Article

Ethanol eDNA Reveals Unique Community Composition of Aquatic Macroinvertebrates Compared to Bulk Tissue Metabarcoding in a Biomonitoring Sampling Scheme

Sadhna Fiona Persaud *[ID], Karl Cottenie [ID] and Jennifer Erin Gleason [ID]

Department of Integrative Biology, University of Guelph, Guelph, ON N1G 2W1, Canada; cottenie@uoguelph.ca (K.C.); jgleason@uoguelph.ca (J.E.G.)
* Correspondence: fiona_persaud@outlook.com

Abstract: Freshwater ecosystems provide essential ecosystem services and support biodiversity; however, their water quality and biological communities are influenced by adjacent agricultural land use. Aquatic macroinvertebrates are commonly used as bioindicators of stream conditions in freshwater biomonitoring programs. Sorting benthic samples for molecular identification is a time-consuming process, and this study investigates the potential of ethanol-collected environmental DNA (eDNA) for metabarcoding macroinvertebrates, especially for common bioindicator groups. The objective of this study was to compare macroinvertebrate composition between paired bulk tissue and ethanol eDNA samples, as eDNA could provide a less time-consuming and non-destructive method of sampling macroinvertebrates. We collected benthic samples from streams in Ontario, Canada, and found that community composition varied greatly between sampling methods and that few taxa were shared between paired tissue and ethanol samples, suggesting that ethanol eDNA is not an acceptable substitute. It is unclear why we did not detect all the organisms that were preserved in the ethanol, or the origin of the DNA we did detect. Furthermore, we also detected no difference in community composition for bioindicator taxa due to surrounding land use or water chemistry, suggesting sites were similar in ecological condition.

Keywords: freshwater; macroinvertebrates; biomonitoring; environmental DNA (eDNA); bulk tissue

1. Introduction

Freshwater ecosystems are one of the most diverse ecosystems on the planet, making them an important component in maintaining ecosystem services [1]. Increasing human pressures in urban and agricultural areas can affect the quality of these environments by polluting surface water [2]. Rivers and streams are important because they support diverse biological communities, as well as provide valuable ecosystem services to humans [2,3]. Thus, it is important to understand how these aquatic ecosystems are influenced by human impacts and how to protect and manage them to secure their benefits for both humans and the environment. Understanding the distribution of organisms in freshwater streams is a necessary step to determine the influence of human-induced changes to habitat condition [2] and advance conservation biology.

Streams are especially vulnerable to human impacts due to their tight connection to the surrounding landscape. For instance, agricultural and urban development along streams can cause habitat degradation and water quality impairment by increases in nutrients (e.g., nitrogen), inorganic sediments, and micropollutants [2]. The use of fertilizers contributes to eutrophication, leading to reduced oxygen levels [3]. These changes in water quality parameters impact aquatic biota and cause changes in the community structure of freshwater organisms [4–7]. Macroinvertebrates have commonly been used as bioindicators to assess stream health due to their established sensitivities to changes in water chemistry and habitat conditions [8]. Most biomonitoring programs focus on two groups

of macroinvertebrates that exhibit broad differences in habitat requirements (i.e., those sensitive to environmental changes versus those more tolerant to disturbances).

The insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies), often referred to as EPTs, are all sensitive to pollutants and disturbances in stream habitats, making their presence an indicator of good water quality [8]. EPTs are representative of good water quality due to their need for high dissolved oxygen concentrations and clean waters [9]. They have thus been used as bioindicator species for decades due to their low tolerance of chemical and organic pollutants [10].

The second category of macroinvertebrates (tolerant organisms) includes Chironomidae (non-biting midges), a family of Dipterans which are incredibly diverse and can inhabit a range of water conditions [11,12]. They are important for stream monitoring due to their wide tolerances of environmental gradients; they can be found in nutrient-rich or -poor environments [11]. Because EPTs often only occur in unpolluted sites, chironomids often dominate communities with poor water quality. However, due to their tolerance ranges and ubiquity, it is important to identify them down to the species level to differentiate their species-specific responses to water quality [11]. For example, Beerman, et al. [13] demonstrated the necessity of identifying Chironomidae past the family-level as they can have distinct responses to the manipulation of water quality parameters such as salinity and the addition of sediment. However, this requires considerable taxonomic expertise, and they are often lumped together as a single group which can mask species-level patterns.

Additionally, freshwater Oligochaeta are a completely aquatic group that are tolerant of low dissolved oxygen levels and high organic matter [8]. They have been used in freshwater biomonitoring since the 1960's due to their wide range of pollution tolerance [14]. Similar to Chironomidae, they are a diverse group, and morphological identification is difficult. Most species in this family cannot be identified at an immature state and require dissection, which is impracticable during routine biomonitoring [14].

One of the most common methods of sampling macroinvertebrates to determine community distribution and diversity is kick-net sampling (Figure 1), followed by sample sorting and identification [15]. However, the morphological identification of organisms associated with kick-net sampling often leads to misidentification due to the lack of diagnostic characteristics in some organisms, such as intermediate larval stages or damaged specimens [16,17]. Furthermore, this method is time-consuming and costly [15,17]. Taxonomic challenges such as these have encouraged technological innovations involving molecular identification (e.g., DNA barcoding, [18]). DNA barcoding involves using sequences of DNA from specific genes to identify organisms. For animals, this is often the cytochrome c oxidase I (COI) gene [18]. DNA barcoding is an efficient and accurate method to identify taxa relevant to biomonitoring programs and can yield higher species richness estimates compared to morphological identification [19].
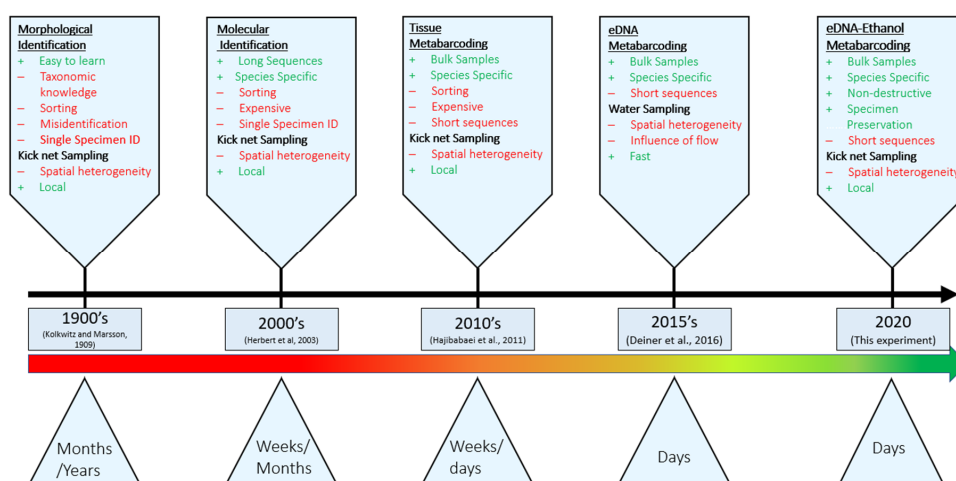

Figure 1. A timeline comparing the pros and cons of aquatic biomonitoring sampling methods.

