DT11 8ST, UK

[15] Environment Agency, Horizon House, Deanery Road, Bristol BS1 5AH, UK

[*] Author for correspondence: MGKelly@bowburn-consultancy.co.uk

Abstract

Fundamental differences in the nature of diatom assemblage composition data generated using light microscopy and molecular barcoding create problems when applying current paradigms and metrics developed for ecological assessment. We therefore describe the development of a new metric designed specifically for diatom *rbcL* barcode data gathered using high throughput sequencing (HTS).

Although the structure of datasets collected using HTS is similar to that collected using light microscopy (LM), differences in the proportions of key species between the two methods mean that the use of metrics designed for LM on HTS data gives biased results. We therefore recalibrated the Trophic Diatom Index in order to produce a version that is sensitive to nutrient pressures in rivers but that can be used with HTS data. Correlation between the LM and HTS metrics is good (r = 0.86 on a cross-validated model), meaning that 30% of sites will change class when the current Water Framework Directive classification approach is applied. Although less than 15% of diatom taxa recorded from UK and Ireland are included in the *rbcL* barcode reference database, gaps in this database are not a major source of variation between the HTS and LM models. We argue that use of metrics calibrated using HTS data is a more realistic option than applying correction factors to enable HTS data to be used with existing indices. We also stress the importance of starting the process of integrating HTS into ecological assessments with relatively conservative approaches. This enables the data collected by HTS to be related to those generated by established approaches, both now and during long-term monitoring, making it possible for scientists, regulators and stakeholders to have an informed conversation about the benefits and challenges of HTS. Overall, the study demonstrates that it is possible to translate the legal requirements of an ecological assessment framework from LM to HTS, though differences in these two approaches mean that there is unlikely to be perfect agreement between their outputs.

Keywords: Ecological assessment, metrics, Water Framework Directive, barcodes, diatoms, phytobenthos, *rbcL*

1. Introduction

Analysis of benthic diatoms forms one part of a suite of ecological methods used to inform decision making associated with European legislation, particularly the Water Framework Directive (WFD: European Union, 2000; Kelly, 2013; Poikane et al., 2016), in rivers and lakes in the United Kingdom (UK). The current UK method assesses ecological status as an Ecological Quality Ratio (EQR) based on the Trophic Diatom Index (TDI: Kelly et al., 2008). It uses light microscopy (LM) to analyse samples and is underpinned by European Standards (CEN, 2014a, b), producing outcomes that have been verified via the European Union's intercalibration exercise (European Union, 2008, 2013). LM-based diatom assessment is a time-consuming process requiring skilled individuals to both analyse samples and interpret data. There are several sources of uncertainty in this pathway, one of which is the process of identifying and enumerating the organisms (Kahlert et al., 2009; 2012). Although uncertainty associated with this stage can be controlled by training and quality control (Kelly, 2013b), these steps add to an already substantial resource commitment. Alternative approaches that offer a similar or better level of precision at a lower cost would be very attractive to government agencies working on tight budgets.

A further complication is that the widespread adoption of diatoms for assessments associated with the WFD (Kelly, 2013a) has taken place alongside a paradigm shift in understanding their taxonomy and phylogenetics. Several studies have shown considerable taxonomic diversity within aggregates formerly thought to be single species (e.g. Evans et al., 2008; Trobajo et al. 2009; Kermarrec et al., 2014; Rovira et al., 2015). This diversity often pushes the capabilities of LM, and analysts, to the limit.

Molecular techniques such as metabarcoding, which combines the principles of DNA barcoding with high throughput sequencing (HTS: sometimes also referred to as "next generation sequencing", NGS) to characterise diatom species, has opened up the possibility for new approaches to ecological assessment using diatoms (Mann et al., 2010; Kermarrec et al., 2014b; Visco et al., 2015; Bailet et al., 2019). These molecular techniques may address issues of cost and precision whilst, at the same time, potentially providing greater taxonomic sensitivity (Mann et al., 2010; Kelly et al., 2015). Previous studies have recommended the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) gene (Mann et al., 2010) as a potential DNA barcode for such purposes. Mann et al. (2010) argue that protein-encoding genes such as *rbcL* pose fewer practical problems than rDNA, once they have been obtained. Benefits of *rbcL* include that there is rarely any intragenomic variation and sequences are very easily aligned and compared. Furthermore, sequencing errors can often be detected by frame shifts or unlikely amino acid changes (e.g. polar by non-polar, basic by

86 acidic, etc). *RbcL* has been exploited for both taxonomy (e.g. Kermarrec et al. 2014a; Carballeira et
87 al. 2017; Kahlert et al. 2019) and ecological assessment (Kermarrec et al., 2014b, Chonova et al.
88 2019). It provides a very practical advantage over its nuclear SSU counterpart (also often used as a
89 barcode: Pawlowski et al. 2012) in the context of characterizing real-world diatom assemblages by
90 metabarcoding, since the amplicon is produced only from autotrophic ecosystem constituents,
91 rather than all eukaryotes.

92 This paper describes the steps taken to develop a metric using HTS data that is suitable for statutory
93 environmental regulation. Although some have advocated more radical approaches ("Biomonitoring
94 2.0": Baird & Hajibabaei, 2012; Woodward et al., 2013; Makiola et al., 2020), we have constructed a
95 molecular analogue of the UK's current approach using the Trophic Diatom Index (TD14: Kelly et al.,
96 2008, UK TAG, 2014). This ensures continuity with existing assessments whilst, at the same time,
97 complying with the normative definitions of the WFD, which refer to "taxonomic composition".
98 Whilst we agree with Baird and Hajibabaei (2012) that there is potential within DNA-based
99 approaches to explore aspects of diversity and ecosystem function that are difficult to measure
100 using traditional approaches, understanding the relationship between data collected by molecular
101 methods and those gathered by traditional approaches is an important first step that lays a
102 foundation upon which more innovative approaches can be built. Relating DNA sequences to the
103 corresponding morphospecies is, in our opinion, a necessary step if stakeholders are to develop trust
104 in these new methods.

105 In theory, both approaches – LM and HTS – yield a list of taxonomic categories, with the relative
106 abundance of each expressed as a proportion of the total. In practice, however, each addresses
107 different entities: LM records diatom valves (i.e. half of a frustule, or cell wall), whilst HTS records
108 *rbcL* genes. Because *rbcL* genes are part of the chloroplast, rather than the nuclear genome, and
109 because the number of chloroplasts varies between genera and the number of *rbcL* copies per
110 chloroplast is also variable, the relationship between LM and HTS data is not 1:1. This will be a
111 potential source of bias if LM methods for data processing are applied to HTS data. Vasselon et al.
112 (2017) suggested applying species-specific correction factors, based on cell biovolume, to bring HTS
113 data into line with expectations based on LM data. However, we believe that, as the LM data also
114 have a number of biases and limitations, it is better to treat HTS data at face value. We therefore
115 start by examining the relationship between LM and HTS data, and then go on to construct a
116 modification of the existing method to estimate ecological status. The motivation for this is not just
117 scientific: the current UK approach to evaluating ecological status using diatoms has been
118 harmonised with methods used elsewhere in the European Union (Kelly 2009; European Commission
119 2013) and a HTS analogue would provide continuity both when evaluating time-series of data and

when comparing UK status classifications with those from neighbouring EU states. In addition, responsibility for the environment has been devolved to the national administrations within the UK (England, Northern Ireland, Scotland, Wales) and close agreement between methods should allow each administration to decide on its approach without prejudice to the management of trans-frontier rivers. A key message from this paper is that metabarcoding data, though telling essentially the same story as LM data, are fundamentally different and require a different set of interpretative paradigms if a full appreciation of their meaning is to be unlocked.

2. Methods

2.1 Study design

Diatom samples were collected during 2014, 2016 and 2017 as part of the UK's routine surveillance monitoring program of rivers (Kelly et al 2018a, b). In most cases, two samples were collected from each site, one in spring and one in autumn. Samples covered a range of ecological conditions along the primary nutrient/organic gradient to which diatoms are known to be particularly sensitive. In total, there were 1223 matched LM and HTS samples from England, 87 from Northern Ireland, 268 from Scotland and 150 from Wales, yielding a total of 1714 samples available for analysis.

2.2 Diatom sample collection

Diatom samples were collected by placing five cobbles in a tray with approximately 50 ml of stream water and then brushing the upper surface of each with a toothbrush in order to remove the biofilm (these are standard procedures: CEN 2014a; Kelly et al. 1998). These samples were then transferred to the laboratory in a cool box. Using a Pasteur pipette 5 ml of the suspension of biofilm and water were transferred to a sterile 15 ml centrifuge tube containing 5 ml nucleic acid preservative based on RNAlater™ storage solution (Merck, Kenilworth, USA'), consisting of 3.5 M ammonium sulphate, 17 mM sodium citrate and 13 mM Ethylenediaminetetraacetic acid (EDTA). The sample was then frozen at –30 °C prior to DNA extraction. The remainder of the sample was preserved using Lugol's iodine for morphological analysis by LM.

2.3 Preparation and analysis of diatoms for light microscopy

Samples for LM were digested with either a mixture of sulphuric and oxalic acids and potassium permanganate or cold hydrogen peroxide (CEN, 2014b). Following digestion, samples were rinsed several times to remove all traces of oxidizing agents. Between rinses samples were either centrifuged or allowed to stand overnight in order to ensure that all diatoms settled to the bottom

of the tube. Permanent slides were prepared using Naphrax (Brunel Microscopes, Chippenham) as a mountant, following Kelly et al. (2008). At least 300 valves on each slide were identified to the highest resolution possible and their abundance recorded. The primary floras and identification guides used were Krammer and Lange-Bertalot (1986, 1997, 2000, 2004), Hartley et al. (1996) and Hofmann et al. (2011). All nomenclature was adjusted to that used by Whitton et al. (1998) which follows conventions in Round et al. (1990) and Fourtanier and Kociolek (1999).

2.4 Development of *rbcL* barcode reference database

2.4.1 Isolation and culture

Samples were collected from a number of locations in England and Scotland, encompassing a wide range of potential ecological diversity in order to establish a reference database of diatom barcodes. A few drops of biofilm/water suspension were placed in Petri dishes and individual cells of diatoms isolated by micropipette or by streaking on 2–3% agar plates. Selected cells (or, in the case of plated material, discrete small colonies of clonal cells) were transferred into small volumes of freshwater medium (WC medium with silicate, adjusted to pH 7; Guillard & Lorenzen 1972) in 96-well plates. After a few days of incubation, the health and clonal nature of each culture were confirmed under an inverted microscope. Successfully established clonal cultures were then grown in 90 mm Petri dishes for DNA extraction and preparation for a voucher slide.

2.4.2 Harvesting for vouchering and DNA extraction

Slurries of cells were transferred to 1.5 ml tubes and centrifuged at 2000 x *g* for 10 minutes. The pellet was transferred to a 1.5 µl tube and kept at –20°C until DNA extraction, leaving a small amount which was resuspended with distilled water and dried onto one 18 mm square coverslip and one 13 mm diameter circular coverslip. The square coverslip was used to prepare a voucher slide for LM ; the circular coverslip was retained in case of the need to examine material with scanning electron microscopy (SEM). For both LM and SEM vouchers, cells were cleaned *in situ* on coverslips by adding nitric acid to the coverslip on a hotplate and heating to oxidize organic material (Trobajo & Mann, 2019). After oxidation, the diatom cell walls, still on the coverslips, were washed with distilled water several times to remove digestion products and then dried again on a hotplate. For LM vouchers, cells were mounted in Naphrax and photographed using a Zeiss Axio-imager photomicroscope using 100x or 63x oil immersion objectives (nominal NA 1.4) and either bright field or Nomarski interference contrast optics. The smaller coverslips are stored in 100-well Eppendorf storage boxes at the Royal Botanic Garden Edinburgh (RBGE, herbarium abbreviation E), UK.

