Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity

Bista, Iliana; Carvalho, Gary; Walsh, Kerry; Seymour, Mathew; Hajibabaei, Mehrdrad; Lallias, Delphine; Christmas, Martin; Creer, Simon

Nature Communications

DOI:
10.1038/ncomms14087

Published: 01/01/2017

Peer reviewed version

Title

Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity.

Authors

Iliana Bista[1], Gary R. Carvalho[1], Kerry Walsh[2], Mathew Seymour[1], Mehrdad Hajibabaei[3], Delphine Lallias[4], Martin Christmas[2], Simon Creer[1*]

[1]Bangor University, School of Biological Sciences, Molecular Ecology and Fisheries Genetics Lab, Bangor, Gwynedd, LL57 2UW

[2]Environment Agency, Horizon House, Deanery Rd, Bristol, UK, BS1 5AH

[3]University of Guelph, Department of Integrative Biology & Biodiversity Institute of Ontario, Guelph, Ontario, Canada, N1G 2W1

[4]GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

***Corresponding author**

Simon Creer, Bangor University, School of Biological Sciences, Molecular Ecology and Fisheries Genetics Laboratory, Bangor, Gwynedd, LL57 2UW, s.creer@bangor.ac.uk, +44(0)1248 382302

Abstract

The use of environmental DNA (eDNA) in biodiversity assessments offers a step-change in sensitivity, throughput and simultaneous measures of ecosystem diversity and function. There remains, however, a need to examine eDNA persistence in the wild through simultaneous temporal measures of eDNA and biota. We used metabarcoding of two markers of different lengths, derived from an annual time-series of aqueous lake eDNA to examine temporal shifts in ecosystem biodiversity and in an ecologically important group of macroinvertebrates (Diptera: Chironomidae). The analyses allow different levels of detection and validation of taxon richness and community composition (β-diversity) through time, with shorter eDNA fragments dominating the eDNA community. Comparisons between eDNA, community DNA, taxonomy and UK species abundance data further show significant relationships between diversity estimates derived across the disparate methodologies. Our results reveal the temporal dynamics of eDNA and validate the utility of eDNA metabarcoding for tracking seasonal diversity at the ecosystem scale.

Introduction

The maintenance of biodiversity underpins the stability of ecosystem processes in constantly changing environments [1]. Consequently, biodiversity loss not only affects ecosystem function and services, but also society as a whole [2]. One major impediment for elucidating the relationship between biodiversity and ecosystem health is a need for robust and detailed understanding of biodiversity processes and dynamics in time and space [3]. To halt or reverse contemporary species loss and habitat degradation, there is a need for increasingly reliable and cost effective methods for biodiversity assessment, since widely employed traditional approaches fall short in many cases [4]. Currently, species identification of individuals at immature life stages and among closely related species is difficult and requires high-level, labour-intensive taxonomic expertise, thereby rendering large scale ecosystem-wide assessments expensive, time consuming and potentially unrepresentative of the ecosystem sampled [5]. However, recent advances in molecular detection techniques, most notably the application of environmental DNA (eDNA), offer exciting new opportunities to improve existing biodiversity assessment procedures.

Environmental DNA (eDNA) is DNA extracted directly from an environment sample (e.g., water, soil or air), without prior isolation of the organisms themselves [6]. Sources of eDNA include sloughed skin cells, urine, faeces, saliva or other bodily secretions [7], and consist of both free molecules (extracellular DNA) and free cells [8]. Furthermore, eDNA collected from water samples has highly sensitive detection capability and is non-invasive to the sampled biota[9], thereby potentially improving environmental management and assessment of freshwater ecosystems [4,10].

Previous work with eDNA of aquatic invertebrates is dominated by PCR-based approaches, which are limited in assessing biodiversity [11–13]. However, high throughput sequencing (HTS) applications, such as metabarcoding, are already advancing prospects in ecology [14], offering comprehensive and efficient tools for measuring and assessing total biodiversity [15]. High throughput sequencing has successfully been used for sequencing whole communities of invertebrates (bulk samples) [16–18], though only a few studies have employed metabarcoding of aqueous eDNA [19,20] , and even fewer for invertebrates [3]. Most aqueous eDNA studies have focused on macroorganisms, including fish and amphibians [19–21], with limited focus on arthropods [22,23]. Nevertheless, the combination of HTS and eDNA is poised to become a prominent tool for ecosystem assessment [10,22] by simultaneously assessing a plethora of organisms, including associated organism interactions, with a throughput sufficient for rapid whole community assessment.

Regardless of the increasing number of eDNA studies, several factors of eDNA research demand clarification, including persistence of eDNA [24]. Persistence of eDNA is the time that eDNA remains detectable (e.g., in the water) after removal or loss of the organism from the environment, which influences the timeframe for biodiversity assessment [6]. Investigating the temporal relationship between community DNA [25] and eDNA is vital, since accurate (extant) biodiversity assessment requires detection of contemporary, and ecologically relevant, biodiversity. The persistence of eDNA for several different species has been studied mainly in artificial systems, including aquaria and mesocosms [6,11,22,26]. Notably, persistence of short eDNA fragments, in artificial environments, was found to vary between days to weeks after removal of the study organisms, depending upon biotic and abiotic factors [27].

Species identity by eDNA is typically undertaken by detection of short DNA fragments [7], a practise possibly influenced by ancient DNA work, which utilises highly fragmented DNA [28]. For the detection of rare and evasive species, short DNA fragments might indeed increase detection, although with some risk of errors if not properly analysed. Possible biases when using short fragments include inadvertently sampling old eDNA fragments which have demonstrated remarkable persistence [8], especially when bound to sediments where the degradation rate is slower, due to protection of DNA molecules and inactivation of extracellular nucleases [27]. Conversely, DNA fragments of several hundred base pairs length are less likely to persist after release into the environment due to rapid degradation [29] and may represent a less abundant, but more contemporary, biodiversity signal [30].

While the ecological value of collecting temporal data is established, most ecological studies focus on spatial data [31]. Similarly, many existing eDNA studies have focused on spatial detection, such as early detection of invasive species [11,32] and presence of rare, or endangered species [33]. Temporal estimates have been relatively neglected in eDNA studies (but see [33] for repeated seasonal sampling), and an understanding of temporal relationships between eDNA and community biodiversity remains a key knowledge gap [3]. Additionally there are no published studies, to our knowledge, employing temporally collected data that incorporate seasonal variation across an annual cycle from aqueous eDNA for ecosystem-wide biodiversity level analysis.

Furthermore, overall ecosystem biodiversity characterisation, using indicator taxonomic groups, can facilitate comparisons between taxonomically identified biodiversity over time (e.g. collection of invertebrate samples) and eDNA detection. One such indicator group is the Chironomidae or non-biting midges (Diptera: Chironomidae), which exhibit specialised

responses to ecological stressors and are acknowledged as one of the most informative macroinvertebrate groups for monitoring lake ecosystem health [34,35] . Importantly, samples can be collected after adult emergence in the form of shed skins of the pupae (pupal exuviae) that float on the water surface. The exuviae technique allows for integrated sampling of lake ecosystems from all aquatic microhabitats of the lake, and sample identification can yield insights on ecosystem-wide biodiversity [34].

Accordingly, here we (a) investigate whether metabarcoding of lake eDNA is effective for the detection of community diversity and temporal shifts in an ecologically important sentinel group of macroinvertebrates, via comparison to the molecular and morphological analysis of chironomid exuvial bulk samples; (b) investigate the use of eDNA analyses for characterising whole-ecosystem biodiversity patterns; and (c) explore the effects of amplicon length on detection of contemporary diversity. We show that freshwater lake eDNA analyses capture seasonally coherent biodiversity patterns across the tree of life and that shorter fragments of eDNA dominate natural ecosystems. Moreover, species incidence measured by metabarcoding of eDNA and DNA derived from communities overlap substantially with traditional taxonomic assessment. Collectively, we examine the ecological relevance of eDNA by exploring mechanisms underpinning the temporal dynamics of eDNA and the biological community at the ecosystem scale in nature.