Metabarcoding has improved the efficiency of DNA barcoding by allowing entire samples to be sequenced at once (instead of single specimens) and lowering the cost substantially [20]. However, one of the drawbacks of this method is the manual separation of organisms when collecting environmental samples due to bycatch in the form of sediment and organic matter [15,20]. Additionally, when using standard methods of bulk tissue sampling, specimens are often homogenized and destroyed during DNA extraction and therefore cannot be further referenced for validation or morphological examination [20].

To address some of these limitations, metabarcoding research began to incorporate environmental DNA (eDNA) [15,17] sampling in lieu of bulk samples. eDNA refers to DNA that is collected from the environment (e.g., water, soil or air), without the isolation of whole organisms [21]. DNA can be transferred to the environment in the form of shed skin cells, reproductive material or waste, and decaying organisms [21]. Several studies have compared traditional bulk tissue samples to eDNA samples collected via water filtration and found a higher species richness with the use of eDNA, and a large overlap in community composition for arthropods and indicator species between the two methods [12,15,22]. However, eDNA is a rapidly evolving field of study and there are still many challenges and unknowns associated with this technique, such as the shedding rates of target taxa and the longevity of DNA in the water column [23]. Hydrolysis of cellular membranes in fast-moving waters such as streams can increase DNA degradation by breaking down the shed DNA [24]. This can account for the absence of some species and a subsequent reduction in biodiversity observed in previous studies comparing sampling methods [24]. While some research suggests that eDNA metabarcoding water samples is an appropriate substitute for kick-net sampling (e.g., [15,21,22]), other studies have preferred bulk sample metabarcoding [24,25]. Gleason et al. [25] compared invertebrate community composition between tissue samples and eDNA samples and found that eDNA mostly amplified algae, with minimal overlap in invertebrate community composition between paired bulk tissue and eDNA samples collected from the same location.

In addition to sequencing DNA collected from filtered water samples, organisms preserved in ethanol can transfer DNA to the ethanol solution [26]. Ethanol is a commonly used preservative and, if the DNA transfer is high enough, could provide a non-destructive technique of sampling macroinvertebrates [26]. Additionally, ethanol in high concentrations ($\geq 95\%$) can denature proteins that cause DNA degradation [27] and can thus allow eDNA to persist longer within a sample. With ethanol-based eDNA, a bulk sample can be used to provide molecular identification without sorting, and sample specimens can be further preserved for future reference [26]. This method may also be more advantageous since some organisms have been observed to regurgitate their stomach content when being preserved, providing detection of prey organisms as a by-product resulting in a more detailed insight into the present community [20].

Although this technique poses a promising alternative to bulk tissue sampling, it has not been thoroughly investigated and it is unclear if ethanol-based DNA can be sequenced successfully in a next-generation sequencing (NGS) workflow since the quantity of DNA being shed into the ethanol solution is unknown [26]. A more recent study addressed this issue by introducing new methods to maximize DNA transfer (ultrasonic irradiation, shaking, and freezing, [20]). Zizka et al. [20] found that not enough DNA from mollusc families was being released, resulting in them being undetected. However, these comparisons have not used the same analysis method as Gleason et al. [25] that compares the different techniques at the local sample level, as opposed to the regional pooled study scale. The local sample level is important to more accurately distinguish differences in community composition between methods.

In this study, we compared the aquatic macroinvertebrate community composition data generated via metabarcoding between eDNA collected from ethanol samples and the associated bulk tissue samples to evaluate the potential of ethanol-based eDNA as a surrogate for bulk sample metabarcoding. We hypothesized that if benthic macroinvertebrates transfer DNA to the ethanol it is stored in, then ethanol eDNA will closely

match the bulk tissue samples because the ethanol samples will contain DNA of only the species collected within the corresponding bulk sample. We tested this prediction using all macroinvertebrates as well as three groups relevant to biomonitoring programs (EPTs, chironomids and oligochaetes).

The second goal of our study was to test whether either method was influenced by environmental condition measured by site type (Conservation area or farm), the percentage of agricultural land use in the watershed, and local water quality parameters. We hypothesized that if aquatic bioindicator species are sensitive to pollutants, then there will be a difference in community structure between impacted and nonimpacted sites. We expected chironomids and oligochaetes would be dominant in impacted streams and EPTs in conservation areas, with water quality varying between the two types of sites.