DNA was extracted using a QIAxtractor (Qiagen). Forward (DPrbcL1: AAGGAGAAATHAATGTCT) and reverse (DPrbcL7: AARCAACCTTGTGTAAGTCTC) primers (Jones et al. 2005) were used to amplify a 1400-bp region of the *rbcL* gene using the following reaction: 10 ng DNA, 1 mM dNTPs, 1 x Roche diagnostics PCR reaction buffer (Roche Diagnostics GmbH, Mannheim, Germany), 1 unit Taq DNA polymerase (Roche), and 0.5 µM of each primer. The final reaction volume was made up with nuclease-free water to 25 µl. Amplification was carried out under the following conditions: 94°C for 3 min, followed by 30–40 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min, with a final extension of 72°C for 5 min. PCR products were visualised by agarose gel electrophoresis against known standards and purified using ExoSAP-IT (USB Corporation, Ohio, USA).

2.4.3 Sanger sequencing

Sequencing was conducted in 10 µl volumes using 0.32 µM of PCR primer or sequencing primers NDrbcL5: CTCAACCATTYATGCG and DrbcL11: CTGTGTAACCCATWAC (Jones et al. 2005) and 1 µl of BigDye v3.1 and 2 µl of sequencing reaction buffer (Applied Biosystems). Sequencing PCR conditions were 25 cycles of 95°C for 30 s, 50°C for 20 s and 60°C for 4 min. Excess dye-labelled nucleotides were removed using the Performa DTR V3 clean-up system (EdgeBio) and sequence products were run on an ABI 3730 DNA sequencer (Applied Biosystems) at the University of Edinburgh.

Sequencing reads were edited and assembled using SeqMan (DNASTAR, Madison, WI). Each *rbcL* region was sequenced by four reads (using primers DPrbcL1, DPrbcL7, NDrbcL5 and DrbcL11) and the whole region sequenced by at least two overlapping reads. Sequences were defined as "high-quality" if all the reads were obtained successfully and resulted in no ambiguous bases, whereas "low-quality" reads were those with at least one read having weak signal(s) and/or noise(s), so that not all of the sequence region was covered by multiple overlapping reads. Because *rbcL* is a translated protein (with almost no variation in sequence length in diatoms) the gene sequences of different taxa were easily aligned manually in BioEdit 7.0.2 (Hall 1999).

The barcode reference database was supplemented by a number of sequences obtained from previous studies at RBGE (e.g. Evans et al. 2008; Trobajo et al. 2009; Rovira et al. 2015) along with sequences from R-Syst::diatom (now Diat.barcode: Rimet et al. 2019) and from GenBank (where the source laboratory had published at least one diatom taxonomy paper in a peer-reviewed journal). Sequences for a few taxa (e.g. *Platessa oblongella*) were identified using a process similar to that of Rimet et al. (2018) in which occurrence frequency in LM and HTS and their relative phylogenetic position using Maximum Likelihood was compared. This resulted in a barcode reference database (now largely incorporated in the publicly available Diat.barcode dataset at https://data.inra.fr/dataset.xhtml?persistentId=doi%3A10.15454%2FTOMBYZ) containing 1232

strains representing 346 species, of which 29 are planktic taxa which are not used for calculation of TDI4 but ensure that as many reads as possible are assigned to taxa (Table S2). The library is available in both Phylip and fasta formats via Supplementary materials.

2.5 Preparation and analysis of diatoms for HTS

2.5.1 DNA extraction

DNA extraction used enzymatic lysis with Proteinase K followed by column purification using Qiagen DNeasy® Blood and Tissue kit according to the manufacturer's instructions (Eland et al., 2012), automated using a BioRobot Universal System (Qiagen). DNA was quantified using a Qubit fluorometer and dsDNA BR Assay kit, again following the manufacturer's instructions (Thermo Fisher Scientific, Cat: Q32850). Genomic DNA was stored at –30°C prior to PCR and sequencing.

2.5.2 PCR amplification of rbcL barcode

The design and properties of the short barcode used for metabarcoding UK river diatoms are described in detail by Kelly et al. (2018a). It differs slightly from the *rbcL* barcode designed and used by Vasselon et al. (e.g. 2017): though covering the same region, it is slightly longer (331 bp rather than 312 bp). Amplification was carried out using the following reaction: 6 µl of HF buffer (NEB, USA), 0.3 µM forward (rbcL-646F: ATGCGTTGGAGAGARCGTTTC) and reverse primers (rbcL-998R: GATCACCTTCTAATTTACCWACAACTG), 0.3 mM of dNTPs, 0.3 µl Phusion high fidelity DNA polymerase (NEB) and 0.5 µl of a 1:10 dilution of sample DNA. The final reaction volume was made up with nuclease-free water to 30 µl. PCR was carried out on a C1000 thermal cycler (Bio-Rad, UK) under the following conditions: 98°C for 2 min, followed by 35 cycles at 98°C for 20 s, 55°C for 45 s, 72°C for 60 s, followed by a final extension at 72°C for 5 min. PCR products were visualized on 1% agarose gels and purified using AMPure Beads following the Illumina 16S Metagenomic Sequencing library preparation protocol (https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). DNA was eluted in 50 µl nuclease-free water.

In order to confirm that the primer did not introduce significant bias, forward and reverse primer sites in all sequences > 1kb (3299 in total) in the most recent release of Diat.barcode (version 8: Rimet et al., 2019) were examined. Allowing for 2 bp mismatch for each primer, only 11 sequences had neither primer site, with another 11 missing forward and 37 missing reverse primer sites. Therefore, of the sequences represented in the current database 98% would be predicted to amplify with the current primers indicating no issue with primer bias.

2.5.3 Index addition

In order to reduce between-run contamination from carry-over of indexed samples, three groups of indices were used sequentially, resulting in indices only being used every third MiSeq run. This enabled any carry over sequences to be identified and eliminated from the analysis.

Illumina Nextera XT sequencing adapters and indices were added to each sample using the following reaction: 10 µl HF buffer, 0.3 mM dNTPs, 1 µM MgCl2, 0.5 µl Phusion polymerase (NEB, USA), 5 µl of each specific 'index 1' and 'index 2' primer and 5 µl of purified sample PCR product. The final reaction volume of 50 µl was made up with nuclease-free water. Amplification was carried out on a C1000 thermal cycler (BioRad, UK) using the following conditions: 98°C for 3 min, followed by 8 cycles of 98°C for 30 s, 55°C for 30 s and 72°C for 30 s, with a final extension of 72°C for 5 min. PCR products were purified with AMPure Beads following the Illumina 16S Metagenomic library preparation protocol. Final libraries were eluted in 25 µl nuclease free water. Quality and quantity of each amplicon library was evaluated with Tapestation (Agilent, USA) along with quantification using Qubit (Life Technologies, CA, USA) prior to sequencing.

2.5.4 Illumina Sequencing (MiSeq)

All samples, including negative (water), positive (a mock-community composed of extracted DNA from 11 diatom species obtained from culture collections: see Kelly et al., 2018 for more details), no-template, index and extraction buffer controls were combined in equal concentrations to produce a 20 nM library, quantified and diluted to produce a final 4 nM library for sequencing. The library was denatured and combined with 5% PhiX sequencing control DNA and loaded onto a MiSeq instrument following the Illumina 16S Metagenomic sequencing library preparation protocol.

2.5.5 Bioinformatic analysis

Quality control consisted of removal of primers using Cutadapt v1.9.1 (Martin 2011), sliding window trimming of poor quality 3' ends of sequences from both strands using Sickle v1.33 (Joshi & Fass 2011) in paired end mode, joining trimmed paired end reads to form one consensus strand using PEAR v0.9.6 (Zhang et al. 2014), followed by a further round of quality assessment for the removal of sequences with an overall accuracy of less than 99.9% using Sickle v1.33 (Joshi & Fass 2011) in single-read mode. Samples with fewer than 3000 reads were either repeated or excluded from the analysis.

Sequences were clustered into operational taxonomic units (OTUs) with UCLUST (Edgar 2010) at 97% similarity. The most abundant sequence in the OTU was selected using QIIME v1.9.1 (Caporaso et al.

2010) and assigned to taxa following BLASTn against the custom-built closed *rbcL* barcode reference database. A similarity threshold of 95% for each BLAST identification was applied and those sequences with hits below 95% were described as having no specific identification. Relative abundance calculations were carried out within QIIME v1.9.1 (Caporaso et al. 2010). QIIME outputs were then converted to species lists with relative abundance estimates. Unknown and planktic taxa were removed from species lists and the proportions of remaining taxa recalculated to give a total of 100% (which expresses the abundance of each taxon in a sample in the way that it is weighted in the TDI4 metric calculations).

2.6 Development of HTS TDI metric

2.6.1 Datasets

Samples for which both LM and HTS data were available (total: 1714) were used to build the HTS metric. Only samples for which at least 3000 reads could be assigned to taxa in the reference library were carried forward for subsequent analyses.

Environmental data were obtained from each of the UK regulatory agencies and are expressed as either the mean (alkalinity and pH) or geometric mean (all other variables) of all available data for the period 2012 to 2016. Table 1 summarises the range of key environmental variables. The dataset has good coverage of the alkalinity and conductivity gradients, with coverage of the latter falling off at about 1000 µS cm^{-1}, suggesting limited coverage of brackish conditions. Most samples have circumneutral pH with just a small number with pH <7.0. The distribution of phosphorus values is curtailed at 0.01 mg L^{-1}, the routine detection limit for this determinand in England, whilst the nitrate-N dataset extends down to 0.1 mg L^{-1}. Ammoniacal-N is included in this summary to show the relatively small number of samples in the dataset with evidence of elevated levels of organic pollution (the limited data for biochemical oxygen demand shows the same trend). A subset of 1505 samples, with matched environmental data, was used to develop indicator values for the HTS metric and used to compare with the LM metric and the nutrient pressure gradient.

303 **Table 1.** Summary statistics of selected environmental variables for the combined LM/HTS dataset.

Variable	Units	N	Mean	Median	Min	Max
PO_4-P	µg L^{-1}	1505	81.1	28.2	1	3600
NO_3-N	mg L^{-1}	1505	2.47	1.45	0.05	27.3
NH_4-N	µg L-1	1029	56	37	5	884
Conductivity	µS cm^{-1}	1357	320	238	26	2162
Alkalinity	mg L^{-1} $CaCO_3$	1505	79.9	56.2	1.7	382
pH		1373	7.7	7.8	5.8	8.4

304

305 *2.6.2 Comparison of LM and HTS datasets*

306 Non-metric multidimensional scaling (NMDS: McCune & Grace 2002) was used to investigate the
307 structure of the LM and HTS datasets using the vegan package in the R software package (R Core
308 Team 2012) (Oksanen et al. 2007) for multivariate analyses. The aim of NMDS was to produce a low
309 dimensional representation of the dissimilarity between samples, measured across all taxa. This
310 examined the consequences of any differences on the structure of the datasets which, in turn, would
311 indicate whether a) ecological status concepts developed for LM can be reliably transferred to HTS
312 and b) inferences derived from HTS data can be compared with older data based on LM.

313 The success of the NMDS is given by the stress, which quantifies the agreement between the (in our
314 case) 2D representation and original dissimilarities, with values < 0.1 representing a good ordination
315 from which inferences may be drawn, 0.1-0.2 representing an ordination that is usable with caution,
316 0.2-0.3 representing an ordination that may be problematic, especially towards the the upper range
317 of the interval, and > 0.3 indicating that the ordination may be misleading (Zuur et al. 2007).

318 The similarity in structure between LM and HTS ordinations was tested using a Procrustes analysis
319 and associated permutation test (Peres-Neto & Jackson 2001) in vegan. In addition, TDI4 was
320 calculated for all samples in both LM and HTS datasets using DARLEQ2 software
321 (http://www.wfduk.org/resources/category/biological-standard-methods-201). Scatterplots,
322 Pearson's correlation coefficient and, where appropriate, Lin's concordance correlation coefficient
323 (Lin 1989) were used to evaluate relationships between ordination axes and metric values . Lin's

concordance correlation coefficient is a modification of correlation analysis which assesses the deviation from a perfect 1:1 relationship between the 2 variables and, as such, is useful for determining whether a change in approach will have a systematic effect on results. It was calculated by means of the epiR package (Stevenson 2010) within R.