Results

Sequencing results. After stringent filtering and quality control, 13,100,236 reads were obtained for: (1) the full-length COI barcoding region (658bp) (amplicon COIF 6,659,598 reads) and (2) a 235bp fragment on the 5'region of the COI barcoding region (amplicon COIS

6,440,638 reads), from 32 samples comprising 16 eDNA and 16 invertebrate community DNA samples. Data for these two amplicons were obtained from a larger dataset including additional amplicon libraries, sequenced on two lanes of MiSeq. Overall, the eDNA samples (extracted from filtered water samples) achieved good sequence coverage (mean number of reads per sample (±SD): COIF: 269,769 ± 57,427; COIS: 259,723 ± 85,437) (for exact number of reads per sample, see Supplementary Table 1). Some of the community DNA samples that contained only small amounts of pupal exuviae resulted in a lower number of reads for both amplicons.

Control samples. During PCR screening of negative controls, no band (no amplification) was observed on agarose gels. Regardless of no visual proof of amplification, each sample was sequenced and a very low number of reads was returned. After PCR and sequencing of the negative control samples, COIS detected only two OTUs, which were BLAST-identified as bacteria. For COIF, again only two OTUs were detected, identified as Gastropoda and Diptera. The Gastropoda OTU was represented by 240 reads in one of the controls while the Dipteran OTU was only represented by 10 reads in total across all types of negative controls.

The positive controls yielded good results for both amplicons, with 547,730 (COIS) and 393,341 (COIF) reads after quality control. Detection success was 100% for COIS (all 30 species detected) and 87% for COIF (26 species detected). Amongst the species that were not detected was a mayfly species (*E. danica*), which also failed to amplify and sequence during individual barcoding of specimens, using the Folmer primers (Supplementary Table 2). BLAST identification and screening of positive control reads resulted in >99.9% of the reads being assigned to the target species known to be present in the positive control. The

relative abundance of OTUs found in the positive control which were attributed to non-target taxa was 0.026% for the COIS and 0.007% for the COIF (Supplementary Table 3).

Abundance filtering and rarefaction analysis. Following investigations of how screening different levels of abundance of rare OTUs affected overall OTU richness (including no filtering, and removal of OTUs that were present at less than 0.01% and 0.02%), a filtering level of 0.01% was set for all ecological analyses. Removal of OTUs present at less than 0.01% yielded equivalent levels of OTU genus richness for the community DNA (37 genera) and eDNA (43 genera) according to year 2014 Chironomidae records of Llyn Padarn (31 genera) (Fig. 1). Furthermore, filtering of reads below 0.01% was within the limits of a small number of non-target reads detected in the positive control samples. The genus richness comparisons employed COIS data to ensure comparability between eDNA and community DNA for the Chironomidae below. According to the analysis of OTU accumulation curves (Supplementary Figs. 1 and 2) versus sequence coverage, a rarefaction depth of 57,869 reads was applied across all water samples (Supplementary Fig. 1a). To subsample Animalia OTUs in our samples a rarefaction depth of 24,914 reads per sample was used (Supplementary Fig. 1b).

Total taxonomic diversity. OTU clustering of the combined eDNA and community DNA datasets at 97% similarity cut-off (after removal of low abundance OTUs) yielded: 442 (eDNA) and 309 (community DNA) OTUs for COIF, and 482 (eDNA) and 394 (community DNA) OTUs for COIS. Taxonomic assignment through BLAST identified the majority of OTUs from Animalia and Protista (Supplementary Fig. 3). From the eDNA samples, COIF identified 170 (35.3%) Animalia OTUs, of which 91 comprised Arthropoda (including 42 Insecta), whilst COIS identified 251 Animalia OTUs (56.8%), of which 212 were Arthropoda (including 167

Insecta) (Supplementary Fig. 4). For the community DNA samples, COIF detected 219 (43.6%) Animalia OTUs, of which 171 were Arthropoda (including 132 Insecta), whilst COIS recovered 227 (73.5%) Animalia OTUs, of which 212 consisted of Arthropoda (including 184 Insecta).

Although not the focus of the study, metabarcoding of the eDNA samples (COIS used here as an example) also yielded matches to fish (*Phoxinus phoxinus*), amphibian and terrestrial OTUs represented at high read frequencies or distributed across numerous independent samples. Of the terrestrial taxa, spider OTUs from the Segestriidae (3,753 reads) and Thomisidae (1,858 reads) families, a millipede OTU (7,312 reads), orthopteran OTU (14,237 reads) and 2,114 reads from domesticated cow (*Bos taurus*) were recovered from multiple samples throughout the year, in addition to a broader diversity of terrestrial groups represented at lower frequencies in the dataset.

Temporal trends of OTU richness from eDNA samples. Measures of OTU richness were calculated exclusively for eDNA samples and plotted against time to detect possible seasonal variations (Supplementary Fig. 5). All samples were rarefied at an equal depth appropriate for each amplicon (total diversity dataset: 57,869 reads per sample, animal diversity dataset: 24,914 reads per sample, for all water samples).

Mean Animalia richness for COIS (±SD) was 37.8 (±10.4), and for COIF, 31.4 (±11.4) (Supplementary Fig. 5a). A significant correlation was detected (Spearman's correlation, $p<0.05$) between the OTU Animalia richness estimates derived from COIF with time and temperature, but not with pH or dissolved oxygen (D.O.). Additionally, mean total richness for COIS (±SD) was 73.1 (±21.2), and for COIF, 88.1 (±26.9) (Supplementary Fig. 5b). A significant correlation was detected (Spearman's correlation, $p<0.05$) between the COIF

(total richness), time, temperature and D.O., but not pH. No significant correlation was found for COIS for the Animalia and total richness and any of the above parameters.

Community structure (β-diversity) from eDNA samples. We used eDNA samples to look into possible changes in community structure over time, for the Animalia identified diversity as well as the total diversity in the dataset. For the eDNA samples, nMDS analysis (Sørensen index) of total diversity for both amplicons (Fig. 2), delimited patterns of seasonal variations driving community composition with qualitatively higher temporal resolution recovered from the smaller amplicon COIS. ANOSIM analyses also supported two main groupings, 'winter' (Nov-April) and 'summer' samples (April–Oct) (COIF: ANOSIM sig. level=0.1%, Global R = 0.717, COIS: ANOSIM sig. level = 0.2%, Global R = 0.475, with outlying samples from winter sampling). Additional analysis of the total diversity supports similar findings [two main groupings: 'winter' (Nov-April) and 'summer' samples (April–Oct) (COIF: ANOSIM sig. level=0.1%, Global R = 0.777, COIS: ANOSIM sig. level = 0.1%, Global R = 0.703)] (Supplementary Fig. 6).

Temporal trends in Chironomidae richness. Analyses of un-trimmed COIF Chironomidae data suggested that temporal richness patterns between eDNA and community DNA samples were comparable (Spearman's correlation $p<0.01$ between eDNA and community DNA for COIF un-trimmed data) (Supplementary Fig. 7). Nevertheless, the sequencing coverage of Chironomidae from the eDNA samples were approximately an order of magnitude lower for COIF than for COIS (Supplementary Fig. 2). Subsequently, in order to maintain a sufficient sequencing depth across samples, COIF was not retained for further Chironomidae related analyses and rarefied incidence based data were used with 4,000 sequencing reads per sample, for COIS only (Supplementary Fig. 2).

220 For the Chironomidae assigned OTUs, COIS identified 103 OTUs from eDNA and 94 OTUs
221 from community DNA samples (138 unique OTUs in total). Using a combination of BLAST ID
222 ≥99% and the online Barcode of Life Database (BOLD) species assignment tool[36] , 73 OTUs
223 (53% out of 138 unique) were assigned species level taxonomic information. Analysis of
224 historical species occurrence data collected by the Environment Agency (EA) (summer
225 surveys 2003 – 2013) in Llyn Padarn (N. Wales, UK) indicated the presence of ≥ 99
226 Chironomidae species from 57 genera. Moreover, Fig.1 illustrates the qualitative overlap
227 between the number of chironomid genera delimited by the current community DNA (65%),
228 eDNA (61%) and taxonomy approaches. Similarly, see Supplementary Fig. 8 for overlap
229 between each method for the four summer time points used for analysis.