2. Materials and Methods

2.1. Stream Sampling

We sampled a total of twenty stream sites in southern Ontario, Canada in May 2019 (Figure 2). We chose sites located either on a Conservation Authority property (the minimally impacted sites) or on privately-owned properties with farms within the local catchment.
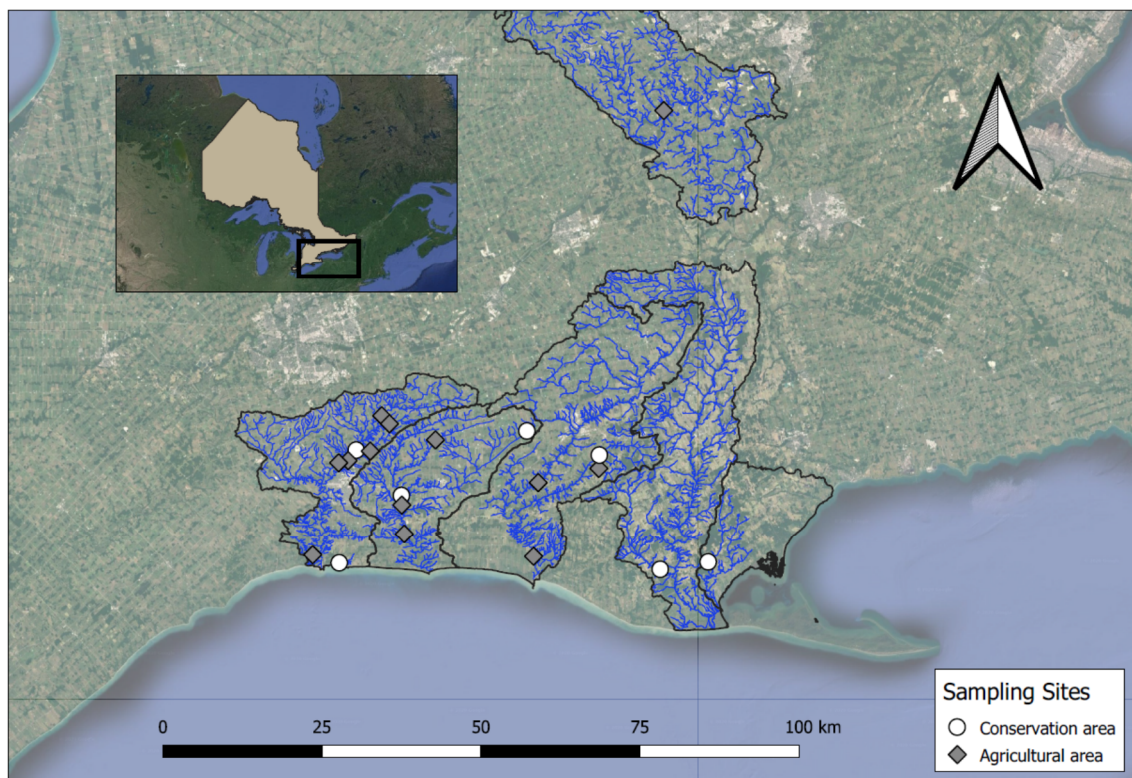
Figure 2. A map of stream sites in southern Ontario, Canada, where aquatic macroinvertebrate samples were collected in May 2019. White circles represent conservation areas and grey diamonds represent privately owned farms.

Prior to collecting benthic samples, we measured water quality (conductivity, total dissolved solids, salinity, dissolved oxygen, pH, dissolved organic matter) in situ using an Exo2 Multiparameter Sonde (YSI Inc., Yellow Springs, OH, USA). We collected benthic macroinvertebrates using a 500 μm mesh D-net and a standard 3-min travelling kick-and-sweep action based on the Ontario Stream Assessment Protocol [28]. We rinsed excess debris from the samples using a 500 μm sieve and preserved the benthic macroinvertebrates bulk samples in 95% ethanol on-site. All sampling equipment was cleaned with a 10% bleach solution and rinsed with de-ionized (DI) water between sites. Samples were

transported back to the laboratory in a cooler and then stored in a fridge at 4 °C until further processing.

2.2. Benthic Sorting and Bulk Tissue DNA Extraction

We separated the ethanol from the bulk sample using a 500 μm sieve that was first cleaned with ELMINase Decontaminant (Thermo Fisher Scientific, Mississauga, Canada) and DI water. The ethanol was drained into a sterilized container before being transferred to falcon tube using a clean funnel. We stored ethanol samples in a −20 °C freezer until further eDNA processing.

After storing the ethanol, we sorted the bulk tissue samples under a dissecting microscope to remove all arthropods, annelids, and molluscs from sample debris and recorded the total abundance and biomass per site. We air-dried the benthic macroinvertebrates after sorting and subsequently ground the samples at 4000 rpm for 30 min using an IKA Tube Mill with 20 mL tubes and ten 4 mm diameter steel beads to ensure homogenization (IKA, Staufen, Germany). Lower abundance samples were ground using a 2 mL tube containing two steel beads using a TissueLyser II (Qiagen, Hilden, Germany) at 30 hz for one minute. We transferred 20 mg (±1 mg) of the ground sample to a sterile 2 mL tube for DNA extraction (the entire sample was used if the total dry biomass was less than 20 mg). We used a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) to extract DNA from the ground tissue powder. We quantified the DNA using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific) and normalized all samples to the same concentrations (5 ng/μL).

2.3. eDNA Filtration and Extraction

Prior to filtration, we sanitized all filtration equipment using a 10% bleach solution for 15 min followed by rinsing with DI water. We inverted the ethanol samples before filtration to resuspend any settled particles. We filtered samples using a mixed cellulose ester membrane filter (pore size 1.0 μm) held within a sampling cup attached to a manifold vacuum pump. We covered the filter cups with Kim Wipes to prevent any particles from falling into the sample. Once samples were filtered (approximately 4–20 min per sample), we placed the filter in a sterile petri dish and stored in the samples at −80 °C until DNA extraction. We used a CTAB extraction protocol for ethanol eDNA (see [25,29]), quantified and normalized DNA concentration of all samples to 5 ng/μL.

2.4. PCR Amplification and Sequencing

We used a two-step PCR approach by first amplifying a 421 base pair region of the mitochondrial CO1 gene (the animal DNA barcode region, [18]) using freshwater macroinvertebrate primers (BF2 + BR2; [30]). Using a Qiagen multiplex PCR kit (Qiagen, Hilden, Germany) the reaction was prepared and had a total volume of 25 μL. Each reaction consisted of 2.5 μL DNA extract (all normalized to 5 ng/μL), 12.5 μL of Qiagen master mix, 9 μL of molecular grade water and 0.5 μL (concentration 0.2 μM) of each primer (BF2 + BR2). We included two technical replicates per sample. We included an additional 12 negative controls that contained 2.5 μL of molecular grade water instead of DNA. The thermocycling profile for the bulk tissue samples was a 95 °C initial denaturation for fifteen minutes, followed by 25 cycles of 94 °C for 30 s, 50 °C for 90 s, 72 °C for 60 s and a final extension at 72 °C for ten minutes. The eDNA samples were run with a modified thermocycling profile that included five cycles at a lower annealing temperature (45 °C) followed by 30 cycles using the above profile (as in Reference [29]). The PCR product was visualized on an agarose gel and purified using NucleoMag NGS clean-up and size select magnetic beads (Macherey-Nagel, Bethlehem, USA) with a ratio of 0.8× [29].