2.6.3 Re-evaluating TDI4 against the pressure gradient

Before deriving the new HTS metric, the relationship between TDI4 and the nutrient pressure gradient was appraised. A weighted average model (WA: Ter Braak, 1986) was derived that directly calculates species indicator values as the weighted mean of their distribution along the pressure gradient. Models were derived by comparing TDI4 to P-PO_4, N-NO_3 and the first component of a principal components analysis of PO_4-P- and NO_3-N (PC1), which in effect, combines the phosphorus (P) and nitrogen (N) gradients into a single variable. All P and N values were $\log_{10}$ transformed before analysis. TDI4 indicator values were then plotted against the WA indicator values (that is, the so-called WA optima) of a model for the PC1 nutrient pressure gradient. WA calculations were performed in R using the package rioja (Juggins 2015). Pearson correlation coefficients were used to describe the relationship between models and pressure variables. Sensitivity values were adjusted in 88 cases and the revised version of the index is henceforth referred to as "TDI5LM".

2.6.4 Derivation of HTS TDI

Having optimised species indicator values for the LM metric, the next step was to derive a new HTS metric ('TDI5NGS'). The objective was to mimic the TDI5LM scores as closely as possible using a WA algorithm to derive HTS taxon indicator values that best predicted TDI5LM values. This was done using a WA regression which used the TDI5LM value for each sample as the independent variable and the HTS species assemblage data as the dependent variable. The outcome was a set of species indicator values that best predicted the TDI5LM value using HTS data. The species indicator values represent weighted centroids or 'optima' of HTS taxa along the TDI5LM gradient. WA regression is known to shrink the range of optima compared with the range of the target gradient (TDI5LM). Therefore species indicator values were expanded using a deshrinking regression of TDI5NGS sample scores on TDI5LM sample scores. This is a usual and necessary step in WA regression and calibration (Birks et al. 1990). Shrinkage is more pronounced at the gradient ends and so non-linear deshrinking, using a monotonic GAM-based smoothing spline (Birks and Simpson 2013), was used to align the original TDI5NGS sample scores to the range of TDI5LM scores. The monotonic regression removes the edge effects inherent in WA calibration, but tends to underestimate values at the low end and overestimate values at the high end of the TDI5LM gradient. A final linear deshrinking was therefore

performed using major axis regression of TDI5NGS scores on TDI5LM scores to optimise TDI5NGS sample scores and avoid under/over prediction at the gradient ends. TDI5NGS was tested against TDI5LM using Pearson's correlation coefficient and Lin's concordance correlation coefficient, as described above. Final TDI5NGS scores are listed in Table S1.

R code (R Development Core Team 2017) that implements the above algorithm is available in the R package DARLEQ3 at https://github.com/nsj3/darleq3.

2.7 Testing and validation of HTS metric

A five-fold cross validation of the HTS metric was used to test its robustness and performance when confronted with new HTS data. The dataset was split at random into five equal-sized fractions and the model trained on four-fifths of the data. The remainder was used to test the model. The process was repeated five times for each left-out group and the TDI5NGS scores aggregated across the five test groups.

Two additional forms of validation of the model were performed. First, a comparison of ecological status classifications produced using LM and HTS classifications in order to demonstrate the extent to which the ecological status for a water body might change if the HTS method was adopted. This provided a broad nationwide perspective so, in order to provide insights at a more local scale, HTS data were plotted alongside LM data for three catchments which had been monitored intensively for a number of years.

To generate classifications from the diatom metrics, an estimate of the value expected if there is no or minimal anthropogenic impact is required. This is predicted by an equation which uses alkalinity to predict the value of TDI4 (Kelly et al. 2013):

$$eTDI4 = 9.933 \times \exp(\log10(Alk) \times 0.81)$$

where: eTDI4 = expected value of the TDI4; and Alk = average alkalinity at the site.

EQRs are then calculated as (100 – observed TDI4) / (100 – expected TDI4) and status classes are assigned as follows: EQR > 1: high status; EQR > 0.75, ≤ 1: good status; EQR > 0.5, ≤ 0.75: moderate status; EQR > 0.25, ≤ 0.5: poor status; EQR ≤ 0.25: bad status.

The method by which the HTS model was derived means that boundaries for (LM) TDI4 also apply to TDI5LM and TDI5NGS. A comparison between classifications obtained by the two methods used normalised versions of class boundaries (high: 0.8, good: 0.6, moderate: 0.4, poor: 0.2). "Bias" is the percentage of samples that are classified in a higher class using one metric when compared with another.

Data for the study of implications for classification on long-term trends in TDI4 came from a study of the variability in LM and HTS described in more detail in Kelly et al. (2018a). Sites chosen for this study had been monitored intensively for other purposes using LM both before and after this study and there was no reason to suppose a temporal trend in ecological status during this time. This permitted both a direct comparison between LM and HTS results for four samples collected over the course of a calendar year (by paired-sample t-test) and a comparison between the four HTS samples and the longer-term dataset (by two-sample t-test). Bartlett tests for homogeneity of variances were also performed These time series provided an indication of the scale of fluctuations encountered when samples were analysed by LM. This, in turn, provided insights into the extent to which agreement between LM and HTS samples in the spatial dataset might be expected to translate into long-term differences in classification results due to a change in method.

3. Results

3.1 Comparison of LM and HTS datasets

An average of 28 taxa were found in the 1714 samples analysed by light microscopy, with a minimum of 2 taxa and a maximum of 72 taxa. HTS samples consisted of an average of 45,544 reads, of which 60% could be assigned to taxa in the reference library and these, typically, recorded more species than were found in LM: an average of 68 and a range of 11 to 148. However, there was considerable scatter in all the relationships between relative abundance of LM and HTS outputs for individual taxa, reflecting uncertainty in both axes associated with the calculation of proportions of single taxa from a pool of many taxa. The general tendency was for small, single-celled species such as *Achnanthidium minutissimum* and *Amphora pediculus* to have lower representation in HTS than LM (Fig. 1) whilst larger cells with two (i.e. *Navicula lanceolata; Ulnaria ulna*) or many (i.e. *Diatoma vulgaris; Melosira varians*) chloroplasts typically had greater representation in HTS compared to LM (Fig. 1). Particular issues were encountered for *Fistulifera saprophila* and *Mayamaea atomus* var*. permitis*, both of which were more abundant in the HTS data but often absent from corresponding LM analyses (Fig. 1). *Planothidium frequentissimum*, by contrast, was abundant in the LM data but barely represented in the HTS data despite being represented in the barcode database, suggesting a more complicated issue relating either to the true identity of the strain currently named as "*Planothidium frequentissimum*" or to differences in the genotypes of field populations compared with sequences in the barcode database.

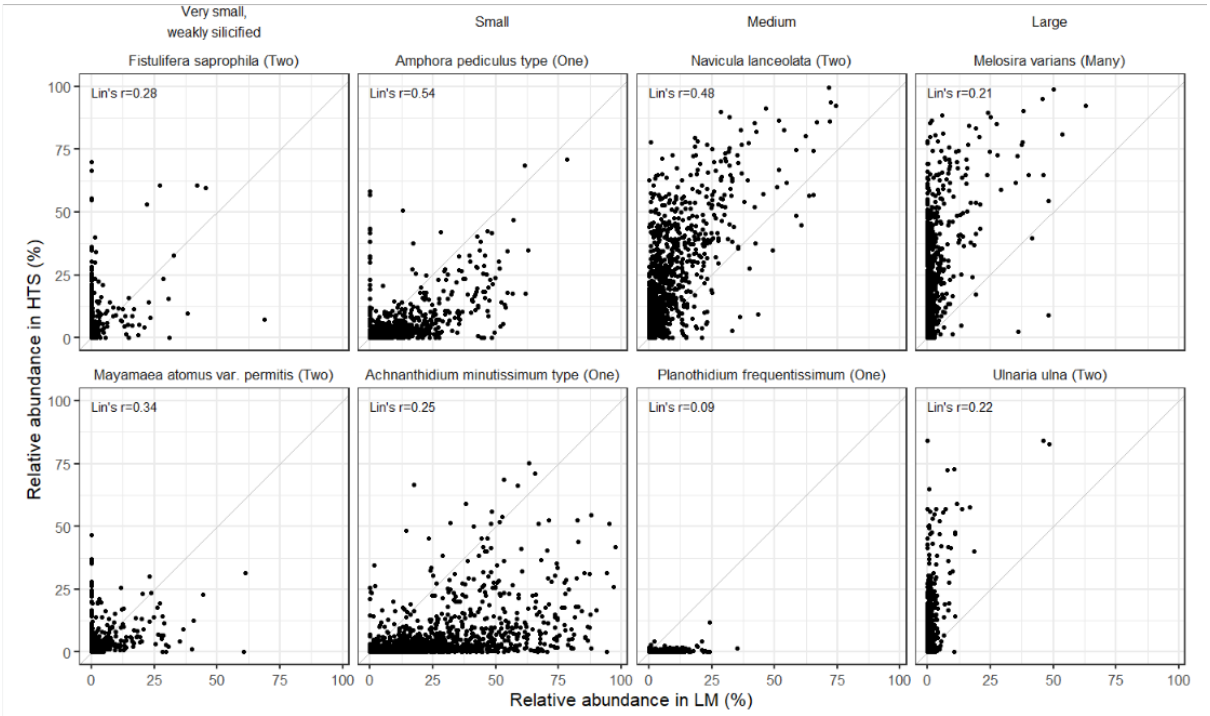


Figure 1. Differences and Lin's concordance correlation coefficient between LM and HTS analyses of selected diatom species that vary in size and (in parentheses after name) chloroplast number. The diagonal line shows 1:1 agreement between the two approaches.

For both LM and HTS datasets, NMDS yielded ordinations with levels of stress just above the threshold of "usable with caution" (LM: 0.24, HTS: 0.22). . However, the two ordinations showed similar structure, demonstrated by the first axes of each being strongly correlated (Pearson correlation coefficient, r = 0.88) (Fig. 2a) and by the correlation between the first two axes assessed by a Procrustes analysis (p = 0.001; 999 permutations). Moreover, the first axis of the NMDS based on LM was strongly correlated with TDI4 calculated using LM data (Pearson correlation coefficient, r = 0.91) (Fig. 2b).

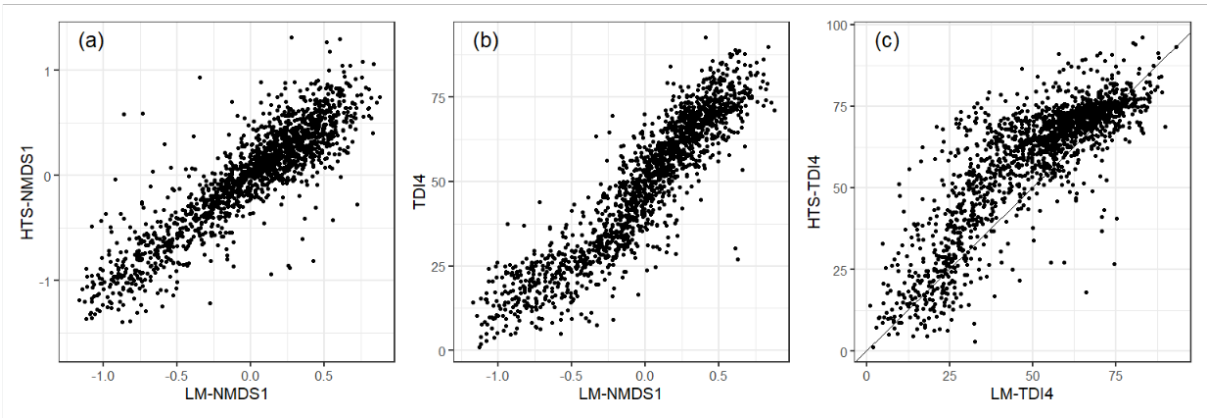


Figure 2. Comparison of LM and HTS data for 1714 samples from UK rivers**,** showing (a) comparison of the first axes of NMDS ordinations performed using LM and HTS data (r = 0.88); (b) axis 1 of NMDS of LM data versus TDI4 (r = 0.91); (c) relationship between TDI4 values calculated using LM and HTS data (r = 0.83; Lin's r = 0.77).