230 To visualise the empirically derived annual diversity patterns, OTU and genus richness was
231 assessed against time (Fig. 3) using a polynomial model. Observed OTU richness ranged
232 from 5-27 OTUs for eDNA and 1-27 OTUs for community DNA over time (Fig. 3a).
233 Conversely, genus level richness ranged from 5-19 for eDNA, 1-16 for community DNA. For
234 the data derived from taxonomic identification of invertebrate (exuviae) community
235 samples, genus level richness ranged from 10-18 (green points, restricted to 4 summer
236 sampling times) (Fig. 3b). Please also note that sampling points spanning the winter months
237 (days 36 -190), which did not yield data, represented samples which contained very low
238 physical numbers of exuviae. Consequently, they were not sequenced to an adequate depth
239 in a mixed Illumina sequencing library, and could not be retained for analysis.

240 Significant associations were detected between time and Chironomidae OTU and genera
241 richness derived from community DNA (OTU richness: polynomial regression, R^2 =0.890, p-
242 value = 0.008; Genera richness: R^2 = 0.849, p-value = 0.017). However Chironomidae OTU

and genera richness derived from eDNA samples did not differ significantly over time (OTU: polynomial regression R^2 = 0.187, p-value = 0.460; Genera: R^2 = 0.128, p-value = 0.635) (Fig. 3). Taxonomic richness (genus level) also did not differ significantly over the limited time points available from seasonal sampling.

Temporal variation of OTU Abundance. We assessed the annual variation in OTU abundance from metabarcoding sequencing reads between eDNA and community DNA sampling methods using a generalized additive model (GAM). To allow across method comparisons we compared OTU abundances for Chironomidae OTUs occurring in both eDNA and community DNA datasets (45 OTUs). Abundances differed significantly among different OTUs (GAM, F = 4.688, p-value <0.001) with a significant effect of the temporal smoothing term (GAM, F = 2.561, p = 0.047) (Table 1). Additionally, abundances did not differ significantly between methods (GAM, F = 0.013, p-value = 0.908), but a significant OTU identity x method interaction (GAM, F = 1.733, p-value = 0.003) was found. The abundance of OTU reads was also found to be significantly positively correlated with expected species frequency (ranging from 0.01 to 0.79) across 97 sites in the United Kingdom (UK) (two-way ANOVA, R^2 = 0.087, p-value = 0.003) (Table 1), using previously catalogued Chironomidae species frequency data [37] (Fig. 4).

Discussion

We present here one of the first temporal studies of aqueous eDNA and community DNA biodiversity from a lake ecosystem, in addition to targeting a specific group of ecological sentinel macroinvertebrates. In contrast to previous analyses that have used PCR (qPCR) to infer presence/absence of a small number of target species (e.g. macroinvertebrates) from eDNA samples [12,13], we employed HTS of amplicon libraries (metabarcoding) to assess

266 temporal trends in total biodiversity. Such methodology allows for the characterisation of
267 the entire community, which is not possible through targeted individual-species sequencing
268 that employs taxon-specific primers. Simultaneously, we provide among the first accounts of
269 temporally collected biodiversity data from an annual series of eDNA samples compared
270 simultaneously with a series of invertebrate community DNA samples. Our findings yield an
271 informative characterisation of temperate lake ecosystem-wide biodiversity, through
272 detection of multiple groups of organisms from invertebrates to macro-organisms, of
273 primarily freshwater, but also terrestrial origins. Furthermore, the biodiversity of the
274 indicator taxon group used (Chironomidae) was successfully detected throughout the year,
275 from both eDNA and community DNA samples, exhibiting substantial overlap with
276 traditional taxonomic data. In addition, OTU sequence abundances were significantly
277 positively associated with expected chironomid species abundance based on UK taxa
278 occurrence data (Table 1, Fig. 4). Such direct coincidence, despite potential biotic and
279 abiotic variability in the release, transport and persistence of eDNA [8], demonstrates the
280 value of eDNA metabarcoding for biodiversity characterisation and ecosystem monitoring [38].

281 Both metabarcoding amplicons detected large amounts of Animal phylum level diversity
282 from eDNA samples, showing broad representation across the freshwater taxonomic
283 biosphere, including the Arthropoda (Supplementary Fig. 4). Within the Arthropoda, the
284 dominance of Insecta, Maxillopoda and Malacostraca (Crustacea) also demonstrates the
285 utility of eDNA metabarcoding for characterisation of freshwater ecosystem-wide
286 biodiversity. There is increasing exposure of the use of eDNA metabarcoding for the
287 detection of fish and amphibians [19,20], as also recorded here. A more novel concept is the
288 ability of freshwater systems to integrate eDNA biodiversity information from terrestrial
289 sources. Terrestrial species found in our dataset, such as spider, millipede and orthopteran

290 species, or the ubiquitous *Bos taurus* (please also note, that no bovine serum products were
291 used in the HTS library preparations), are all commonplace in the surrounding area of the
292 study site and were detected by the analysis of eDNA residing in the lake water samples.
293 The ability of freshwater catchments to contain eDNA from broader habitat biodiversity
294 therefore presents an opportunity for further research regarding the relationship between
295 aqueous eDNA and biodiversity at the landscape scale.

296 Focusing on the Chironomidae richness estimates derived from the analysis of the short COI
297 fragment (Fig. 3), we can see that the COIS amplicon yielded 138 unique OTUs from both
298 sample types throughout the year. The analysis of the COIS amplicon therefore provided
299 valuable comparative qualitative and quantitative data both within the metabarcoding
300 datasets and between the historically collected data for Llyn Padarn and the rest of the UK
301 [37]. Other eDNA studies have focused mainly on macro-organisms such as fish or amphibians
302 whereby skin cells and mucus are a likely primary source of eDNA [8]. While aquatic
303 invertebrates such as chironomids are individually typically much smaller, the accumulated
304 biomass of the community clearly produces sufficiently detectable and persistent amounts
305 of eDNA (from natural shedding, moulting and death) for meaningful biodiversity
306 assessment. Additional quantitative studies are required to determine the effects of
307 invertebrate community biomass on levels of eDNA in environmental samples [21].

308 Sequencing of the complete COI region (COIF ~658bp) from eDNA samples was successful in
309 detecting several genera of chironomids and provided biodiversity estimates comparable
310 with community DNA biodiversity patterns (Supplementary Fig. 7). However, it was not
311 possible to retain the COIF locus throughout all analyses after applying strict abundance
312 filtering of OTUs. Low sequence coverage of the COIF for the Chironomidae (primarily in the

313 water eDNA and not the community DNA samples, Supplementary Fig. 2) meant that more
314 robust, ecological comparisons were more effectively achieved using the short eDNA
315 fragment (COIS). Possible reasons for the discrepancies in coverage of the two amplicons
316 could be related to variations in primer specificity, with the COIS primers being more
317 successful than COIF primers in amplifying Chironomidae [39] (please also see the limitations
318 of the Folmer COI barcoding primers for metabarcoding analyses in [40]). Nevertheless, we did
319 not detect substantial phylogenetic biases in OTUs recovered from the two primer pairs
320 (Supplementary Fig. 9) and coverage of the Chironomidae was only depleted in the water
321 eDNA samples for the COIF. Alternatively, the discrepancy in different amplicon success may
322 be due to the reduced availability of longer sized eDNA fragments in a natural ecosystem [30].