The second PCR used Illumnia indexing primers (set D) to tag samples based on Illumina's standard protocol with the following volumes: 5 μL of the PCR product from the first PCR, 25 μL of Qiagen master mix, 5 μL of each forward and reverse indexing primers (10 μM) and 10 μL of molecular grade water for a total volume of 50 μL. The

thermocycling profile included an initial denaturation at 95 °C for fifteen minutes, 8 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s and a final extension of 72 °C for five minutes. We visualized the PCR product on an agarose gel and purified it using NucleoMag beads at a 0.6× ratio [29]. The final product was again visualized, and the prepared library was submitted to the Advanced Analysis Center at the University of Guelph, Ontario for normalization, pooling, and sequencing. The library was sequenced using 300 paired-end sequencing with 5% PhiX spike in on the Illumina MiSeq platform.

2.5. Bioinformatics

We assessed the raw sequence data quality following the protocol in Gleason et al. [25], using the bio-informatics platform JAMP version 0.67 (http://github.com/VascoElbrecht/JAMP). We used USEARCH (v. 11.0.6668, [31]) to merge the paired-end reads, followed by cutadapt (v. 1.15; [32]) to trim primer sequences, and filtered out low-quality sequences (expected error ≥ 1). Operational Taxonomic Units (OTUs) were created using USEARCH to cluster all reads together at a 97% threshold and singletons were removed. We matched OTU sequences to the Barcode of Life Database (BOLD; [33]) reference library to assign taxonomic information. As in Gleason et al. [25], the maximum number of sequence reads of each OTU present in the negative controls were subtracted from the rest of the dataset, followed by filtering out any non-target taxa (e.g., keeping only Arthropods, Molluscs and Oligochaetes in the dataset). Poor quality taxonomic matches (less than 90%) were also filtered out. The process was also repeated to obtain datasets only including EPTs, chironomids, and oligochaetes (for a total of four data sets). Technical replicates were pooled together. The four final datasets and taxonomic metadata are available as supplementary data (Tables SI1–SI5).

2.6. Statistical Analysis

Data manipulation and statistical analyses were performed using R version 3.5 [34]. Three samples were removed from further analysis due to very low sequence yield, indicating poor sequencing performance. Using the raw data, we calculated the total number of sequences and OTUs belonging to each of our target groups (all macroinvertebrates, EPTs, chironomids, oligochaetes) for bulk tissue samples and ethanol samples. We used stacked bar charts to visualize the overall taxonomic composition of each sample type. To compare community composition between the sample types, we used a paired sample [25] approach (i.e., each pair consisted of a bulk sample and its associated ethanol eDNA sample) and calculated the number of taxa unique to each method and the number of taxa shared (i.e., the taxonomic overlap) between the paired samples. We repeated this analysis for all of the target groups (e.g., all four datasets).

Using presence/absence data, we performed Redundancy Analyses (RDAs) using the 'rda' function in the R package vegan [2] with sample type (e.g., bulk tissue or ethanol) as a constraint to evaluate differences in community composition between the two sampling methods. To test for significant differences, we used an ANOVA on the results of the RDA and incorporated the paired analysis into the model. We repeated these analyses for all of the taxonomic groups of interest.

We used the RDA and ANOVA approach to test three sets of environmental predictor variables for a potentially significant influence on the community composition: (1) local land use (whether the site was on Conservation Authority property or a privately owned farm), (2) regional land use (the percentage of agricultural land use in the catchment) and (3) water quality (using the in situ measurements we took). Here, we treated bulk tissue and ethanol samples separately and performed these analyses for all four target groups and each sampling method. However, we could not use RDAs for the EPT groups as too many sites lacked these taxa and sparsity became an issue. Finally, we used the same statistical approach to determine whether water quality was influenced by site type or land use.

3. Results

The MiSeq run resulted in a total of over 14 million raw sequence reads, and after filtering out low-quality reads and non-target taxa (i.e., keeping only sequences that matched Arthropod, Mollusc and Annelid OTUs with at least 90% similarity), we retained 5,534,585 sequences (Table 1; Figure 3). Our 12 negative controls contained a relatively high proportion of sequences (mean 191,272.2 per sample $\pm$ 97,280.83 SD); however, most of these sequences (86%) were either not macroinvertebrates or did not meet our quality filtration thresholds. All sequences have been deposited in NCBI's Sequence Read Archive (PRJNA649592). We observed substantially more sequences belonging to tissue samples (over four million) compared to the ethanol samples (under one million), despite a closer similarity in raw read amount (Table 1; Figure 3). Tissue samples subsequently also had a higher number of high-quality reads for all bio-monitoring relevant groups (EPTs, chironomids, oligochaetes).

Table 1. Sequence reads and total OTU counts based on raw sequence reads, after quality control and for all four of the final dataframes (*): all aquatic macroinvertebrates, chironomids, oligochaetes and EPTs. Values are presented for all sequences combined (Overall) and separately for bulk tissue samples and ethanol eDNA samples.