TDI4, calculated using the current version but with HTS data, was also strongly correlated with the TDI4 calculated using LM data (Figure 2c; Pearson correlation coefficient, r = 0.83) but the line deviated from 1:1 (Lin's concordance correlation coefficient: 0.77), with many HTS analyses returning higher values for the same sample than LM when TDI4 was low or moderate.

3.2 Optimising the LM model and deriving the HTS metric

Although there was good agreement in the structure of the LM and HTS datasets, initial comparisons of the distribution of species within the LM and HTS datasets led to the suspicion that some indicator values used in the current TDI4 model may be overestimates or underestimates of their sensitivity to the nutrient pressure gradient. Therefore, before attempting to derive a new version of the HTS metric, the fit of the LM TDI4 model to the pressure gradients PO_4-P, NO_3-N and PC1 (combined PO_4-P and NO_3-N) was plotted against each of these (Fig. 3).

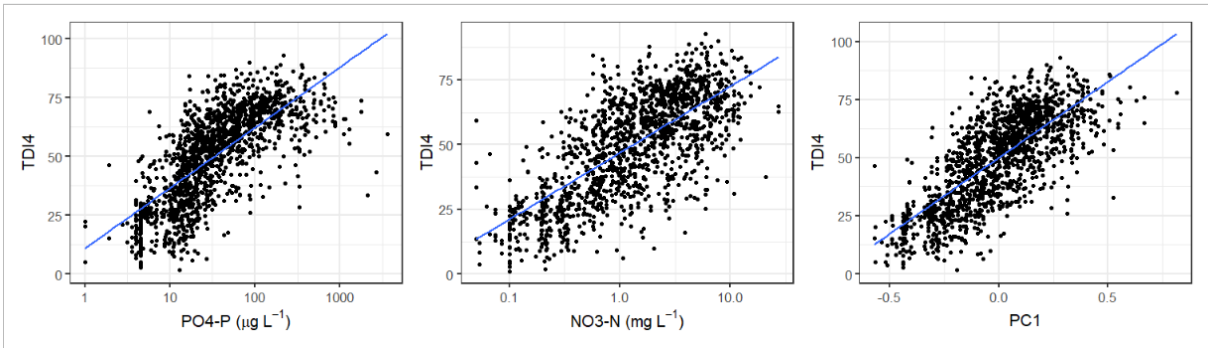


Figure 3. Relationship between TDI4 and the three nutrient pressure variables PO_4-P (r=0.71), NO_3-N (r=0.71) and PC1 (r=0.75).

The relationships between TDI4 and P-PO_4 and N-NO3 are similar (r = 0.71) although the relationship with P-PO_4 appears to be weaker at low phosphorus concentrations, where there may be measurement and detection limit issues. The relationships between TDI4 and pressures exhibit some non-linearity. This is informative insofar as it indicates points along the gradient where the model may be less sensitive to changes in pressure, but this does not compromise the quality of the model *per se*. The correlation between TDI4 and PC1 is the strongest (r = 0.75). PC1 was therefore used to represent the nutrient pressure gradient in all subsequent analyses. TDI5LM has only a small effect on the overall relationship with PC1 (Fig. 4b), . An HTS-specific metric ("TDI5NGS") derived using a WA algorithm and a monotonic generalised additive rescaling model (GAM) was strongly correlated with TDI5LM (Pearson's correlation and Lin's concordance coefficients both 0.89) but TDI5NGS had a slight tendency to overestimate at low TDI5LM values and underestimate at high values. However, when the relationship between TDI5LM and TDI5NGS was tested against the pressure gradient of PC1, TDI5NGS had a slightly weaker relationship to the nutrient pressure gradient than TDI5LM (r = 0.70 and 0.75 respectively (Figs. 4b and 4d).

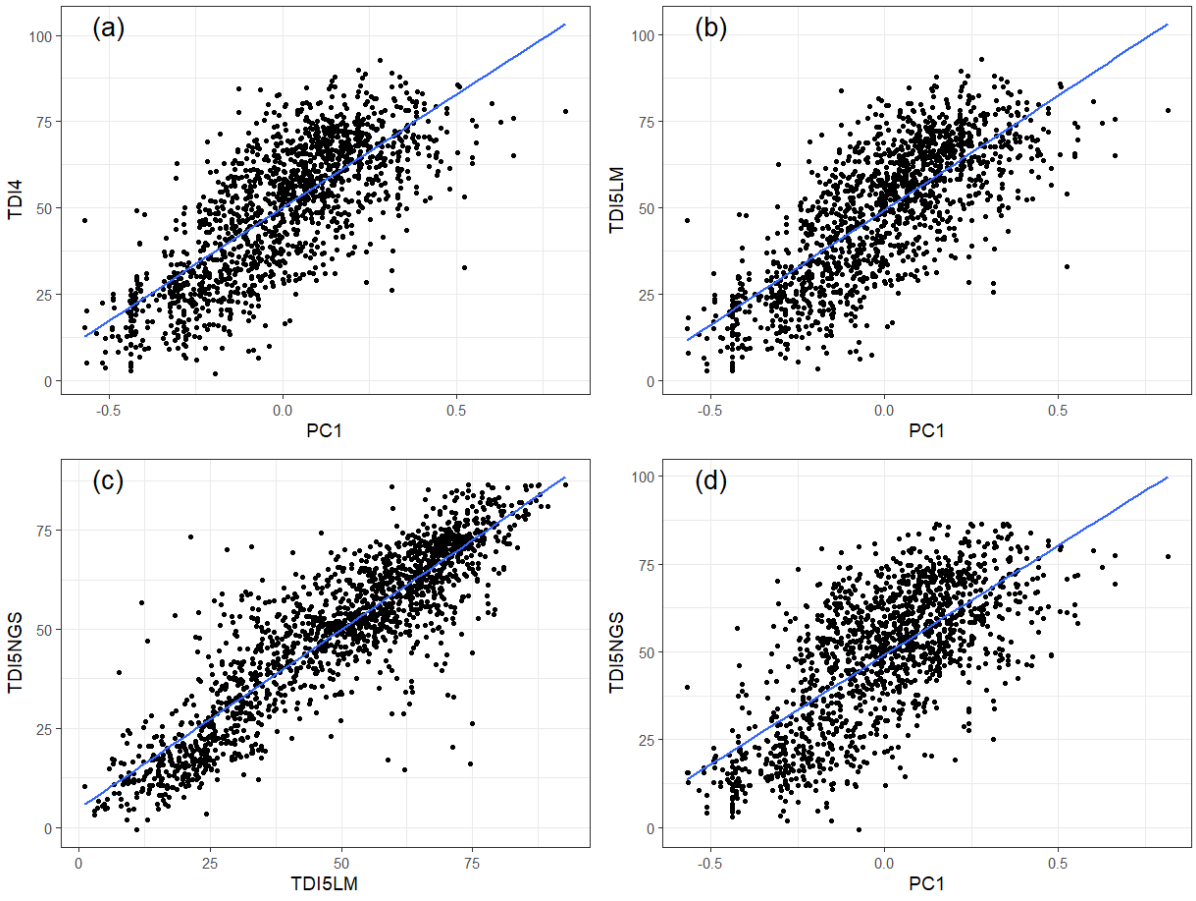


Figure 4. Relationship between (a) TDI4 and the combined nutrient pressure gradient PC1 (Pearson correlation coefficient, r = 0.75); (b) TDI5LM and PC1 (r = 0.75); (c) TDI5LM and TDI5NGS (r = 0.89); and (d) TDI5NGS and PC1 (r = 0.70)

3.3 TDI5NGS model performance

Under five-fold cross-validation the correlation between TDI5LM and TDI5NGS is only marginally lower (Pearson's correlation coefficient, r = 0.86) than that for the metric without cross-validation (r = 0.89). This means that the metric is robust and that the correlation cited above between LM and HTS methods is a good guide to the expected agreement between the two methods when applied to new data

It was important to understand how much of the observed difference between metrics calculated with LM and HTS data could be due to gaps in the barcode reference database, which currently represents just 346 of over 2500 species recorded from UK and Ireland freshwaters. Fig. 5 shows the relationship between TDI5LM calculated with all available taxa (x axis) and TDI5LM calculated with just those taxa included in the barcode database. The high correlation between the two variants (Pearson's correlation coefficient, r = 0.994) suggests that most of the relevant biological variation within diatom assemblages is captured by the barcode database, although there are some outliers where the variation is greater.

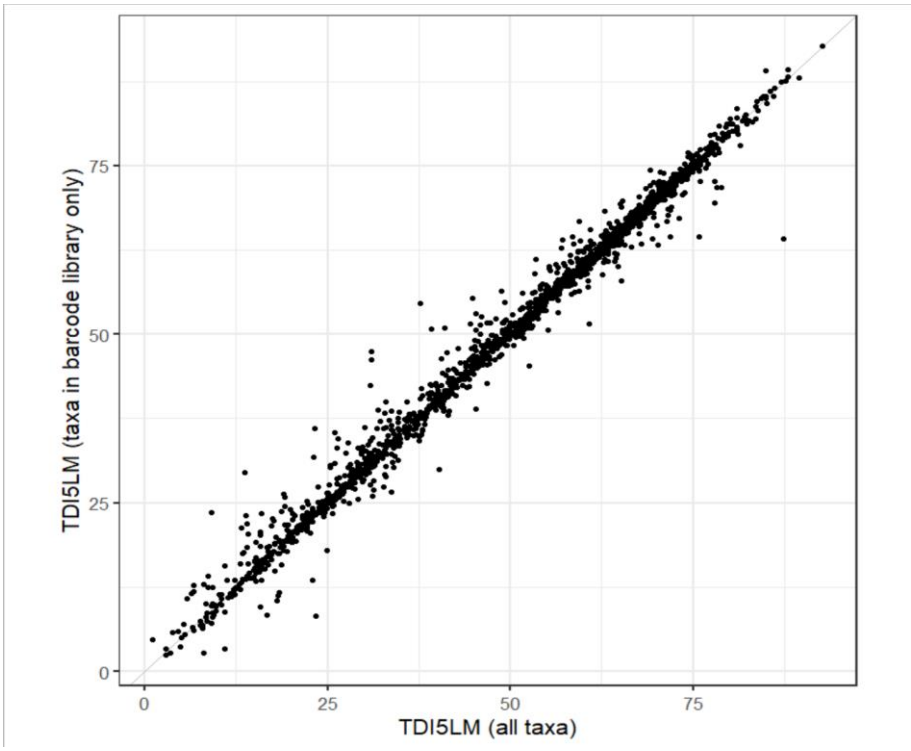


Figure 5. Difference between TDI5LM calculated with all taxa and with just those taxa that are represented in the barcode database (r=0.994).

3.4 Implications of changing to HTS analyses for classification of ecological status

Comparisons between status calculated using TDI4LM and TDI5LM results in only 7 % of sites changing class (Fig. 6a). Moving from LM to HTS, however, results in around 30% of sites changing class, although only around three percent would shift more than one class (Fig. 6b & 6c). Overall, the bias (i.e. the tendency to classify into a higher or lower class) between ecological status calculated with the current metric ("TDI4") and TDI5LM or TDI5NGS is slightly negative (–6.7% TDI5LM, –5.0% TDI5NGS), indicating that the new classifications are slightly less likely to result in a precautionary classification for both LM and HTS methods. By contrast, bias between TDI5LM and TDI5NGS is low (0.9%), suggesting that these may be more interchangeable than TDI4LM and TDI5NGS.