323 After DNA is released into the environment, the degradation process likely begins, breaking
324 down DNA and yielding shorter fragments. It has been shown that ~400bp length fragments
325 remain detectable in water for days to weeks [6,11], with the rate of degradation depending
326 upon various biotic and abiotic factors [27]. Overall, smaller fragments degrade slower
327 compared to longer fragments, suggesting an enhanced probability of detection by studies
328 targeting shorter DNA fragments [41]. The present data support the enhanced detection of
329 shorter eDNA fragments, as evidenced by higher sequence coverage of the Chironomidae by
330 the shorter COIS amplicon in the water eDNA samples. Nevertheless, the data additionally
331 show that longer fragments are available at likely lower concentrations in the wild [30]
332 (represented by the COIF amplicon) (Supplementary Fig. 2). Using time vs. DNA
333 fragmentation as a working hypothesis for eDNA degradation, longer fragments are
334 predicted to represent more recently living cellular material. It is also therefore noteworthy
335 that among the water eDNA analyses, only the biodiversity delimited by the COIF amplicon
336 yielded significant associations with time/temperature (Spearman's correlation, $p<0.05$)

337 (Supplementary Fig. 5), most likely representing more rapid breakdown of longer eDNA
338 fragments in the lake environment. Nevertheless, higher sequence coverage, or methods
339 that preferentially amplify longer amplicons, are needed to enhance amplification
340 probability for potentially smaller concentrations of longer eDNA fragments in natural
341 systems. Such solutions include the combination of multiple primer pairs[17], or use of taxon
342 specific/blocking primers. Other suggested strategies for enhancing HTS of eDNA (where
343 concentrations are sufficiently high) involve direct shotgun sequencing or use of capture
344 probes [28,42].

345 Amongst the concerns regarding the utility of eDNA to assess biodiversity, is whether or not
346 species detection represents living or recently living organisms, or communities of 'zombie'
347 DNA (i.e. historically distant DNA from organisms that previously lived in the ecosystem a
348 substantial time ago)[38]. If eDNA exhibited long persistence times in the wild, temporal
349 patterns of β-diversity would be predicted to be extremely low (i.e., non-existent), especially
350 when derived from smaller fragments. However, here we have clearly shown that temporal
351 turnover (β-diversity) was observed for both the animal level (Fig. 2), and total diversity
352 derived eDNA biodiversity analysis (Supplementary Fig. 6), including temporal patterns of
353 seasonal biodiversity groupings over the year. Similar temporal results were observed for
354 both amplicons, with the short eDNA amplicon providing higher temporal resolution. Some
355 winter samples (Nov 25th and Dec 17th) in the COIS nMDS analysis displayed high levels of
356 β-diversity, since they either contained higher richness (Supplementary Fig. 5, days 57 and
357 79) or additional cohorts of taxa not present in the remaining samples (Supplementary Fig.
358 4). In the absence of technical artefacts, the additional turnover in β-diversity observed
359 could be the consequence of extreme storm events that coincided with the winter 2013-
360 2014 sampling [43], inputting additional allochthonous eDNA from outside the study area. The

361 time points defining the separation of the two main seasonal biodiversity groups were
362 identified over November and late April, times which also correspond to water temperature
363 below 8 ˚C (winter samples) and above 10 ˚C (summer samples). Changes in observed
364 community composition (β-diversity) over April and November (Fig. 2, Supplementary Fig. 6)
365 most likely reflect seasonal turnover, possibly attributed to lake inversion effects [44]. It is
366 known that changes in water temperature around these times of the year (Spring and
367 Autumn), can trigger the loss of water column stratification by mixing due to changes in
368 surface water temperature[44]. Collectively, the demonstration of seasonal turnover of lake
369 eDNA β-diversity supports empirical studies using model ecosystems[44]. Previous laboratory
370 and mesocosm studies have demonstrated the short-term temporal decay of eDNA in
371 artificial environments (e.g. 2-6 weeks)[8,22,26] and the present data show that the eDNA signal
372 in the wild is of a contemporary nature.

373 Metabarcoding sequencing of invertebrate communities directly reveals the
374 presence/absence of living, or recently living communities [28]. Hence, the insights provided
375 by community DNA samples here offered an essential benchmark to serve as a proxy for the
376 contemporary invertebrate community. The biodiversity estimates derived from
377 metabarcoding of the community DNA (Fig. 3, Supplementary Fig. 7, orange lines) matched
378 literature-based estimations of seasonal variation of Chironomidae for Northern
379 Hemisphere temperate latitudes [45], with a decrease in species richness over winter (often
380 represented by 'null' samples due to low numbers of collected exuviae) and a summer
381 increase related to rising water temperature (Fig. 3). Since the emergence patterns of
382 Chironomidae through the year are strongly related to changes in temperature and
383 photoperiod [45] (Supplementary Methods), rapid turnover in emerging communities are
384 apparent and can yield biased estimates of ecological status due to short-term shifts of

385 species emergence [46]. One of the advantages of metabarcoding over traditional analysis is
386 the ability to analyse many samples simultaneously, and so using molecular approaches for
387 biodiversity assessment presents the opportunity to intensify ecological assessment and
388 derive greater precision in ecosystem health assessment [3].

389 The companion analysis of the chironomid eDNA did not follow the expected emergence
390 pattern in richness, despite detecting Chironomidae turnover throughout the year from
391 community DNA samples (Fig. 3). The combination of the β-diversity turnover in eDNA
392 composition (Fig. 2), seasonally fluctuating community DNA richness (Fig. 3, orange lines)
393 and a lack of coherent seasonal shifts in eDNA richness (Fig. 3, blue line) thereby provides an
394 annual model of 'community DNA – eDNA' dynamics. The data thereby suggest that there
395 will likely be a standing resource of eDNA for biodiversity detection in lake ecosystems that
396 experience annual species turnover [44] (Fig. 2). Compositional turnover is thereby expected
397 to result from seasonal variation in species abundances, increasing sources of contemporary
398 eDNA, and environmental degradation decreasing levels of past eDNA accumulation.

399 Using GAM modelling facilitated comparison between read abundances of individual OTUs
400 derived from eDNA and community DNA analyses. Numbers of read abundances differed
401 between OTUs over time and between eDNA and community DNA abundances at the
402 individual OTU level (Table 1). There was also a significant positive association between the
403 abundance of sequencing reads derived from the present study and species frequency at
404 the national scale (Fig. 4). Therefore, lower abundance OTUs from the present study occur
405 at lower frequencies and more abundant OTUs are more common, according to an
406 extensive database of Chironomidae occurrence across the UK [37] (Fig. 4).

In combination, the analyses provide an overview of chironomid lake eDNA dynamics. Some species will inevitably yield higher levels of eDNA than others, in relation to life history stage, moulting rates/frequency, abundance, biomass, or cellular content/mitochondrial densities[3,7,8]. In addition, the relationship between eDNA and community DNA is affected by biophysical characteristics and interactions between biotic and abiotic factors (e.g. microbial activity, UV radiation and temperature) that affect persistence and degradation rates throughout the year[8,27]. Despite such dynamic interactions, numerous broad quantitative associations have been reported for a range of taxa and their eDNA profiles, including data from artificial, semi-natural and natural aquatic ecosystems [47–52]. Here also, regardless of which methodology was employed, metabarcoding of both eDNA and community DNA reflected general Chironomidae species frequencies across the UK [37] (Fig. 4) and overlapped with biodiversity estimates derived from taxonomy analyses (Fig. 1).

In summary, we have shown that eDNA from water samples collected consecutively over an annual cycle in a lake ecosystem reveals ecologically representative species and community-level shifts in diversity. Importantly, such patterns were validated both by independent assessments of changes in physical presence in a key indicator group of macroinvertebrates, as well as coinciding with established seasonal trends in indicator species emergence and traditional taxonomy. Collectively, the findings address key outstanding questions related to the ecological relevance and temporal persistence of freshwater eDNA in a natural ecosystem, with significant implications for biomonitoring and the future investigation of biodiversity ecosystem functioning relationships.

Methods

Field sampling. Samples (chironomid pupal exuviae and water samples) were collected during Sept 2013 – Sept 2014 from Llyn Padarn, UK (Supplementary Methods), an oligotrophic lake ecosystem located in Snowdonia National Park (53.130051, -4.135567), N. Wales, UK (Supplementary Fig. 10) (Approximate surface area is 97.6 ha, maximum depth 27m). The site has been monitored regularly by the UK Environment Agency (EA), and more recently by Natural Resources Wales (NRW) for indicator species of Chironomidae and other invertebrate communities, providing important historical data. Two sites at opposite sides of the lake were selected for sampling: Site 1 (S1: NW: 53.139106, -4.153975) and Site 2 (S2: SW: 53.122414, -4.126761) (Supplementary Fig. 10). Using two locations increases potential for species detection based on both eDNA and invertebrate sampling. Sampling was conducted at approximately three-week intervals for 1 year (16 time points), using standardised sampling methodology, and collecting simultaneously water and Chironomidae samples. The two sites were sampled always in the same sequence (S1, then S2) between 8:30am–11:30am, including consecutive collection of water samples, invertebrate samples, followed by water metadata (pH, Dissolved Oxygen (D.O.), conductivity and water temperature), using a calibrated YSI Pro Plus multi-meter. As only water and exuviae (shed skins) were collected and the work was performed in collaboration with the EA and NRW, a permit was not required.