|	Overall		Tissue		Ethanol	
	Sequences	**OTUs**	**Sequences**	**OTUs**	**Sequences**	**OTUs**
Raw	14,369,501	7792	6,518,181	599	4,840,919	6438
Filtration	8,033,945	6227	4,998,042	476	3,035,903	5960
Invertebrates *	5,534,585	841	4,628,695	421	905,890	602
Chironomids *	3,153,709	228	2,810,963	191	342,746	123
Oligochaetes *	997,964	125	582,943	56	415,021	115
EPTs *	751,336	54	691,597	42	59,739	27
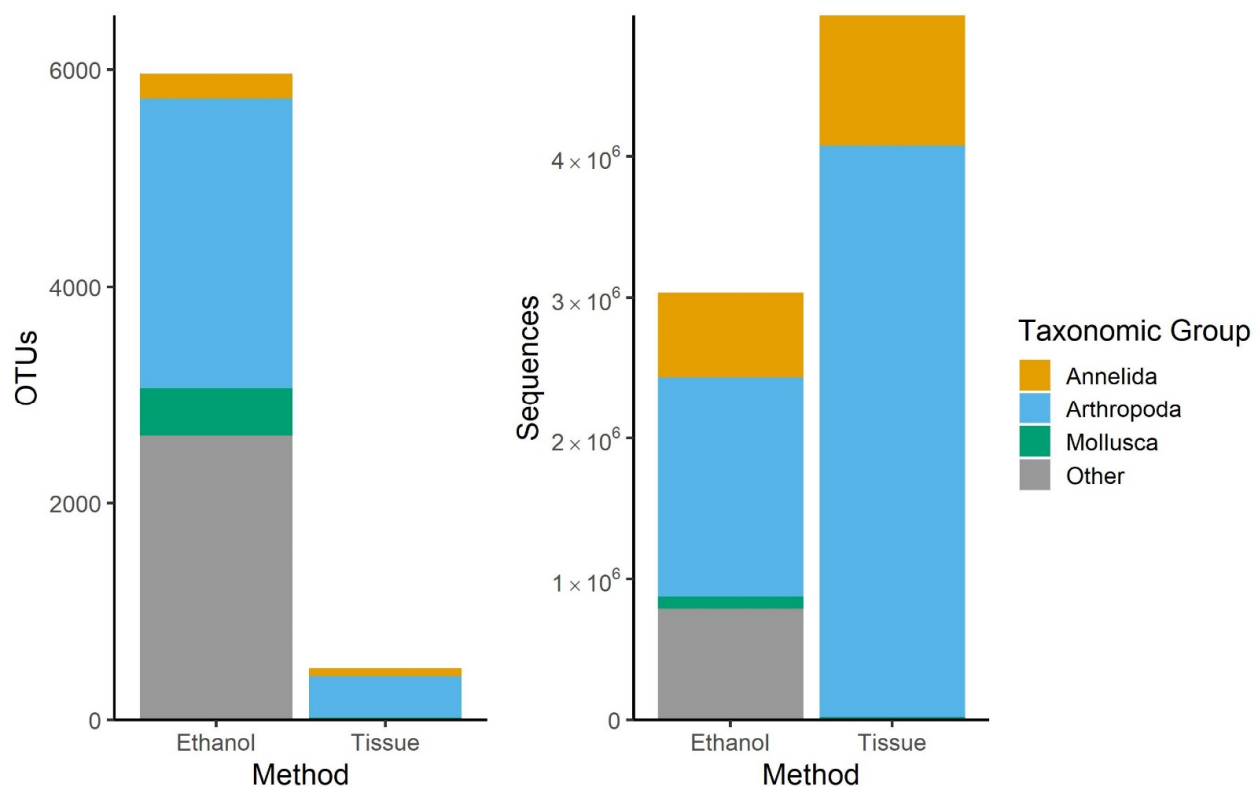
Figure 3. The number of OTUs and total sequences made up by the target taxa (Annelida, Arthropoda, Mollusca) for both ethanol and tissue samples.

While ethanol samples contained ten times as many raw OTU counts than tissue samples (approximately 6000 vs. 600), many of these OTUs were removed during the quality control process (poor match similarity) or were not benthic macroinvertebrates. While 70% of the tissue OTUs were retained in the final dataset, less than 10% of the ethanol OTUs were kept (see Table 1). After all quality control and filtration steps, there was a total of 841 macroinvertebrate OTUs. Out of this total, 54 of the OTUs were EPTs, 228 were chironomids, and 125 were oligochaetes. Tissue samples had more EPT and chironomid OTUs, while ethanol samples contained more oligochaete OTUs.