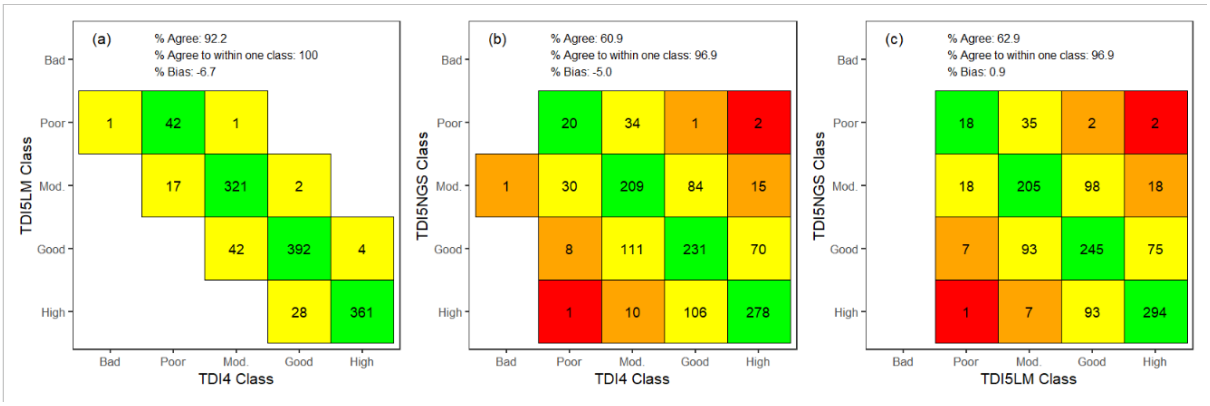


Figure 6. Comparisons between ecological status classes for samples computed by TDI5LM and TDI4 (a.), TDI5NGS and TDI4 (b.) and TDI5NGS and TDI5LM (c.). Green shading: identical classification for both metrics; yellow shading: agreement to within one class; orange shading: agreement to within two classes; red shading: greater than two class difference between methods. Embedded text shows summary statistics for each set of comparisons. N=1211 for each of the three sets.

3.5 Consequences of changing to HTS analyses on long-term trends in TDI

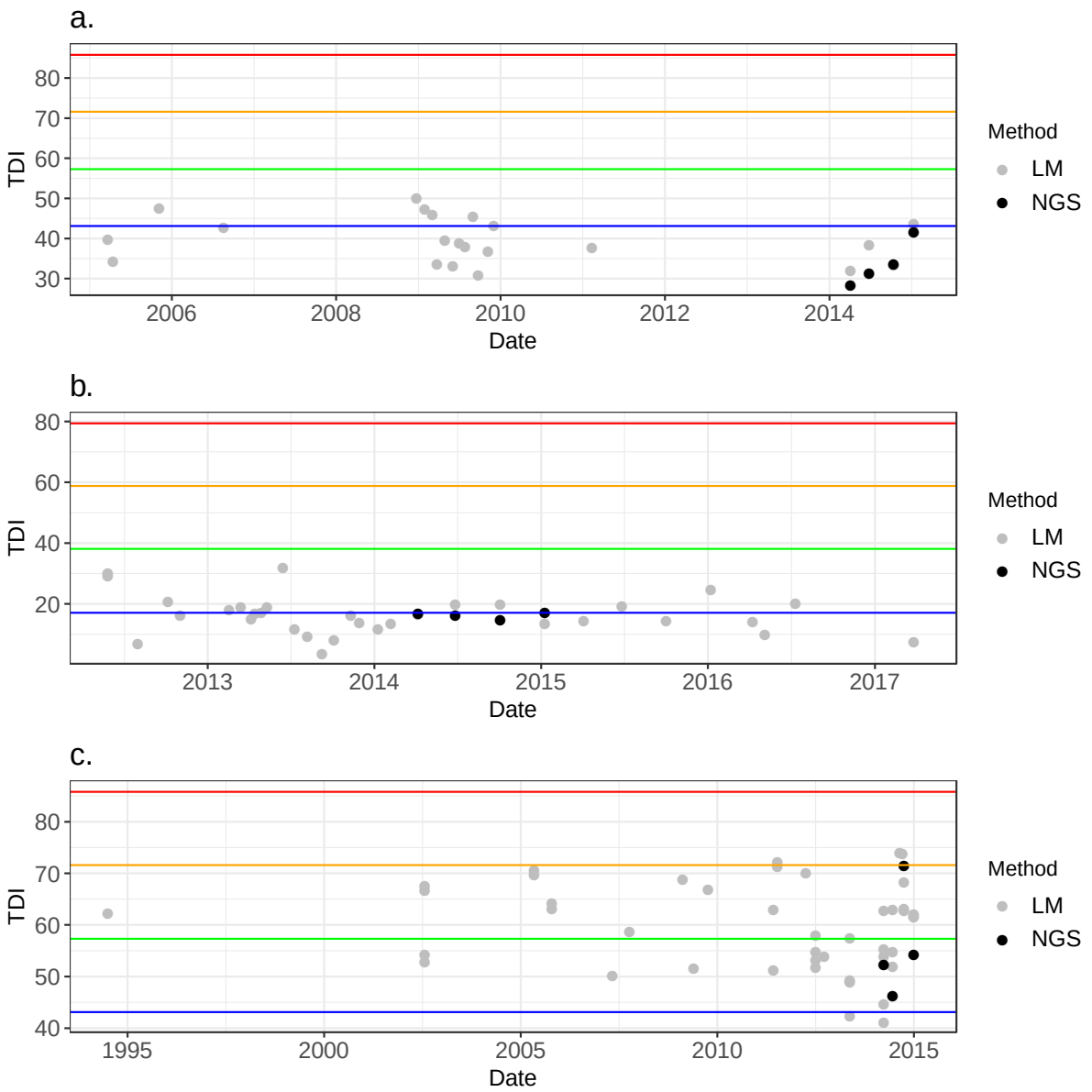
Long-term average TDI4 for the River Wear at Wolsingham in County Durham at the eastern edge of the Pennines, with relatively low population density and low intensity agriculture upstream, was 40, indicating high status but close to the boundary with good status (Fig. 7a). Mean TDI5NGS for four samples collected during 2014 was 34, also indicating high status. There was no significant difference in mean or variance between TDI5NGS and either long-term or 2014 data (Table 2). All 2014 TDI5NGS samples were consistent with high-status but 7 TDI4 samples from the long-term dataset indicated good rather than high status.

Long-term average TDI4 at the River Ehen near Ennerdale Bridge in Cumbria was lower but, as the alkalinity at this site is lower, the status class boundaries are also lower and the long-term average is very close to the boundary, with 14 of 32 samples indicating good, rather than high status (Fig. 7b). In this case, the variance of the long-term dataset was significantly higher than that for the 2014 HTS data but, once again, there was no significant difference between HTS results and either the long-term or 2014 LM data (Table 2).

Finally, the long-term average TDI4 for the River Derwent at Ebchester in County Durham, England was 60, indicating moderate status (due to enrichment with nutrients and organic matter from Consett sewage treatment works), and the river is consequently enriched with nutrients and organic matter. The average value of both LM and HTS analyses in 2014, however, was slightly lower, indicating good status (Fig. 7c), although differences were not significant (Table 2); however, both LM and HTS data fluctuate across good and moderate status, with occasional results indicating poor status (Fig. 7c). 21 of 44 LM records indicate a different classification to the long-term data, as does the 2014 LM subset, highlighting the need for caution when interpreting the differences reported in Fig. 6.

Table 2. Temporal differences between LM (TDI4) and HTS (TDI5NGS) results for three rivers in northeast England. n = number of samples in LM dataset; N.S. = not significant ($p \geq 0.05$); ** = $p < 0.01, \geq 0.001$.

		mean						
	n	TDI4		TDI5NGS	TDI5NGS v all TDI4		TDI5NGS v 2014 TDI4	
		All	2014	2014	Bartlett	t-test	Bartlett	t-test
Wear	21	40	37	34	N.S.	N.S.	N.S.	N.S.
Ehen	32	17	18	17	**	N.S.	N.S.	N.S.
Derwent	44	60	56	56	N.S.	N.S.	N.S.	N.S.



528

529 **Figure 7.** Long-term trends in diatom-based classifications in three rivers in northern England using
530 LM and HTS approaches: a. River Wear, Wolsingham; b. River Ehen, Ennerdale Bridge; c. River
531 Derwent, Ebchester. LM data are expressed as TDI4 (the current regulatory approach) whilst HTS
532 data are expressed as TDI5NGS, Horizontal lines show the position of high/good (blue),
533 good/moderate (green) and moderate/poor (orange) status class boundaries. The position of status
534 class boundaries is determined primarily by alkalinity (see Kelly et al. 2008).

Discussion

We have shown a significant correlation between LM- and HTS-based diatom metrics (Fig. 4c) despite an incomplete *rbcL* DNA reference database and observed variability in the relative abundance of certain taxa evaluated using LM and HTS. While other studies have shown strong relationships between LM and HTS metrics (see Kermarrec *et al*. 2014b; Visco et al. 2015; Zimmerman et al. 2014), the present study is much larger in scale (over 1000 samples).. Some of the other studies also showed deviations from a 1:1 relationship when comparing metric outputs generated by LM and HTS (cf. Fig. 2c). However, we have gone one step further than in previous studies, with the development of an HTS-specific diatom metric (Fig. 4c).

Though now well-established as part of the ecological assessment toolkit in Europe and beyond (Kelly 2013a; Poikane et al. 2016), diatom analysis requires highly-trained individuals to spend considerable lengths of time with expensive microscopes. There are a number of uncertainties associated with assessments (Prygiel et al. 2002; Kelly et al. 2009), a significant part of which is associated with the analytical process itself (Kahlert et al. 2009, 2012). There is, therefore, a strong case for exploring alternatives and we have demonstrated that HTS is one that shows great promise.

We set out to establish an HTS analogue of the existing diatom assessment method and, to the extent that there is 97% percent agreement to within one class (Fig. 6) we believe that we have largely succeeded. However, this result needs to be set in context: there was only 63% exact agreement (i.e. the same status class predicted by both LM and HTS). In adopting the same principles for HTS as are used for LM, we inevitably bring across all the uncertainties that are not associated with the analytical process itself (such as spatial and temporal variability in diatom assemblage composition). Furthermore, a sample-by-sample comparison in a dataset of spatially-discrete samples also has to take account of fluctuations in assessment results over time (Fig. 7). In retrospect, were appropriate data available, such assessments would need to consider “excess changes” (i.e. the proportion of samples that have changed class overall, minus the proportion expected to change class for purely stochastic reasons).

Having now established this molecular analogue of the existing diatom assessment method, we can begin to consider how to access the added value contained within the HTS data, exploiting the extra information on diversity represented in those OTUs or Amplified Sequence Variants (ASVs: Callahan et al., 2017) that have no exact equivalent in the traditional taxonomic classification (or which do not represent diatoms). So long as new metrics can be linked to legislative drivers such as the WFD, then

567 there is huge potential for HTS within ecological assessment.It is, however, important to bear in
568 mind that the "traditional" LM approach is itself an imperfect reflection of reality (albeit one with
569 which practitioners are familiar). The two approaches offer alternative views of the stream
570 ecosystem that need to be reconciled; it is rarely as simple as deciding that one method is "right" or
571 that it is "better" than the alternative.

572 **4.1 Relationship between LM and HTS data**

573 We show that the occurrence of individual species in the HTS output is different to that in
574 corresponding LM analyses (Fig. 1). To some extent, differences are predictable: larger cells with
575 multiple chloroplasts tend to have greater representation in HTS output than small single-
576 chloroplast taxa (discussed also by Vasselon et al., 2017). In the case of small, lightly-silicified taxa
577 such as *Fistulifera* and *Mayamaea,* their greater representation in HTS may be due to their valves
578 not surviving the preparation process used in LM (Zgrundo et al. 2013; Perez-Burillo et al. 2020). In
579 this respect, metabarcoding output might give a more accurate indication of the contribution made
580 by different diatom species to ecosystem processes than conventional analyses based on "cleaned"
581 diatom valves.

582 Although LM-based analysis of diatoms provides the benchmark against which metabarcoding
583 approaches are being judged, diatoms are, in most parts of Europe, proxies for the whole
584 phytobenthos community. Whilst cleaning diatoms offers greater taxonomic sensitivity compared
585 with analysis of live diatoms, this comes at the expense of information about non-diatom algae (an
586 important component of many biofilms) as well as extracellular structures such as mucilage stalks
587 and tubes, and about which individuals of which species were alive at the time of sampling (Gillett et
588 al. 2008; Kelly 2013a; Kelly et al. 2019). Moreover, methods for data analysis focus on enumeration
589 of individuals, regardless of cell size. There can be, for example, a 100× difference in the biovolume
590 of a single cell of *Achnanthidium minutissimum* compared to one of *Ulnaria ulna* (Vasselon et al.,
591 2018), yet both have equal influence on a TDI4 or TDI5LM calculation.