Chironomid exuviae collection and eDNA filtration. Invertebrate samples in the form of chironomid exuviae (shed pupal skins) were collected using the field collection protocols for the Chironomid Pupal Exuviae Technique (CPET) [53], with a 250µm mesh collection net (Supplementary Methods). The floating insect skins were collected on the leeward side (accumulation area) of each sampling site following described methods [34] and placed in a sterile container. Upon returning to the lab, the sample was coarsely sorted to remove

excessive plant debris, fixed in 100% ethanol and stored at 4°C on the same day of collection, until further processing.

For eDNA samples, one litre of surface water was collected using sterile glass Nalgene bottles from each site, which was transferred on ice and placed at 4°C immediately after return to the laboratory. Filtration was completed within 6 hours in a PCR-free separate room. Sterilised, reusable funnel filtration units (Nalgene filter holders with funnel) were used with 0.45µm cellulose nitrate filter membranes and a high-pressure vacuum pump. The filter membranes were stored in sterile 15ml falcon tubes at - 80°C until DNA extraction.

Equipment Sterilization and negative control samples. All equipment was thoroughly sterilized between sampling visits. The glass Nalgene bottles used for water collection, filtration units and forceps would undergo consecutive cleaning rounds including wash and overnight soak with 10% Trigene (Ammonium chloride & hydrochloride, Medichem Int.), thorough rinse, UV treatment for 5 min and autoclaving. All additional equipment used for invertebrate collection (net, meters, boots) was also thoroughly washed with 10% Trigene. For eDNA extractions, single-use pre-sterilised scissors and forceps were used to handle the filter membranes, and the exterior of storage tubes was wiped with 10% Trigene before handling. During field surveys, to minimise cross contamination from consecutive sampling points, the water samples were collected first, before any other samples or measurements were taken and prior to invertebrate collection. **Negative controls** were collected by filtration of 1 litre of distilled water through the filtration funnels and filter membranes processed. Blank extractions of reagents (reagent controls) and filters (filter controls) were also extracted with the same Phenol Chloroform extraction protocol (PCI) [54]. The negative control equipment would undergo the same cleaning steps as stated above. All negative

476 controls were amplified with both primer pairs and MiSeq library preparation steps (as
477 below), and sequenced on an Illumina MiSeq.

478 **DNA extractions for eDNA filter membranes and invertebrate samples.** Environmental DNA
479 (eDNA) was extracted from the filter membranes, using a modified Phenol Chloroform
480 protocol (PCI), adapted from Renshaw et al [54], with an added digestion step with the
481 addition of 20µl Proteinase K (20mg/µl) (Sigma – Aldrich) and incubation at 60°C for 1 hour.
482 This protocol was selected after rigorous in-house testing of available eDNA capture and
483 extraction protocols (Supplementary Methods). In Renshaw *et al.* [54] it was demonstrated
484 that the latter protocol yielded the highest number of DNA copies of targeted eDNA
485 fragments. Furthermore, the combination of filtration and PCI has been shown to optimise
486 DNA yields, performing equally well in eukaryotes and prokaryotes, with enhanced
487 detection of diversity than other methods [55]. Two individual extractions were performed for
488 each sample, which were subsequently pooled. Extractions were performed in a different
489 building to PCR library construction where no invertebrate DNA had been handled
490 previously. Extracts were stored in a clean room with no post PCR processing.

491 DNA extraction from the bulk pupal exuviae samples (community DNA) was performed
492 using a modified QIAmp Blood Maxi Kit protocol, with an added Proteinase K overnight
493 incubation step. Due to seasonal variation of chironomid emergence [45], the mass of the
494 collected invertebrate skin material varied, with some of the winter samples containing
495 smaller amounts of tissue. In order to optimise extraction efficiency, 1g of dry invertebrate
496 material was subsampled from large samples. Conversely, for some low-density winter
497 samples, 1g of exuviae was not available and so in these instances, the whole sample was
498 used for analysis. DNA extraction was performed in standard Qiagen Blood and Tissue kit

499 columns for small winter samples and QIAmp Blood Maxi Kit columns for all other samples
500 with an added 20µl Proteinase K (20mg/µl) overnight incubation step. Both kits are verified
501 by Qiagen to use the same chemistry and differ with respect to the use of columns of
502 different volume capacity to prevent clogging of the membrane. Following separation from
503 the ethanol preservative, the community samples were allowed to air-dry for approximately
504 1 hour and then were homogenised using a sterile mechanical drill and pestle. For detailed
505 information on each extracted sample, see Supplementary Tables 4 and 5.

506 **Primer selection and MiSeq Library preparation.** To fulfil the overarching aims of the study,
507 we required (a) metabarcoding primers that would amplify across a broad range of taxa (in
508 particular, lake occurring taxa), (b) a marker enabling the best annotation power for
509 macroinvertebrates and in particular, the Chironomidae, (c) a combination of two primer
510 pairs providing different length amplicons.

511 Accordingly, two amplicons of different sizes of the mitochondrial Cytochrome Oxidase I
512 gene (COI) were selected for sequencing. The full-length COI barcoding region (658bp),
513 using the universal Folmer primers LCO1490 - HCO2198 [56] (amplicon COIF) and a 235bp
514 fragment (amplicon COIS) using the forward primer LCO1490 and the reverse COIA-R primer
515 (reversed forward COI-A primer by [39]) (See Supplementary Table 6 for primer sequences).
516 Initially, the forward COI-A primer was designed by [39] specifically for amplification of
517 Chironomidae from environmental samples. Two Illumina MiSeq dual indexed amplicon
518 libraries were prepared using a two-step PCR protocol [57]. The first round amplification was
519 performed using template-specific primers with 5' Illumina tails [TruGrade, by IDT,
520 Integrated DNA Technologies (Coralville, USA)], followed by Agencourt AMPure magnetic
521 bead purification. A second round amplification was performed using Illumina adapters with

8-nucleotide Nextera indexes (Supplementary Table 6). A 5N sequence was implemented between the forward universal tail and the template specific primer, which is known to improve clustering and cluster detection on MiSeq sequencing platforms [57]. Using primers with identical tails in the first step and indexed primers in the second, is a protocol specifically developed by Illumina to reduce bias caused by variable index sequences in mixed environmental samples [58,59].

Each sample was amplified in triplicate, the final products were pooled and purified with AMPure beads and quantified using a dsQubit assay. Final library pooling was performed in equimolar quantities for all samples. Sequencing was performed at the Liverpool Centre for Genome Research, distributed across two independent lanes (for the COIS and COIF amplicons) of Paired-end Illumina MiSeq (2x300) sequencing.

PCR protocols for MiSeq Library Preparation. PCRs were performed in 25µl reaction volumes containing, for Round 1: 12.5µl Q5® Hot Start High-Fidelity 2X Master Mix, 10.5µl PCR water, 0.5µl (10nmole/µl) of each forward and reverse primer and 1µl DNA (10ng/µl). For Round 2: 12.5µl Q5® Hot Start High-Fidelity 2X Master Mix, 6.5µl PCR water, 0.5µl of each forward and reverse primer and 5µl Purified PCR product from Round 1. The following thermo-cycling parameters were used: Round 1: COIF: Denaturation at 98°C for 30 sec, 20 cycles of: 98°C for 10 sec, 46°C for 30 sec, 72°C for 40 sec, followed by a 10min extension at 72°C, hold at 4°C. COIS: Denaturation at 98°C for 30 sec, 20 cycles of: 98°C for 10 sec, 45°C for 30 sec, 72°C 30 sec, followed by a 10min extension at 72°C, hold at 4°C. Round 2: both amplicons: Denaturation at 98°C for 30 sec, 15 cycles of: 98°C for 10 sec, 55°C for 30 sec, 72°C for 30 sec, followed by a 10min extension at 72°C, cool at 4°C for 10min. Round 1 PCRs were performed using Illumina-tailed primers and Round 2 using Illumina indexes.