3.1. Comparison of Sampling Methods

When comparing all macroinvertebrate OTUs, we observed a very small overlap in the identity of OTUs between paired bulk tissue and ethanol samples (0 to 12 OTUs overlap, Figure 4A). All target groups (EPTs, chironomids, oligochaetes) similarly displayed very minimal overlap in community composition at the paired sample level (Figure 4B–D). EPTs had no shared taxa between paired tissue and ethanol samples. We tested for differences in community composition between tissue and ethanol samples for all four groups (all invertebrates, EPTs, chironomids, oligochaetes) and observed significant differences for all invertebrates ($F = 1.5917$, $p = 0.001$; Figure 5A), and oligochaetes ($F = 1.5154$, $= 0.017$; Figure 5C). EPTs did not have enough OTUs present for this analysis, but based on the absence of overlap in Figure 4D, the community composition is likely significantly different. However, there was no significant difference in Chironomidae community composition between paired tissue and ethanol samples ($F = 1.33$, $p = 0.111$; Figure 5B)
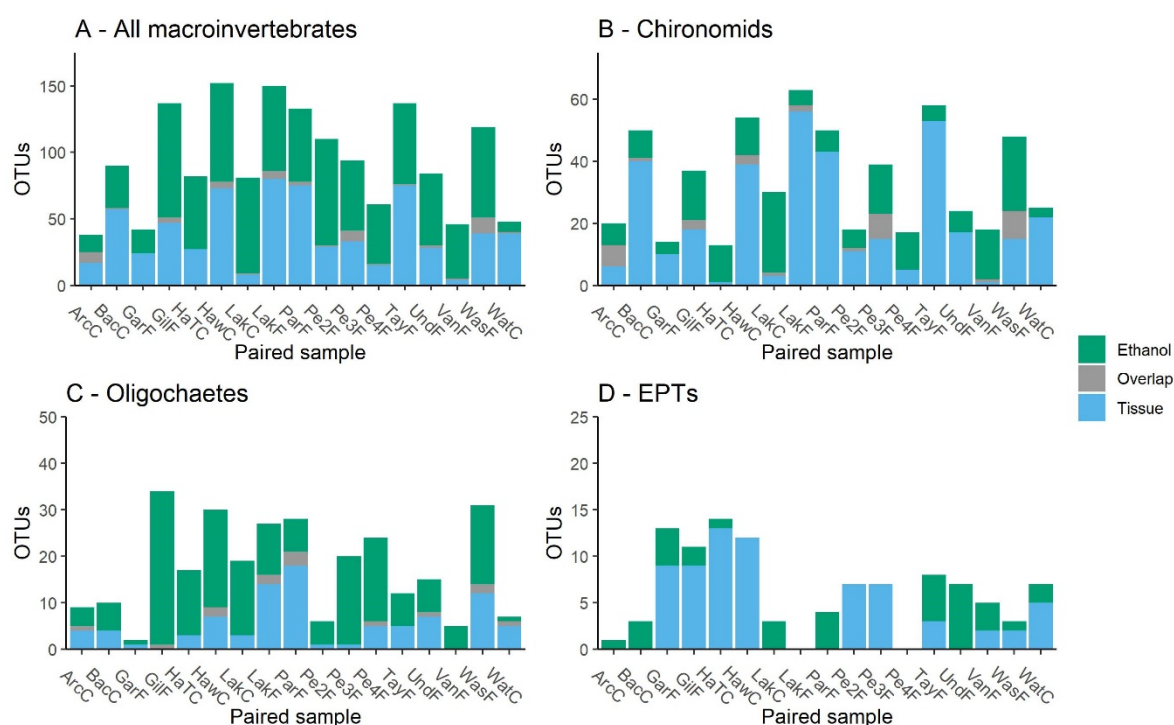
Figure 4. Each bar represents a paired sample (a bulk tissue sample and the associated ethanol eDNA sample from its preservative) for (**A**) all macroinvertebrates, (**B**) chironomids, (**C**) oligochates, and (**D**) EPTs. OTUs unique to the tissue sample are represented by a blue bar and those unique to the ethanol sample are represented by a green bar. The number of OTUs shared between a paired tissue and ethanol sample are represented by a grey bar in the middle of a column. The absence of a grey middle bar indicates that no OTUs were shared.
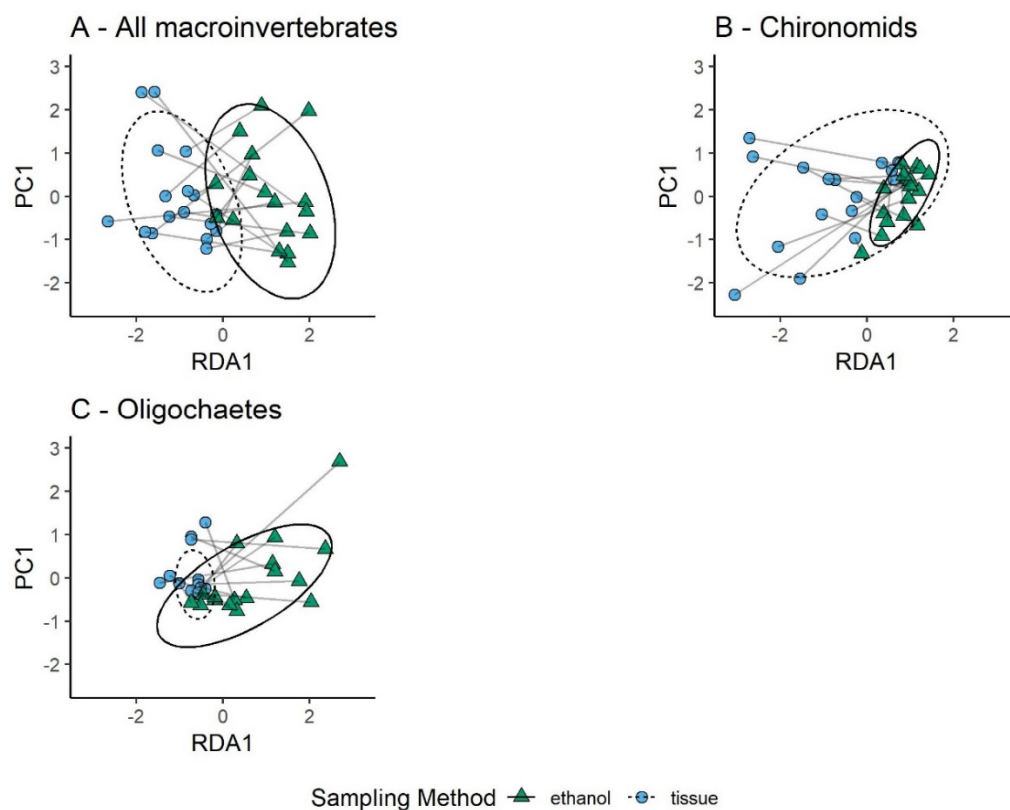
Figure 5. The redundancy analysis plots for (**A**) all macroinvertebrates, (**B**) chironomids, and (**C**) oligochaetes. Tissue samples are represented by blue circles and ethanol samples are represented by green triangles, each with 90% confidence ellipses. Paired samples (e.g., the ethanol preservative from its associated bulk tissue sample) are joined by a grey line. EPTs are not represented here as tissue and ethanol sampling did not share any EPT OTUs.

3.2. Agricultural Gradient Comparison

We also tested whether each of the groups had significant relationships to site type (e.g., privately owned land or conservation area), the percentage of agricultural land use (e.g., crops) in the watershed, or the water quality parameters we measured and found no significant relationships for any group or parameter (see Table 2).

Table 2. The results of all RDA-based ANOVAs for sampling method (tissue or ethanol), site type (Conservation area or farm), land use percentage, and water quality variables. **S** represents a significant *p*-value (<0.05) and NS represents a *p*-value >0.05. We could not perform RDAs for EPTs as the dataset was too sparse, but based on the overlap plot (Figure 4D) we have listed the sampling method as significant (S*) as no OTUs were shared for EPTs between tissue and ethanol samples.

	All Invertebrates		Chironomids		Oligochaetes		EPTs	
Method	**S**		**NS**		**S**		**S***	
	Tissue	**Ethanol**	**Tissue**	**Tissue**	**Ethanol**	**Ethanol**	**Tissue**	**Ethanol**
Site Type	NS	NS	NS	NS	NS	NS	n/a	n/a
Land use %	NS	NS	NS	NS	NS	NS	n/a	n/a
Water quality	NS	NS	NS	NS	NS	NS	n/a	n/a

3.3. Water Quality and Land Use

We found no influence of water quality parameters on the RDA model using site type (e.g., farm or conservation area) as a predictor variable (F = 9.162, *p* = 0.418). However, when the same parameters were fit using the percentage of agricultural land use in the surrounding catchment (as opposed to the binary farm or conservation area category), we found that water quality measurements were correlated with land use (F = 3.996, *p* = 0.021).

4. Discussion

Freshwater ecosystems provide essential services to humans, such as fresh drinking water, and their maintenance crucial [1]. As increased water pollution in agricultural areas can alter community composition in freshwater streams [3,35], biomonitoring programs provide crucial information about the status of a stream and its inhabitants [9]. Previously used biomonitoring techniques using bulk tissue samples can be time-consuming and costly [15,17], and we investigated the efficacy of ethanol collected eDNA as a substitute with a focus in bioindicator taxa.

4.1. Ethanol and Bulk Tissue Community Composition

We found that overall, ethanol samples had more OTUs compared to tissue samples (Figure 3; Table 1). Chironomids were the only group represented by more OTUs in tissue samples than eDNA samples (Table 1) and we are unsure why they are the only group which differed here. Of the biomonitoring groups, chironomids had the greatest OTU richness, followed by oligochaetes, and then EPTs. In addition to calculating OTUs, we also determined the number of sequences for each target group. We found that tissue samples had the most sequences, and most of those sequences were chironomids for tissue and oligochaetes for ethanol (Table 1). Similar to a previous study conducted using water eDNA samples, ethanol eDNA detected more OTUs than tissue samples; however, most of them were from non-target groups [25]. Similar to Gleason et al. [25], we determined that tissue-based metabarcoding contained more target sequence reads than eDNA, despite a smaller number of OTUs (Figure 3).