592 Differences in size are, to some extent, reflected by the *rbcL* data (Vasselon et al. 2018), suggesting
593 that metabarcoding output using this marker may give a more accurate indication of the relative
594 contribution of each taxon to diatom productivity than LM. In practice, different enumeration
595 concepts in HTS and LM analyses may well explain the non-linear response observed in Fig. 2c. For
596 example, *Achnanthidium minutissimum*, which has a high LM:HTS ratio (Fig. 1), tends to be very
597 abundant in low nutrient (low TDI4) sites, whilst taxa such as *Navicula lanceolata* and, in particular,
598 *Melosira varians*, which are abundant where TDI4 is high, have much lower LM:HTS ratios (Fig. 1).
599 To some extent, these individual differences balance each other out in the final metric calculations,

yielding a reasonable agreement between the two methods (Fig. 6). Nonetheless, that such explainable differences exist is a justification for deriving new indices directly from HTS data, rather than the approach adopted by Vasselon et al. (2018) of providing "correction factors" to align HTS output with LM metrics. Their approach appears to us to embed the recognised biases of LM analyses into HTS rather than treating this new technology as an opportunity to move ecological assessment methods forward. Using TDI5LM as the benchmark against which TDI5NGS is derived is a stronger approach because TDI5LM offers the optimum picture of community turnover along the pressure gradient but without making any assumptions about the behaviour of any individual species. Ideally, we would have derived TDI5NGS directly from physico-chemical data; however, problems with detection limits, particularly for phosphorus, in a large database merging records from four separate organisations, made this option less attractive.

Although differences in classifications derived from LM and HTS data were observed (Fig. 6), it is important that these results are placed in context. Fig. 6 shows variation in classifications derived from individual samples, whilst Fig. 7 shows how those samples can vary over time at a single site, often crossing status class boundaries. To some extent, the stability of a classification will depend upon the number of replicates on which this is based, and also on the distance of the mean value of the metric from a status class boundary (Kelly et al., 2009). Thus, some "noise" is to be expected in comparisons such as that shown in Fig. 6 and classifications should, ideally, be derived from means of several replicates. Second, this "noise" is only one of a number of sources of uncertainty in ecological status assessments and, in the case of the UK phytobenthos assessment, a larger systematic source of misclassification is likely to arise from weaknesses in the determination of reference values for metrics (Kelly et al., 2020).

It is also important to remember that HTS, too, has intrinsic biases, not least of which are the decisions involved in bioinformatics (Baillet et al., this issue). Although various developments exist to improve the analysis of metabarcoding data, such as open-reference clustering (Rideout et al., 2014) or graph theory based algorithms such as Swarm (Foster et al., 2016), the fundamental question of what molecular criteria define a species in micro-eukaryotic communities remains to be answered. The sequence similarity cutoff of 97% chosen in the HTS dataset represents a pragmatic solution to group sequences for ease of bioinformatic analysis. This approach has been used successfully for 18S rDNA of other protists such as Cercozoa (Fiore-Donno et al., 2018). Our own preliminary analyses indicated an intra-species diversity of rbcL that supports the cutoff at 97% sequence similarity. Rivera et al. (2017) applied a cutoff of 95% rbcL sequence similarity, though the rationale behind the choice remains elusive, highlighting the lack of consensus between sequence differences and species delineation.

At the same time, it is important to recognise that ecological status assessment forms part of the UK's legal framework of environmental regulation and as such it is sensible to ensure that principles are, as far as possible, consistent between LM and HTS. In any case, the changes involved in the transition from LM to HTS are of a smaller scale than were encountered when the reference model underpinning ecological status assessments was improved (Kelly et al., this issue), emphasising the need to keep the impact of HTS adoption in proportion. Similarly, the HTS-based metric had a slightly weaker relationship with the pressure gradient than the existing metric, a cause for concern for regulators who will use these data to make decisions about catchment management. The reason for this is not clear, but the barcode reference database, the key link between HTS output and the wider ecological knowledge base, is one aspect of the study that deserves a closer look.

4.2 The significance of the barcode reference database

Correct assignment of HTS data to the appropriate Linnaean binomial is of prime importance to the development of a viable HTS-based ecological assessment procedure that is consistent with the requirements of the WFD. The situation for diatoms is complicated by the number of new developments in the underlying taxonomy, many of which are themselves driven by the insights that molecular biology has provided. In some cases, these clarify differences between species that present challenges to traditional analyses (Trobajo et al., 2013), which in turn allow ecological differences to be unravelled (Kelly et al. 2015). In other cases, such studies throw doubt on species defined on morphological criteria alone (Kermarrec et al. 2013; Rovira et al. 2015; Duleba et al. 2016; Kahlert et al. 2019).

Despite the substantial effort that went into the development of the barcode database, it still represents only about 12 per cent of the total number of species recorded from British and Irish freshwaters. On the other hand, this list includes representatives of most of the commonly encountered taxa and should be sufficient to account for much of the variation between samples (Fig. 5). A further potential source of bias is the reliance on PCR amplification of a barcode gene which will likely have been selected based on studies performed on available reference sequences . Major "gaps" in coverage could, in theory, embed biases and perpetuate errors. Approaches such as metagenomics (or environmental sequencing) which omit PCR amplification might be better suited to detect all diatom species present, though sensitivity might be lower unless sufficient sequencing coverage is applied to detect rare taxa. Such method comparison should be considered in future studies in connection to its impact for environmental assessments.

In practice, however, Linnaean binomials provide a link between both LM and HTS data and autecological information, from which the final status assessment is derived. The assumption is that the information associated with each binomial adds substantial value to the assessment. In theory, a system based purely on OTUs or ASVs (i.e. bypassing Linnaean binomials completely) could work as efficiently (Apothéloz-Perret-Gentil et al., 2017), once it had been calibrated against the principal environmental gradients. In practice, our experiments (unpublished) with an OTU-based approach have not yielded appreciably stronger relationships and, moreover, as Article 14 of the WFD specifically addresses "Public information and consultation", we believe that it is important to explain observed changes in assessments in terms of ecosystem function, for which an understanding of the diatom species involved (their life-form characteristics, size, motility, etc) will continue to be necessary (Rimet & Bouchez, 2012; Tapolczai et al., 2016). A truly taxonomy-free approach is, in any case, inconsistent with legislation that requires assessment of "species composition".

We learned an important lesson about reference databases during the earlier phases of the UK study: that it is foolish to constrain the database to the particular region of a barcode gene that is appropriate for the current HTS technology. When we began developing an HTS approach to ecological assessment using diatoms, the current HTS technology was Roche 454 pyrosequencing and we accordingly used a barcode region of >500 bp within *rbcL*. During development of the HTS approach, however, 454 sequencing became obsolete, requiring a new barcode region to be selected that was suitable for Illumina technology. Fortunately, we had decided at the outset that for the reference database we would sequence almost full-length *rbcL*, principally so that any associated phylogenetic work (to clarify taxonomic distinctions) could be done with an adequate genetic marker. Hence when 454 sequencing was discontinued, we were able to use the redesigned barcode without having to generate new *rbcL* sequences for the reference database. We suggest from this that, until the availability and affordability of accurate long reads make the current HTS technology obsolete, the gene sequences made for the reference databases should be of a longer length than is currently necessary for HTS metabarcoding (e.g. ±full-length *rbcL*, 18S rDNA), and the DNA extracted from cultured clones should be stored securely to allow for step changes in approach.

4.3 Further development of HTS-based phytobenthos metrics

Overall, the outcomes from this study are positive: a procedure has been developed that is compatible with the leading HTS technologies. Procedures for extracting, amplifying and analysing DNA sequences in these samples have been developed and tested, and bioinformatics procedures have been devised to produce data that are compatible with outputs from current LM analyses. This,

698 in turn, has allowed the similarities and differences between the two approaches to be evaluated
699 and a new metric -a variant of the current TDI4, optimised for HTS - to be developed. This has been
700 achieved using a barcode database that still only includes a small proportion of the diatom species
701 that have been described from the UK. Although this situation is improving as more laboratories
702 contribute barcodes to online databases, the enormous diversity of diatoms (Mann &
703 Vanormelingen, 2013) means that it is unlikely that fully comprehensive coverage of all diatom
704 species will be achieved at a sufficiently high quality in the near future. It is also important not to
705 mistake breadth of coverage of morphologically-defined species with depth of coverage of
706 genotypes within complexes, the absence of which may contribute to mismatches between LM and
707 HTS, and which may convey potentially-useful information about ecosystem resilience.

708 We set out to establish an HTS analogue of the existing diatom assessment method as this ensures
709 continuity with existing methods although it also risks enshrining aspects of the process that are
710 artefacts of a particular mode of working. Indeed, even the exclusive focus on diatoms – which arose
711 because the sensitivity of this group could best be exploited by LM only at the expense of all other
712 algae in a sample – deserves to be questioned (Kelly et al., 2015). Current methods are optimised to
713 give strong correlations with water chemistry (Poikane et al., 2016) and will miss shifts in the balance
714 of different algal phyla, which in turn may impart valuable information on ecosystem functioning
715 (e.g. Schneider & Lindstrøm, 2011). Whilst it is possible that new metrics could be developed,
716 drawing on the potential of HTS to capture diversity along stressor gradients, most proposals to date
717 (e.g. Baird & Hajibabei 2012) are still variants of the "name-and-count" approach of traditional
718 applied ecology and rarely address the many sources of uncertainty beyond analytical precision.
719 The major limitation of current benthic algae and plant-based assessments, however, is their limited
720 capability to evaluate biomass and predict secondary effects of eutrophication (Kelly 2013a).

721 The aspiration of producing HTS analogues of existing techniques was, we believed, a sensible
722 starting point as it forces a close examination of the relationship between HTS and "traditional" data
723 (see, also, Hering et al., 2018). To the extent that there is 97% percent agreement to within one
724 class (Fig. 6) we believe that we have largely succeeded in producing an HTS analogue. However,
725 this result needs to be set in context: there was only 63% exact agreement (i.e. the same status class
726 predicted by both LM and HTS). In adopting the same principles for HTS as are used for LM, we
727 inevitably bring across all the uncertainties that are not associated with the analytical process itself
728 (such as spatial and temporal variability in diatom assemblage composition).

729 Having now established this molecular analogue of the existing diatom assessment method, we can
730 begin to consider how to access the added value contained within the HTS data, exploiting the extra

information on diversity represented in those OTUs that have no exact equivalent in the traditional taxonomic classification. So long as new metrics can be linked to legislative drivers such as the WFD, then there is huge potential for HTS within ecological assessment.

Given the latest advancements especially with regards to long read technology a myriad of new possibilities arise to refine ecological assessments using HTS (e.g. Jamy et al., 2019). Seen from a bureaucrat's perspective, however, the relationships shown in Fig. 4 translate into an appreciable amount of change to established classifications (Fig. 6) with little perceived benefit when seen from the perspective of stakeholders and managers. To some extent, this highlights shortcomings in the ability of the current UK regulatory framework to manage change (Kelly, 2019). A final lesson from this project is that adoption of HTS is not a simple transaction in which LM is replaced by a better technology, but requires a deeper level of institutional transformation than had hitherto been anticipated.

Acknowledgements

This research received financial support from the following UK regulatory agencies: Environment Agency, Scottish Environmental Protection Agency, Natural Resources Wales and Northern Ireland Environment Agency. We would like to thank all operational staff from the regulatory authorities for the collection of diatom samples for molecular analysis and Environment Agency staff for carrying out the light microscopy analysis on the calibration dataset samples. We also thank Sarah Pritchard (Beacon Biological) for help with preparing diatom samples. The Royal Botanic Garden Edinburgh (RBGE) is supported by the Scottish Government's Rural and Environment Science and Analytical Services Division. We also thank two reviewers for detailed comments on the manuscript.

References

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., Pawlowski, J., 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Molecular Ecology Resources 17, 1231–1242. https://doi.org/10.1111/1755-0998.12668

Baird, D.J., Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular Ecology 21, 2039–2044.