545 **Positive control samples.** To account for efficiency of amplification protocols and
546 sequencing, a composite positive control sample comprising 30 invertebrate DNA extracts,
547 including Amphipoda, Coleoptera, Diptera, Ephemeroptera, Gastropoda, Hemiptera,
548 Isopoda and Trichoptera, was also amplified in triplicate with both primer pairs, and
549 sequenced alongside eDNA and community samples on MiSeq (Supplementary Table 2).

550 **Bioinformatics and statistical analysis**. Sequences, including positive and negative controls,
551 were de-multiplexed and Illumina adapters trimmed using Cutadapt [60] and Sickle [61]. A 10%
552 level of mismatch (2 bases) was allowed for primer removal. Filtering and quality control
553 were then performed using USEARCH v7 [62]. Sequence quality was visualised using FastQC
554 (www.bioinformatics.babraham.ac.uk) and only sequences with a Phred quality score >25
555 were retained for analysis. Using USEARCH (fastq_maxee = 1) sequences with a maximum
556 expected error (maxee) > 1 were discarded. Maxee is the expected number of errors as sum
557 of the error probabilities (provided by Phred scores). Filtering was performed after merging
558 of R1 and R2 reads (minimum overlap 25bp), which allows recalculation of the error
559 probabilities for the combined sequences and increased accuracy. Sequences shorter than
560 100bp were discarded. The remaining sequences were de-replicated and sorted by cluster
561 size (cluster abundance) and sequences with <2 clusters (singletons) were removed. For the
562 COIF amplicon, the whole barcoding region was amplified and sequenced, but because of
563 the current limitations of MiSeq sequencing read lengths, only the forward reads (R1) were
564 used for analysis. Consequently, the per base quality drop expected in Illumina MiSeq data
565 at the tail of the forward reads was inspected in FastQC and all reads were truncated at
566 250bp and then quality filtered as above. Next, chimeras were removed (uchime_denovo)
567 using a *de novo* delimitation approach. An operational taxonomic unit (OTU) table was
568 created using OTU clustering at 97% similarity (USEARCH). Clustering at 97% similarity level

569 was chosen based on existing knowledge of intraspecific diversity for Chironomidae [39], since
570 previous studies suggest that chironomid intraspecific diversity ranges between 0-4.2% [39] or
571 0-4.9% [63].

572 Taxonomy was then assigned to the OTU table using BLAST+ (megablast) [64] against a
573 reference COI database. The reference library was compiled from NCBI GenBank, by
574 downloading all COI sequences, >100bp, excluding environmental sequences (20th June
575 2015, N = 807,388 sequences) and higher taxonomic level information was edited using the
576 GALAXY online software platform [65]. Taxonomic assignment of the OTU tables and
577 subsequent analysis was performed in QIIME [66]. All analyses involving USEARCH, QIIME and
578 BLAST+ were performed using the High Performance Computing (HPC) Wales systems.

579 Given the potentially sensitive nature of eDNA metabarcoding, low frequency sequences
580 can either represent less abundant taxa, or possible false positives and low level
581 contaminant OTUs [67]. In order to reduce the error associated with low frequency sequences,
582 and also focus analyses on predicted levels of richness [68], we used two types of analysis.
583 First, we identified the frequency of potential contaminant reads in the positive control.
584 Second, we compared chironomid eDNA richness with variable levels of relative abundance
585 filtering (no filtering, 0.01% and 0.02%), against historical records of richness (genus level
586 only available) for Llyn Padarn (based on summer surveys for Llyn Padarn, 2003 – 2013).
587 Consequently, abundance filtering was performed on the OTU tables at the level that most
588 closely emulated expected chironomid richness and within the limits associated with
589 empirically observed low-level contamination in the sequencing dataset.

590 The validity of the Chironomidae OTUs identified by BLAST and retained after abundance
591 filtering was checked using a phylogenetic approach. The BLAST identified Chironomidae

OTUs were aligned with barcodes from 24 Chironomidae and 40 Trichoptera species obtained herein, sequenced from UK samples using universal primers [56]. Alignment, testing for the presence of stop codon and insertions and bootstrapped phylogenetic tree construction were performed in MEGA [69]. Ultimately, only the OTUs that grouped closely with known chironomid sequences on the phylogenetic tree were included in further analysis.

For downstream analyses, the appropriate depth of coverage per sample was determined according to OTU accumulation vs. sequence coverage curves generated in QIIME. Samples were subsequently normalised using rarefaction in QIIME at appropriate depth for each amplicon [70].

Taxonomic identification of invertebrate community samples. To provide a comparison with community DNA and eDNA sequenced samples, chironomid exuviae community samples from 4 time points (T10: April 30, T11: May 20, T14: July 23, T16: September 04) were taxonomically identified according to standard CPET methodology used by the EA. More specifically, 200 chironomid exuviae were subsampled from the total community sample and identified to the highest possible level (genus or species) by specialised EA staff. The results of the taxonomic identification were used to compare chironomid richness at the genus level with metabarcoding-generated richness (see below).

Calculation of diversity measures. OTU richness (total diversity and Chironomidae diversity) was calculated in QIIME. Furthermore, for Chironomidae with good taxonomic identification, richness was also calculated at the genus level. To assess variation of richness over time polynomial regression was performed using R version 3.2.4 (2016).

The PRIMER-E software [71] was used to calculate β-diversity based on the Sørensen index for total diversity and Animalia only diversity detected from aqueous eDNA samples and for Chironomidae OTUs for both sample types. Non-metric multi-dimensional scaling (nMDS) and Hierarchical Clustering (HC) analysis were used to represent community similarity between samples. Analysis of similarity (ANOSIM) was used to test for significant effects of time in relation to community composition.

Chironomidae OTU read abundance (eDNA vs community DNA). In order to explore relationships between the numbers of metabarcoding sequence reads, individual OTUs and methodology (eDNA vs. community DNA), we used a generalized additive model (GAM), with time as a smoothing term, using the R-package mgcv [72]. In the GAM model, abundance, calculated as total normalised reads per OTU and standardized per method (to allow for across method comparison), was assessed in relation to OTU identity and method (eDNA vs community DNA). Additionally, we assessed the ecological relationship between OTU abundance (log transformed) in Llyn Padarn and species frequency (i.e. abundances derived from ecological assessment) across the UK, by performing a two-way ANOVA, using the lm function in R. UK species frequencies were derived from a Chironomidae inventory of 435 species across 220 UK lakes [37]. We restricted the species frequency data to 97 sites where species frequency was inventoried at the national level and observed in this study.

Data availability. Sequencing data reported here have been deposited in GenBank (Submission IDs: 1966226, 1966195) and the European Nucleotide Archive (ENA) (Accession number: PRJEB13009).

637 **References**

638

639 1. Loreau, M. & de Mazancourt, C. Biodiversity and ecosystem stability: A synthesis of
640 underlying mechanisms. *Ecol. Lett.* **16,** 106–115 (2013).

641 2. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **489,** 326–
642 326 (2012).

643 3. Thomsen, P. F. & Willerslev, E. Environmental DNA - An emerging tool in conservation
644 for monitoring past and present biodiversity. *Biol. Conserv.* **183,** 4–18 (2015).

645 4. Lawson Handley, L. How will the 'molecular revolution' contribute to biological
646 recording? *Biol. J. Linn. Soc.* **115,** 750–766 (2015).

647 5. Yu, D. W. *et al.* Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity
648 assessment and biomonitoring. *Methods Ecol. Evol.* **3,** 613–623 (2012).

649 6. Dejean, T. *et al.* Persistence of environmental DNA in freshwater ecosystems. *PLoS*
650 *One* **6,** e23398 (2011).