The community composition varied greatly between sampling methods, with very little taxonomic overlap between paired samples. In general, there were more unique chironomids and EPT OTUs in bulk tissue samples, whereas ethanol samples contained more unique oligochaete OTUs. There was minimal overlap between paired samples for all groups, and EPTs did not share any OTUs in common between sampling methods. Thus, our prediction that metabarcoding bulk tissue samples and ethanol-eDNA will be similar in taxonomic composition due to the release of DNA into ethanol is not supported by this study. Given the discordance between methods, biomonitoring procedures could consider both methods to accurately detect the taxonomic composition of a site. However, the origin of the DNA detected only in samples associated with ethanol is not known. The most likely possibility is that it is present in the sediment substrate associated with the sampling. If that is the case, though, then using ethanol-associated eDNA (solely or in combination with other methods) will need additional studies to determine the origin of sediment-associated DNA.

Previous studies comparing water-collected eDNA to kicknet sampling also detected differences in community composition, with eDNA having higher richness for some groups and kicknet for others [15,22]. This difference in community composition was likely due to the influence of water movement, transporting DNA from upstream sites covering a larger sampling area compared to the more local scale kicknet sample [22,25,36]. In comparison to water eDNA research, methods using eDNA collected from ethanol are scarce. Additionally, there has not been another study using the same direct comparison at the site level as in our study. Previous studies using ethanol collected eDNA have detected a substantial overlap in taxa found both in bulk tissue and ethanol samples, suggesting ethanol is a suitable substitute for bulk tissue sampling [20,26,37]. However, in the study conducted by Zizka et al. [20], although there was a large overlap found between methods for the EPT taxa, an overall comparison of all detected taxa without filtering out any families revealed very distinct communities between sampling methods. The majority of OTUs from the ethanol sample in this study were from algae, and there were more unique macroinvertebrates in the bulk tissue samples [20]. Similar to the results of our study, Martins et al. [38] found differences in taxonomic communities between sampling methods. This study found that taxon detection was positively related to abundance in ethanol samples [38]. Our results are thus concordant with previous studies which

display unique communities between tissue and ethanol metabarcoding. The reason for the discrepancy on the success of ethanol between those studies and our results is unclear. It is currently unknown why we did not detect all the organisms that were preserved in the ethanol, or where the DNA that we did detect in ethanol but not in the associated tissue sample came from. The lack of detection may be caused by specimens with a smaller body size releasing less DNA into the ethanol preservation or the presence of an exoskeleton preventing DNA from being released [20,39]. Biomass is potentially a source of error, and the overamplification of larger taxa may add to the discordance between sampling methods [38,39]. Additionally, the amount of DNA being shed by each organism is unclear. Although it has been previously demonstrated that DNA can successfully be obtained from ethanol preservation [40], it is possible that not enough DNA is being released from specimens, or that they are being released at a slow rate, resulting in a difference in OTU richness. On the other hand, a potential source of the DNA found in ethanol but not tissue samples is that it comes from the substrate and vegetation within the bulk sample that could have macroinvertebrate DNA on it and released it into the ethanol preservation. If there is macroinvertebrate DNA within the substrate in addition to whole specimens, then this may account for the difference in community composition within samples.

In addition to these biological unknowns surrounding the transfer of eDNA to ethanol, there are also technical biases which can also result in the discrepancy between community composition between our sample methods (see Reference [38] for summary). For example, the choice of DNA extraction method can result in biases in community data [41]. We used two different extraction methods here (Qiagen DNEasy kits for tissue and a CTAB buffer extraction for eDNA) to maximize respective yield and quality; however, this may have introduced some of the differences we observed in our dataset. We also selected a lower annealing temperature for the ethanol samples to optimize amplification [29], which can also result in biases. Finally, the choice of primer pair and marker is essential to any metabarcoding study [42], and ultimately any comparison of methods will be specific to these choices (e.g., COI and BF2 + BF2 here). We selected our primer set as it has been very successful at representing aquatic macroinvertebrate communities (see Reference [30]); however, it is a degenerate primer that, when used with eDNA, can result in the high-levels of non-target amplification that we see here [25,43]. Leese et al. [40] had success using a COI primer set with limited degeneracy to reduce non-target amplification and improve macroinvertebrate yield with water-based eDNA samples. While our selected primer set (BF2 + BR2) worked well for tissue samples, future work sequencing sample preservative could test more conserved primer sets. However, for the purpose of this study, using the same primer set for both sample types facilitated a more direct OTU comparison.

Future research to resolve these differences should incorporate a mock community and observe the effects of time on DNA release into ethanol. The mock communities should ideally be composed of bioindicator species to reflect relevance to biomonitoring programs. Future studies may also want to compare species richness by using a mock community without any substrate in the preservative to observe the effectiveness of DNA transfer in ethanol, followed by another sample with only substrate for comparison. The substrate may be removed and added to a mock community to observe the effects of DNA transfer from sampling debris. Although ethanol collected DNA has provided promising benefits for freshwater biomonitoring, the results of this study demonstrate that there is a range of unanswered questions that currently prevent it from acting as an acceptable substitute for bulk tissue sampling.

4.2. Influence of Agriculture and Water Quality

In addition to these methodological comparisons, we studied the potential impacts of agriculture and water quality on benthic communities. However, we found that there was no significant difference in community composition between site types (agriculture/conservation area) for all taxonomic groups and sampling methods (Table 2). Additionally, there were no significant influences of agricultural land use within the watershed

or local water quality on community composition. If the water quality parameters in this study are reflective of stream health, then our study suggests that these streams are similar in ecological condition or that the variability is determined by processes not measured by one of the variables included in our study. Thus, the results of this study indicate a lack of agricultural gradient across study sites. Further research is required to determine if these macroinvertebrates communities are adapted to agricultural water conditions and if there is a difference in community composition between varying water qualities. Studies have shown that as agricultural land cover increases within a watershed, habitat quality tends to degrade at the watershed scale, resulting in a regional change in community composition [44,45]. Based on the results in our study, the sites appeared to have a similar amount of agricultural land use, resulting in communities with seemingly random variation in community composition.