760 Bailet, B., Bouchez, A., Franc, A., Frigerio, J.-M., Keck, F., Karjalainen, S.-M., Rimet, F., Schneider, S.C.,
761 Kahlert, M., 2019. Molecular versus morphological data for benthic diatoms biomonitoring in
762 Northern Europe freshwater and consequences for ecological status. Metabarcoding and
763 Metagenomics 3, 21–35. https://doi.org/10.3897/mbmg.3.34002

764 Bailet, B, Apothéloz-Perret-Gentil, L., Baričević, A., Chonova, T., Franc, A., Frigerio, J.M., Kelly,
765 M.G., Morai, D.,Pfannkuchen, M., Proft, S., Ramon, M., Vasselon, V., Zimmermann, J., Kahlert, M.,
766 2020, Diatom DNA metabarcoding for ecological assessment: comparison among bioinformatics
767 pipelines used in six European countries reveals the need for standardization. *Ecological Indicators*
768 (this issue)

769 Birks, H.J.B., Line, J.M., Juggins, S., Stevenson, A.C., Ter Braak, C.J.F., 1990. Diatoms and pH
770 reconstruction. Philosophical Transactions of the Royal Society of London, Series B, 327, 263–278.

771 Birks, H. J. B., Simpson, G. L., 2013. "Diatoms and pH reconstruction" (1990) revisited. Journal of
772 Paleolimnology 49, 363–371. https://doi.org/10.1007/s10933-013-9697-7

773 Caporas, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N.,
774 Gonzalez Pena, A., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley,
775 R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J. , Sevinsky, J.R., Turnbaugh.
776 P,J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis
777 of high-throughput community sequencing data. Nature Methods 7, 335–336

778 Callahan, B., McMurdie, P., Holmes, S., 2017. Exact sequence variants should replace operational
779 taxonomic units in marker-gene data analysis. ISME Journal 11, 2639–2643.

780 Carballeira, R., Trobajo, R., Leira, M., Benito, X., Sato, S., Mann, D. G., 2017. A combined
781 morphological and molecular approach to *Nitzschia varelae* sp. nov., with discussion of symmetry in
782 Bacillariaceae. European Journal of Phycology 52, 342–359.
783 https://doi.org/10.1080/09670262.2017.1309575

784 CEN, 2014a. Water quality – Guidance standard for the routine sampling and pretreatment of
785 benthic diatoms from rivers. EN 13946: 2014. Comité European de Normalisation, Geneva.

786 CEN, 2014b. Water quality – Guidance standard for the identification, enumeration and
787 interpretation of benthic diatom samples from running waters. EN 14407:2014. Comité European de
788 Normalisation, Geneva.

789 Chonova, T., Kurmayer, R., Rimet, F., Labanowski, J., Vasselon, V., Keck, F., Illmer, P & Bouchez, A.,
790 2019. Benthic diatom communities in an alpine river impacted by waste water treatment effluents
791 as revealed using dna metabarcoding. Frontiers in Microbiology 10, 653.
792 Https://doi.org/10.3389/fmicb.2019.00653

793 Duleba, M., Kiss, K.T., Földi, A., Kovács, J., Borojević, K.K., Molnár, L.F., Ács, É., 2015. Morphological
794 and genetic variability of assemblages of *Cyclotella ocellata* Pantocsek / *C. comensis* Grunow
795 complex (Bacillariophyta, Thalassiosirales)., Diatom Research 30, 283–306.

796 Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST, Bioinformatics 26,
797 2460–2461.

798 European Union, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23rd
799 October 2000 Establishing a Framework for Community Action in the Field of Water Policy Official
800 Journal of the European Communities, European Commission, Brussels (2000) (22 December, L 327,
801 1-80.

802 European Union, 2008. Commission Decision of 30 October 2008 establishing, pursuant to Directive
803 2000/60/EC of the European Parliament and of the Council, the values of the Member State
804 monitoring system classifications as a result of the intercalibration exercise. Official Journal of the
805 European Union L 332, 20–44.

806 European Union, 2013. Commission Decision of 20 September 2013 establishing, pursuant to
807 Directive 2000/60/EC of the European Parliament and of the Council, the values of the Member
808 State monitoring system classifications as a result of the intercalibration exercise and repealing
809 Decision 2008/915/EC. Official Journal of the European Union Series L 266, 1–47.

810 Evans, K.M., Wortley, A.H., Simpson, G.E., Chepurnov, V.A., Mann, D.G., 2008. A molecular
811 systematic approach to explore diversity within the *Sellaphora pupula* species complex
812 (Bacillariophyta). Journal of Phycology 44, 215–231.

813 Fiore-Donno, A. M., Rixen, C., Rippin, M., Glaser, K., Samolov, E., Karsten, U., … Bonkowski, M. 2018.
814 New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput
815 environmental diversity surveys, with emphasis on worldwide biological soil crusts. Molecular
816 Ecology Resources. https://doi.org/10.1111/1755-0998.12729

817 Forster, D., Dunthorn, M., Stoeck, T., Mahé, F., 2016. Comparison of three clustering approaches for
818 detecting novel environmental microbial diversity. PeerJ. https://doi.org/10.7717/peerj.1692

819 Fourtanier, E., Kociolek, J.P., 1999. Catalogue of the diatom genera. Diatom Research 14, 1–190.

820 Gillett, N., Pan, Y., Parker, C., 2009. Should only live diatoms be used in the bioassessment of small
821 mountain streams? Hydrobiologia 620, 135-147. https://doi.org/10.1007/s10750-008-9624-5

822 Guillard, R.R.L., Lorenzen, C.J., 1972. Yellow-green algae with chlorophyllide c. Journal of Phycology
823 8, 10–14.

824 Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program
825 for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

826 Hartley, B., Barber, H.G., Carter, J.R., 1996. An Atlas of British Diatoms. Biopress, Bristol.

827 Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakere, S., Hänfling, B.,
828 Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., Kelly, M.G., 2018.
829 Implementation options for DNA-based identification into ecological status assessment under the
830 European Water Framework Directive. Water Research 138, 192–205.

831 Hofmann, G., Werum, M., Lange-Bertalot, H., 2011. Diatomeen im Süsswasser-Benthos von
832 Mitteleuropa. A.R.G. Gantner Verlag K.G., Ruggell.

833 Jamy, M, Foster, R, Barbera, P, Czech, L., Kozlov, A., Stamatakis, A., Bending, G., Hilton, S., Bass, D.,
834 Burki, F., 2019. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and
835 taxonomically resolve environmental diversity. Mol Ecol Resour. 20: 429-443.
836 https://doi.org/10.1111/1755-0998.13117

837 Jones, H.M., Simpson, G.E., Stickle, A.J., Mann, D.G., 2005. Life history and systematics of *Petroneis*
838 (Bacillariophyta) with special reference to British waters. European Journal of Phycology 40, 61–87.

839 Joshi, N.A., Fass, J.N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ
840 files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

841 Juggins, S., 2015. rioja: Analysis of Quaternary science data. R package version 0.9-6. http://cran.r-
842 project.org/package=rioja.

843 Kahlert, M., Kelly, M.G., Albert, R.-L., Almeida, S., Bešta, T., Blanco, S., Denys, L., Ector, L., Fránková,
844 M., Hlúbiková, D., Ivanov, P., Kennedy, B., Marvan, P., Mertens, A., Miettinen, J., PIcinska-
845 Fałtynowicz, J., Rosebery, J., Tornés, E., Van Dam, H., Vilbaste, S., Vogel, A., 2012. Identification is a
846 minor source of uncertainty in diatom-based ecological status assessments on a continent-wide
847 scale: results of a European ring-test. Hydrobiologia 695, 109–124.

Kahlert, M., Albert, R.L., Anttila, E.L., Bengtsson, R., Bigler, C., Eskola, T., Galman, V., Gottschalk, S., Herlitz, E., Jarlman, A., Kasperoviciene, J., Kokocinski, M., Luup, H., Miettinen, J., Paunksnyte, I., Piirsoo, K., Quintana, I., Raunio, J., Sandell, B., Simola, H., Sundberg, I., Vilbaste, S., Weckstrom, J., 2009. Harmonization is more important than experience – results of the first Nordic-Baltic diatom intercalibration exercise 2007 (stream monitoring). Journal of Applied Phycology 21, 471–482.

Kahlert, M., Kelly, M.G., Mann, D.G., Rimet, F., Sato, S., Bouchez, A., Keck, F., 2019. Connecting the morphological and molecular species concepts to facilitate species identification within the genus *Fragilaria* (Bacillariophyta). Journal of Phycology 55, 948–970.

Kelly, M.G., 2013a. Data rich, information poor? Phytobenthos assessment and the Water Framework Directive. European Journal of Phycology 48, 437–450.

Kelly, M.G., 2013b. Building capacity for ecological assessment using diatoms in UK rivers. Journal of Ecology and Environment 36, 89-94.

Kelly, M.G., 2019. Adapting the (fast-moving) world of molecular ecology to the (slow-moving) world of environmental regulation: lessons from the UK diatom metabarcoding exercise. Metabarcoding and Metagenomics 3: 127–135. https://mbmg.pensoft.net/article/39041/

Kelly, M.G., Bennion, H., Burgess, A., Elllis, J., Juggins, S., Guthrie, R., Jamieson, B.J., Adriaenseens, V., Yallop, M.L., 2009. Uncertainty in ecological status assessments of lakes and rivers using diatoms Hydrobiologia 633, 5-15.

Kelly, M.G., Boonham, N., Juggins, S., Kille, P., Mann, D.G., Pass, D., Sapp, M., Sato, S., Glover, R.,2018. A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. Science Report SC140024/R, Environment Agency, Bristol.

Kelly M.G., A. Cazaubon, E. Coring, A. Dell'Uomo, L. Ector, B. Goldsmith, H. Guasch, J. Hürlimann, A. Jarlman, B. Kawecka, J. Kwandrans, R. Laugaste, E.-A. Lindstrøm, M. Leitao, P. Marvan, J. Padisák, E. Pipp, J. Prygiel, E. Rott, S. Sabater, H. van Dam, J. Vizinet, 1998. Recommendations for the routine sampling of diatoms for water quality assessments in Europe. Journal of Applied Phycology 10, 215-224.

Kelly, M.G., Juggins, S., Guthrie, R., Pritchard, S., Jamieson, J., Rippey, B., Hirst, H., Yallop, M., 2008. Assessment of ecological status in U.K. rivers using diatoms. Freshwater Biology 53, 403–422.

Kelly, M.G., Juggins, S., Phillips, G.P., Willby, N.J., 2020. Re-evaluating expectations for river phytobenthos assessment and understanding the relationship with macrophytes. Ecological Indicators 117: 106582

879 Kelly, M.G., King, L., Yallop, M.L., 2019. As trees walking: pros and cons of partial sight in the analysis
880 of stream biofilms. Plant Ecology and Evolution 152, 120-130.

881 Kelly, M.G., Schneider, S.C., King. L., 2015. Customs, habits and traditions: the role of non-scientific
882 factors in the development of ecological assessment methods. WIRES Water 2: 159-165.

883 Kelly, M.G., Trobajo, R., Rovira, L., Mann, D.G., 2015. Characterizing the niches of two very similar
884 *Nitzschia* species and implications for ecological assessment. Diatom Research 30: 27–33.

885 Kelly, M.G., Willby, N.J., Phillips, G., Benstead, R., 2013. The integration of macrophyte and
886 phytobenthos surveys as a single biological quality element for the Water Framework Directive,
887 Science Report: SC070034/T4, Environment Agency, Bristol.

888 Kermarrec, L., Bouchez, A., Rimet, F., Humbert, J.-F., 2014. First evidence of the existence of semi-
889 cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kützing) Kützing
890 complex (Bacillariophyta). Protist 164, 686–705.

891 Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J.M., Humbert, J.F., Bouchez, A., 2014. A
892 next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshwater
893 Science 33, 349–363. https://doi.org/10.1086/675079.

894 Krammer, K., Lange-Bertalot. H., 1986. Die Süsswasserflora von Mitteleuropa 2: Bacillariophyceae. 1
895 Teil: Naviculaceae. Stuttgart: Gustav Fischer-Verlag.