651 7. Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M. & Gough, K. C. The
652 detection of aquatic animal species using environmental DNA - a review of eDNA as a
653 survey tool in ecology. *J. Appl. Ecol.* **51,** 1450–1459 (2014).

654 8. Barnes, M. A. & Turner, C. R. The ecology of environmental DNA and implications for
655 conservation genetics. *Conserv. Genet.* **17,** 1–17 (2016).

656 9. Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity
657 monitoring. *Trends Ecol. Evol.* **29,** 358–367 (2014).

658 10. Kelly, R. P. *et al.* Harnessing DNA to improve environmental management. *Science.*
659 **344,** 1455–1456 (2014).

660 11. Goldberg, C. S., Sepulveda, A., Ray, A., Baumgardt, J. & Waits, L. P. Environmental
661 DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus*
662 *antipodarum*). *Freshw. Sci.* **32,** 792–800 (2013).

663 12. Deiner, K. & Altermatt, F. Transport distance of invertebrate environmental DNA in a
664 natural river. *PLoS One* **9,** e88786 (2014).

665 13. Mächler, E., Deiner, K., Steinmann, P. & Altermatt, F. Utility of environmental DNA for
666 monitoring rare and indicator macroinvertebrate species. *Freshw. Sci.* **33,** 1174–1183
667 (2014).

668 14. Chave, J. The problem of pattern and scale in ecology: What have we learned in
669 20 years? *Ecol. Lett.* **16,** 4–16 (2013).

670 15. Ji, Y. *et al.* Reliable, verifiable and efficient monitoring of biodiversity via
671 metabarcoding. *Ecol. Lett.* **16,** 1245–1257 (2013).

672 16. Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C. & Baird, D. J. Environmental
673 barcoding: A next-generation sequencing approach for biomonitoring applications
674 using river benthos. *PLoS One* **6,** e17497 (2011).

675 17. Gibson, J. *et al.* Simultaneous assessment of the macrobiome and microbiome in a
676 bulk sample of tropical arthropods through DNA metasystematics. *Proc. Natl. Acad.*

677 *Sci. U. S. A.* **111,** 8007–12 (2014).
678 18. Gibson, J. F. *et al.* Large-scale biomonitoring of remote and threatened ecosystems
679 via high-throughput sequencing. *PLoS One* **10,** 1–15 (2015).
680 19. Hänfling, B. *et al.* Environmental DNA metabarcoding of lake fish communities
681 reflects long-term data from established survey methods. *Mol. Ecol.* **25,** 3101–3119
682 (2016).
683 20. Valentini, A. *et al.* Next-generation monitoring of aquatic biodiversity using
684 environmental DNA metabarcoding. *Mol. Ecol.* **25,** 929–942 (2016).
685 21. Evans, N. T. *et al.* Quantification of mesocosm fish and amphibian species diversity via
686 environmental DNA metabarcoding. *Mol. Ecol. Resour.* **16,** 29–41 (2016).
687 22. Thomsen, P. F. *et al.* Monitoring endangered freshwater biodiversity using
688 environmental DNA. *Mol. Ecol.* **21,** 2565–2573 (2012).
689 23. Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J.C. & Altermatt, F. Environmental
690 DNA reveals that rivers are conveyer belts of biodiversity information. *Nat. Commun.*
691 **7,** 12544 (2016).
692 24. Lodge, D. M. *et al.* Conservation in a cup of water: Estimating biodiversity and
693 population abundance from environmental DNA. *Mol. Ecol.* **21,** 2555–2558 (2012).
694 25. Creer, S. *et al.* The ecologist's field guide to sequence-based identification of
695 biodiversity. *Methods Ecol. Evol.* **56,** 68–74 (2016).
696 26. Strickler, K. M., Fremier, A. K. & Goldberg, C. S. Quantifying effects of UV-B,
697 temperature, and pH on eDNA degradation in aquatic microcosms. *Biol. Conserv.* **183,**
698 85–92 (2015).
699 27. Barnes, M. A. *et al.* Environmental conditions influence eDNA persistence in aquatic
700 systems. *Environ. Sci. Technol.* **48,** 1819–1827 (2014).
701 28. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-
702 generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21,** 2045–
703 2050 (2012).
704 29. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715
705 (1993).
706 30. Deagle, B. E., Eveson, J. P. & Jarman, S. N. Quantification of damage in DNA recovered
707 from highly degraded samples--a case study on DNA in faeces. *Front. Zool.* **3,** 11
708 (2006).
709 31. Magurran, A. E. *et al.* Long-term datasets in biodiversity research and monitoring:
710 Assessing change in ecological communities through time. *Trends in Ecology and*
711 *Evolution* **25,** 574–582 (2010).
712 32. Dejean, T. *et al.* Improved detection of an alien invasive species through
713 environmental DNA barcoding: The example of the American bullfrog Lithobates
714 catesbeianus. *J. Appl. Ecol.* **49,** 953–959 (2012).
715 33. Biggs, J. *et al.* Using eDNA to develop a national citizen science-based monitoring
716 programme for the great crested newt (*Triturus cristatus*). *Biol. Conserv.* **183,** 19–28
717 (2015).

718 34. Wilson, R. & Ruse, L. *A guide to the identification of genera of chironomid pupal*
719 *exuviae occurring in Britain and Ireland*. (Freshwater Biological Association
720 Publication 13, Ambleside, UK., 2005).
721 35. Ruse, L. Lake acidification assessed using chironomid pupal exuviae. *Fundam. Appl.*
722 *Limnol. / Arch. für Hydrobiol.* **178,** 267–286 (2011).
723 36. Ratnasingham, S. & Hebert, P. D. N. BOLD: The Barcode of Life Data System:
724 Barcoding. *Mol. Ecol. Notes* **7,** 355–364 (2007).
725 37. Ruse, L. P. Chironomid (Diptera) species recorded from UK lakes as pupal exuviae. *J.*
726 *Entomol. Acarol. Res.* **45,** 13 (2013).
727 38. Baird, D. J. & Hajibabaei, M. Biomonitoring 2.0: A new paradigm in ecosystem
728 assessment made possible by next-generation DNA sequencing. *Mol. Ecol.* **21,** 2039–
729 2044 (2012).
730 39. Carew, M., Pettigrove, V., Metzeling, L. & Hoffmann, A. Environmental monitoring
731 using next generation sequencing: rapid identification of macroinvertebrate
732 bioindicator species. *Front. Zool.* **10,** 45 (2013).
733 40. Deagle, B. E. *et al.* DNA metabarcoding and the cytochrome c oxidase subunit I
734 marker: not a perfect match. *Biol. Lett.* **10,** 1789–1793 (2014).
735 41. Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. Environmental DNA. *Mol.*
736 *Ecol.* **21,** 1789–1793 (2012).
737 42. Liu, S. *et al.* Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free
738 mitogenomics biodiversity analysis. *Mol. Ecol. Resour.* **16,** 470–479 (2016).
739 43. Met Office. Hadley Centre for Climate Prediction and Research (JCHMR), Maclean
740 Building, Wallingford OX10 8BB, UK. (2016).
741 44. Moss, B. R. *Ecology of Fresh Waters: A View for the Twenty-First Century*. (Wiley-
742 Blackwell, 2010).
743 45. Armitage, P. D., Pinder, L. C. & Cranston, P. *The Chironomidae: biology and ecology of*
744 *non-biting midges*. (Chapman and Hall, 1995).
745 46. Raunio, J., Paasivirta, L. & Hämäläinen, H. Assessing lake trophic status using spring-
746 emerging chironomid pupal exuviae. *Fundam. Appl. Limnol. / Arch. für Hydrobiol.* **176,**
747 61–73 (2010).
748 47. Kelly, R. P., Port, J. A., Yamahara, K. M. & Crowder, L. B. Using environmental DNA to
749 census marine fishes in a large mesocosm. *PLoS One* **9,** e86175 (2014).
750 48. Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. N. & Kawabata, Z. Surveillance
751 of fish species composition using environmental DNA. *Limnology* **13,** 193–197 (2012).
752 49. Pilliod, D. S., Goldberg, C. S., Arkle, R. S., Waits, L. P. & Richardson, J. Estimating
753 occupancy and abundance of stream amphibians using environmental DNA from
754 filtered water samples. *Can. J. Fish. Aquat. Sci.* **70,** 1123–1130 (2013).
755 50. Thomsen, P. F. *et al.* Detection of a Diverse Marine Fish Fauna Using Environmental
756 DNA from Seawater Samples. *PLoS One* **7,** e41732 (2012).
757 51. Klymus, K. E., Richter, C. A., Chapman, D. C. & Paukert, C. Quantification of eDNA
758 shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp

759 Hypophthalmichthys molitrix. *Biol. Conserv.* **183,** 77–84 (2015).

760 52. Lacoursière-Roussel, A., Rosabal, M. & Bernatchez, L. Estimating fish abundance and
761 biomass from eDNA concentrations: variability among capture methods and
762 environmental conditions. *Mol. Ecol. Resour.* **16,** 1401–1414 (2016).