896 Krammer, K., Lange-Bertalot, H., 1997. Die Süßwasserflora von Mitteleuropa, 2. Bacillariophyceae.
897 Teil 2: Bacillariaceae, Epithemiaceae, Surirellaceae. 2te Auflage, mit einem neuen Anhang. Gustav
898 Fischer Verlag, Stuttgart.

899 Krammer, K., Lange-Bertalot, H., 2000. Die Süsswasserflora von Mitteleuropa 2: Bacillariophyceae. 3
900 Teil: Centrales, Fragilariaceae, Eunotiaceae. 2nd edition. Gustav Fischer Verlag, Stuttgart.

901 Krammer, K., Lange-Bertalot, H., 2004. Süsswasserflora von Mitteleuropa 2, Bacillariophyceae. Teil 4
902 : Achnanthaceae. Kritische Ergänzungen zu Achnanthes s.l., Navicula s. str., Gomphonema.Spektrum
903 Akademischer Verlag/Gustav Fischer, Heidelberg.

904 Kuroiwa, T., Suzuki, T., Ogawa, K., Kawano, S., 1981. The chloroplast nucleus: distribution, number,
905 size, and shape, and a model for the multiplication of the chloroplast genome during chloroplast
906 development. Plant and Cell Physiology 22, 381–396.

907 Lin, L. I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45,
908 255–268.

909 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., Brennan, G., Bush,
910 A., Canard, E., Cordier, T., Creer, S., Curry, R.A., David, P., Dumbrell, A.J., Gravel, D., Hajibabaei, M.,
911 Hayden, B., van der Hoorn, B., Jarne, P., Jones, I., Karimi, B., Keck, F., Kelly, M., Knot, I.E., Krol, L.,
912 Massol, F., Monk, W.A., Murphy, J., Pawlowski, J., Poisot, T., Porter, T.M., Randall, K.C., Ransome, E.,
913 Ravigné, V., Raybould, A., Robin, S., Schrama, M., Schatz, B., Tamaddoni-Nezhad, A., Trimbos, K.B.,
914 Vacher, C., Vasselon, V., Wood, S., Woodward, G. & Bohan, D.A., 2020. Key questions for next-
915 generation biomonitoring. Frontiers in Environmental Science 7, 197.
916 https://doi.org/10.3389/fenvs.2019.00197

917 Mann, D.G., Vanormelingen, P., 2013.) An inordinate fondness? The number, distributions, and
918 origins of diatom species. Journal of Eukaryotic Microbiology 60, 414–420.

919 Mann, D.G., Sato, S., Trobajo, R., Vanormelingen, P., Souffreau, C., 2010. DNA barcoding for species
920 identification and discovery in diatoms. Cryptogamie, Algologie 31, 557–577.

921 Martin, M., 2011, Cutadapt removes adapter sequences from high-throughput sequencing reads.
922 EMBnet Journal 17, 10–12.

923 Mayama, S., Mayama, N., Shihira-Ishikawa, I., 2004. Characterization of linear-oblong pyrenoids with
924 cp-DNA along their sides in *Nitzschia sigmoidea* (Bacillariophyceae). Phycological Research 52, 129–
925 139.

926 Oksanen, J., Kindt, R., Legendre, P., O'Hara, R.B., 2006. vegan: Community Ecology Package Version
927 1.8-5. HTTP://CRAN.R-PROJECT.ORG/.

928 Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S.S., Cepicka, I., Decelle, J.,
929 Dunthorn, M., Fiore-Donno, A.M., Gile, G.H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P.J., Kostka,
930 M., Kudryavtsev, A., Lara, E., Lukeš, J., Mann, D.G., Mitchell, E.A.D., Nitsche, F., Romeralo, M.,
931 Saunders, G.W., Simpson, A.G.B., Smirnov, A.V., Spouge, J., Stern, R.F., Stoeck, T., Zimmermann, J.,
932 Schindel, D., de Vargas, C.,2012. CBOL Protist Working Group: Barcoding eukaryotic richness beyond
933 the animal, plant and fungal kingdoms. PLoS Biology 10, e1001419

934 Peres-Neto, P., Jackson, D., 2001. How well do multivariate data sets match? The advantages of a
935 Procrustean superimposition approach over the Mantel test. Oecologia 129, 169–178.

936 Pérez-Burillo, J., Trobajo, R., Vasselon, V., Rimet, F., Bouchez, A. & Mann, D.G., 2020. Evaluation and
937 sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of Mediterranean rivers.
938 Science of the Total Environment 727, 138445.

939 Poikane, S., Kelly, M.G., Cantonati, M., 2016. Benthic algal assessment of ecological status in
940 European lakes and rivers: challenges and opportunities. Science of the Total Environment 568, 603–
941 613.

942 Prygiel, J., Carpentier, P., Almeida, S., Coste, M., Druart, J.-C., Ector, L., Guillard, D., Honoré, M.-A.,
943 Iserentant, R., Ledeganck, P., Lalanne-Cassou, C., Lesniak, C., Mercier, I., Moncaut, P., Nazart, M.,
944 Nouchet, N., Peres, F., Peeters, V., Rimet, F., Rumeau, A., Sabater, S., Straub, F., Torrisi, M.,
945 Tudesque, L., Van der Vijver, B., Vidal, H., Vizinet, J., Zydek, N., 2002. Determination of the biological
946 diatom index (IBD NF T 90-354): results of an intercomparison exercise. Journal of Applied Phycology
947 14, 27–39.

948 R Development Core Team, 2006. R: A language and environment for statistical computing. R
949 Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-
950 project.org.

951 Rideout, J. R., He, Y., Navas-Molina, J. A., Walters, W. A., Ursell, L. K., Gibbons, S. M., …Gregory
952 Caporaso, J., 2014. Subsampled open-reference clustering creates consistent, comprehensive OTU
953 definitions and scales to billions of sequences. PeerJ. https://doi.org/10.7717/peerj.545

954 Rimet, F., Abarca, N., Bouchez, A., Kusber, W.-F., Jahn, R., Kahlert, M., Keck, F., Kelly, M.G., Mann,
955 D.G., Piuz, A., Trobajo, R., Tapolczai, K., Valentin, V., Zimmerman, J., 2018. The potential of high
956 throughput sequencing (HTS) of natural samples as a source of primary taxonomic information for
957 reference libraries of diatom barcodes. Fottea 18, 37-54.

958 Rimet, F., Gusev, E., Kahlert, M., Kelly, M.G., Kulikovskiy, M., Maltsev, Y., Mann, D.G., Pfannkuchen,
959 M., Trobajo, R., Vasselon, V., Zimmermann, J., Bouchez, A., 2019. Diat.barcode, an open-access
960 curated barcode library for diatoms. Scientific Reports 9: 15116.

961 Rivera, S. F., Vasselon, V., Jacquet, S., Bouchez, A., Ariztegui, D., & Rimet, F. (2018). Metabarcoding
962 of lake benthic diatoms: from structure assemblages to ecological assessment. Hydrobiologia 807,
963 37-51. https://doi.org/10.1007/s10750-017-3381-2

964 Rovira, L., Trobajo, R., Sato, S., Ibáñez, C., Mann, D.G., 2015. Genetic and Physiological Diversity in
965 the Diatom *Nitzschia inconspicua*. Journal of Eukaryotic Microbiology 62, 815–832.

966 Round, F.E., Crawford, R.M., Mann, D.G., 1990. The diatoms: Biology and morphology of the genera.
967 Cambridge University Press, Cambridge.

968 Schneider S. C., Lindstrøm, E. A., 2011. The periphyton index of trophic status PIT: A new
969 eutrophication metric based on non-diatomaceous benthic algae in Nordic rivers. Hydrobiologia 665,
970 143-155.

971 Stevenson, M.,2010. epiR: Functions for analysing epidemiological data. Version 0.9-27 [available on
972 internet at http://cran.r-project.org/web/packages/epiR].

973 ter Braak,C.J.F., 1986. Canonical Correspondence Analysis: a new eigenvector technique for
974 multivariate direct gradient analysis. Ecology 67, 1167–1179.

975 Trobajo, R., Clavero, E., Chepurnov, V.A., Sabbe, K., Mann, D.G., Ishihara, S., Cox, E.J., 2009.
976 Morphological, genetic and mating diversity within the widespread bioindicator *Nitzschia palea*
977 (Bacillariophyceae). Phycologia 48, 443–459.

978 Trobajo, R., Mann, D.G., 2019. A rapid cleaning method for diatoms. Diatom Research 34, 115–124.

979 Trobajo, R., Rovira, L., Ector, L., Wetzel, C.E., Kelly, M.G., Mann, D.G., 2013. Morphology and identity
980 of some ecologically important small Nitzschia species. Diatom Research 28, 37–59

981 UK TAG, 2014a. UK TAG river assessment method: macrophytes and phytobenthos: phytobenthos –
982 Diatoms for Assessing River and Lake Ecological Quality (River DARLEQ2).
983 http://www.wfduk.org/resources/rivers-phytobenthos

984 Vasselon, V., Bouchez, A., Rimet, F., Jacquet, S., Trobajo, R., Corniquel, M., Tapolczai, K., Domaizon,
985 I., 2018. Avoiding quantification bias in metabarcoding: application of a cell biovolume correction
986 factor in diatom molecular biomonitoring. Methods in Ecology and Evolution 9, 1060–1069.

987 Visco, J.A., Apothéloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillett, L., Pawlowski, J., 2015.
988 Environmental monitoring: inferring diatom index from next-generation sequencing data.
989 Environmental Science and Technology 9, 7597–7605.

990 Whitton, B.A., John, D.M., Johnson, L.R., Boulton, P.N.G., Kelly, M.G., Haworth, E.Y., 1998. A coded
991 list of freshwater algae of the British Isles. LOIS publication number 222. Institute of Hydrology,
992 Wallingford.

993 Woodward, G., Gray, C., Baird, D.J., 2013. Biomonitoring for the 21st Century: new perspectives in
994 an age of globalisation and emerging environmental threats. Limnetica 32, 159–174.

995 Zhang, J., Kobert, K., Flouri, T., Stamatakis, A. 2014. PEAR: a fast and accurate Illumina Paired-End
996 read merger. Bioinformatics 30, 614–620.

997 Zgrundo, A., Lemke, P., Pniewski, F., Cox, E.J., Latala, A., 2013. Morphological and molecular
998 phylogenetic studies on *Fistulifera saprophila*. Diatom Research 28, 431–443.

999 Zimmermann, J., Glöckner, G., Jahn, R., Enke, N., Gemeinholzer, B., 2015. Metabarcoding vs.
1000 morphological identification to assess diatom diversity in environmental studies. Molecular Ecology
1001 Resources https://doi.org/10.1111/1755-0998.12336

1002 Zuur, A., Ieno, E., Smith, G., 2007. Analysing Ecological Data. Springer.

1003

1004 Supplementary material

1005 Kelly_et_al_2020_Tables_S1_&_S2.xlsx
1006 Table S1. List of diatom taxa with their associated sensitivity values in the TDI4 (current
1007 implementation of Trophic Diatom for light microscopy), TDI5LM (updated version for light
1008 microscopy) and TDI5NGS (new version optimised for high throughput sequencing).

1009 Table S2. Validated taxonomy of diatom species used for OTU classification. This table provides an
1010 identifier (ID) for each sequence linked to its taxonomic classification. Where appropriate a linkage
1011 to the diatom diatom species codes used in water classification is also provided. Non-diatoms
1012 species, such as those derived from Xanthophyceae, are provided but not validated at lower
1013 taxonomic levels.

1014 Kelly_et_al_2020_diatom_seqeunces.phy
1015 Kelly_et_al_2020_diatom_sequences.fasta
1016 Alignment files for the RbcL region used for OTU classication Nucleotide alignment files are provided
1017 in Phylip and fasta formats respectively. The sequence identifier cross-reference to the Taxaminic
1018 database file provided in supplementary material XX. Sequences have been phylogenetically ordered
1019 using a Neighbor-Joining consensus tree to assist visualisation.

1020
1021