763 53. Ruse, L. Classification of nutrient impact on lakes using the chironomid pupal exuvial
764 technique. *Ecol. Indic.* **10,** 594–601 (2010).

765 54. Renshaw, M. A., Olds, B. P., Jerde, C. L., Mcveigh, M. M. & Lodge, D. M. The room
766 temperature preservation of filtered environmental DNA samples and assimilation
767 into a phenol-chloroform-isoamyl alcohol DNA extraction. *Mol. Ecol. Resour.* **15,** 168–
768 176 (2015).

769 55. Deiner, K., Walser, J. C., Mächler, E. & Altermatt, F. Choice of capture and extraction
770 methods affect detection of freshwater biodiversity from environmental DNA. *Biol.*
771 *Conserv.* **183,** 53–63 (2015).

772 56. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for
773 amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan
774 invertebrates. *Mol. Mar. Biol. Biotechnol.* **3,** 294–299 (1994).

775 57. Miya, M. *et al.* MiFish, a set of universal PCR primers for metabarcoding
776 environmental DNA from fishes: detection of more than 230 subtropical marine
777 species. *R. Soc. Open Sci.* **2,** 150088 (2015).

778 58. Berry, D., Mahfoudh, K. Ben, Wagner, M. & Loy, A. Barcoded primers used in
779 multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* **77,**
780 7846–7849 (2011).

781 59. O'Donnell, J. L., Kelly, R. P., Lowell, N. C. & Port, J. A. Indexed PCR primers induce
782 template- Specific bias in Large-Scale DNA sequencing studies. *PLoS One* **11,**
783 e0148698 (2016).

784 60. Martin, M. Martin, M. Cutadapt removes adapter sequences from high-throughput
785 sequencing reads. EMBnet.journal. 2011. Date of access 05/08/2015. *EMBnet* **17,** 10–
786 12 (2011).

787 61. Joshi, N. & Fass, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for
788 FastQ files (Version 1.33) [Software]. *Available at https://github.com/najoshi/sickle.*
789 2011 (2011).

790 62. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
791 *Bioinformatics* **26,** 2460–2461 (2010).

792 63. Ekrem, T., Willassen, E. & Stur, E. A comprehensive DNA sequence library is essential
793 for identification with DNA barcodes. *Mol. Phylogenet. Evol.* **43,** 530–542 (2007).

794 64. Camacho, C. *et al.* BLAST plus: architecture and applications. *BMC Bioinformatics* **10,**
795 1 (2009).

796 65. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for
797 supporting accessible, reproducible, and transparent computational research in the
798 life sciences. *Genome Biol.* **11,** R86 (2010).

799 66. Caporaso, J. G. *et al.* QIIME allows analysis of high- throughput community
800 sequencing data Intensity normalization improves color calling in SOLiD sequencing.

Nat. Publ. Gr. **7,** 335–336 (2010).

67. Murray, D. C., Coghlan, M. L. & Bunce, M. From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. *PLoS One* **10,** e0124671 (2015).

68. Fonseca, V. G. *et al.* Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.* **1,** 98 (2010).

69. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24,** 1596–1599 (2007).

70. Magurran, A. E. & McGill, B. J. *Biological diversity: frontiers in measurement and assessment*. *Challenges* (Oxford University Press, 2011).

71. Clarke, K.R., & Gorley, R.N. Primer v6: User Manual/Tutorial. 192 (2006).

72. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73,** 3–36 (2011).

Acknowledgements

This work was funded by the Environment Agency (EA) UK, a Knowledge Economy Skills Scholarship (KESS), a Natural Environment Research Council (NERC) NBAF pilot project grant (NBAF824 2013-14) and the Freshwater Biological Association (FBA) (Gilson Le Cren Memorial Award 2014). We thank the EA and Bangor University for support and in particular, Wendy Grail, John Evans, Emlyn Roberts and EA staff for facilitating provision of eDNA grade laboratory working spaces, equipment, and taxonomic identification of chironomid specimens; HPC Wales for allowing use of their systems; Les Ruse and APEM for identification of Chironomidae specimens for Barcoding; Natural Resources Wales for providing historical data. We would also like to thank Florian Leese and two anonymous reviewers whose feedback greatly enhanced earlier versions of the manuscript. We also acknowledge the support of NERC Highlight Topic grant NE/N006216/1. Knowledge Economy Skills Scholarships (KESS) is a pan-Wales higher-level skills initiative led by Bangor University on behalf of the HE sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) convergence programme for West Wales and the Valleys.

Author contributions

IB, SC, GRC: Designed experiment. IB: Performed lab work, fieldwork, bioinformatics and statistical analysis. MS: Performed statistical analysis and data modelling, DL: Contributed in optimisation of analytical pipelines. MH, MC, KW: Participated in experimental design. IB, SC, GRC: Wrote manuscript. IB, SC, GRC, MS, KW, DL, and MH: Edited manuscript.

Additional information

Competing financial interests. The authors declare no competing financial interests.

Figures legends

Figure 1| Venn diagram showing genera richness detected for all three methods. Number of Chironomidae Genera per sample type (purple: eDNA, orange: community DNA, green: taxonomic identification) and the number of genera common between sample types (overlap area). (Four time points, 45 Chironomidae genera).

Figure 2| Animal eDNA β-diversity – nMDS (Sørensen index). a. COIF, **b**. COIS amplicon (eDNA samples only) (N = 32). Solid green circles: 30% similarity cut-off (corresponding to "winter" –"summer" groups), dashed blue circles: 40% similarity cut-off.

Figure 3| Richness patterns for Chironomidae OTUs and genera. a: OTU richness. **b**: Genera richness. Points represent richness values to individual sampling points for eDNA (blue), community DNA (orange) and taxonomic identification of chironomid exuviae (green). Sampling points spanning the winter months (days 36 -190) did not yield data due to very low physical numbers of exuviae. Best fitted, significant lines from polynomial regressions for eDNA samples (blue) and community DNA (orange), plotted against time (x –axis: Sep. 2013 – Sep 2014) (a.) eDNA: p-value = 0.46, community DNA: p-value= 0.008, (b.) eDNA: p-value=0.017, community DNA p-value=0.635 (N = 29 datapoints).

Figure 4 | Abundance patterns for Chironomidae. The sequencing identified Chironomidae OTUs plotted against species frequency across the UK, according to historical data. Environmental DNA (eDNA) samples (blue) and community DNA (orange) are shown along with the best fitted, significant, linear regression model (black line) (R^2 = 0.087, p-value= 0.003).

Tables

875 **Table 1 | Sequence reads versus OTU and sampling method.** Generalized additive model
876 (GAM), explaining OTU sequence abundance relative to OTU taxonomic ID (OTU) and
877 sampling method (eDNA or community DNA - Method) over time. Model estimates and
878 significances of the smoothing terms are given for the most parsimonious models. (R^2 =
879 0.18, df: degrees of freedom, edf: estimated degrees of freedom).

880

	Df	F	p-value
OTU	44	4.688	<0.01
Method	1	0.013	0.908
OTU x Method	44	1.733	0.003

Approximate significance of smooth terms:	**edf**	**F**	**p-value**
s(Time)	2.899	2.561	0.047

881