The lack of local effects may be a result of the large-scale agricultural land use within the Lake Erie watershed. Previous studies observing an agricultural gradient in Southern Ontario suggested that macroinvertebrate communities in these streams may be relatively homogenous due to intense agriculture within the watershed [46,47]. Water quality results in this study were influenced by the percentage of agricultural land use rather than site type. Southwestern Ontario is a very developed area of the province and the impacts of agriculture have been observed for a long time [48]. Therefore, all streams in our study area, including those in conservation areas, can be influenced by agriculture in the watershed. Thus, lumping sites into categories based on the immediate surrounding landscape (e.g., conservation area or farm) is potentially not the best reflection of agricultural impact. This suggests that a more robust look at regional land-use composition is required to create a more representative gradient. We did detect differences in water quality parameters that were related to the percentage of agricultural land use in the catchment, thus a more accurate reflection of land use might be to include the percentage of cropping in the regional catchment, rather than a binary category of local impacts (e.g., some conservation areas are small and surrounded by farms upstream).

Additionally, it is possible we did not include the relevant water quality parameters. For instance, we did not quantify nutrient levels or pesticide concentrations. Streams may become polluted with nutrients leading to eutrophication as a consequence of agriculture, and therefore, measuring nutrients such as nitrogen and phosphorus may be beneficial in future studies [3,35]. Observing the levels of pesticides may also be beneficial to understand the impacts of agriculture on stream communities, as they have been shown to alter community structure in freshwater streams [48]. Observing additional water quality parameters may provide more insight into the agricultural gradient in Southern Ontario, as well as the impacts of agriculture on stream communities.

5. Conclusions

In conclusion, the results of this study indicate that ethanol collected eDNA metabarcoding is not an effective substitute to kicknet-based bulk tissue metabarcoding as there was very little overlap in community composition between paired samples. Although ethanol eDNA provides a promising future to freshwater biomonitoring, it still requires further research to ensure that it is effective at representing the bulk sample as it is currently unclear why all taxa within the bulk sample are not detected in the ethanol. Developing a time and cost-efficient method of macroinvertebrate sampling is essential to monitor the impacts of agriculture in developing areas such as southern Ontario, and we suggest that while bulk tissue samples are currently the most reliable option, preservative-based eDNA presents many opportunities for future research.

ethanol) for each of the four taxonomic groups (all macroinvertebrates, chironomids, oligochaetes, EPTs). SI5 contains the relevant taxonomic metadata for each OTU.

References

1. Bolund, P.; Hunhammar, S. Ecosystem Services in Urban Areas. *Renew. Agric. Food Syst.* **1999**, *29*, 293–301. [CrossRef]
2. Burdon, F.J.; Munz, N.A.; Reyes, M.; Focks, A.; Joss, A.; Räsänen, K.; Altermatt, F.; Eggen, R.I.L.; Stamm, C. Agriculture versus Wastewater Pollution as Drivers of Macroinvertebrate Community Structure in Streams. *Sci. Total Environ.* **2019**, *659*, 1256–1265. [CrossRef]
3. Xu, M.; Wang, Z.; Duan, X.; Pan, B. Effects of Pollution on Macroinvertebrates and Water Quality Bio-Assessment. *Hydrobiologia* **2014**, *729*, 247–259. [CrossRef]
4. Munn, M.D.; Black, R.W.; Gruber, S.J. Response of Benthic Algae to Environmental Gradients in an Agriculturally Dominated Landscape. *J. N. Am. Benthol. Soc.* **2002**, *21*, 221–237. [CrossRef]
5. Wang, H.; Hondzo, M.; Xu, C.; Poole, V.; Spacie, A. Dissolved Oxygen Dynamics of Streams Draining an Urbanized and an Agricultural Catchment. *Ecol. Modell.* **2003**, *160*, 145–161. [CrossRef]
6. Morgan, A.M.; Royer, T.V.; David, M.B.; Gentry, L.E. Relationships among Nutrients, Chlorophyll-a, and Dissolved Oxygen in Agricultural Streams in Illinois. *J. Environ. Qual.* **2006**, *35*, 1110–1117. [CrossRef] [PubMed]
7. Olson, J.R.; Hawkins, C.P. Effects of Total Dissolved Solids on Growth and Mortality Predict Distributions of Stream Macroinvertebrates. *Freshw. Biol.* **2017**, *62*, 779–791. [CrossRef]
8. Mcdonald, B.S.; Mullins, G.W.; Lewis, S. Macroinvertebrates as Indicators of Stream Health. *Am. Biol. Teach.* **1991**, *53*, 462–466. [CrossRef]
9. Johnson, R.K.; Wiederholm, T.; Rosenberg, D.M. *Freshwater Biomonitoring Using Individual Organisms, Populations, and Species Assemblages of Benthic Macroinvertebrates*; Chapman Hall: New York, NY, USA, 1993; pp. 40–125. [CrossRef]
10. Lenat, D.R. Water Quality Assessment of Streams Using a Qualitative Collection Method for Benthic Macroinvertebrates. *J. N. Am. Benthol. Soc.* **1988**, *7*, 222–233. [CrossRef]
11. Nicacio, G.; Juen, L. Chironomids as Indicators in Freshwater Ecosystems: An Assessment of the Literature. *Insect Conserv. Divers.* **2015**, *8*, 393–403. [CrossRef]
12. Bista, I.; Carvalho, G.R.; Walsh, K.; Seymour, M.; Hajibabaei, M.; Lallias, D.; Christmas, M.; Creer, S. Annual Time-Series Analysis of Aqueous EDNA Reveals Ecologically Relevant Dynamics of Lake Ecosystem Biodiversity. *Nat. Commun.* **2017**, *8*. [CrossRef] [PubMed]
13. Beermann, A.J.; Zizka, V.M.A.; Elbrecht, V.; Baranov, V.; Leese, F. DNA Metabarcoding Reveals the Complex and Hidden Responses of Chironomids to Multiple Stressors. *Environ. Sci. Eur.* **2018**, *30*. [CrossRef]
14. Vivien, R.; Wyler, S.; Lafont, M.; Pawlowski, J. Molecular Barcoding of Aquatic Oligochaetes: Implications for Biomonitoring. *PLoS ONE* **2015**, *10*, 1–15. [CrossRef] [PubMed]
15. Mächler, E.; Deiner, K.; Steinmann, P.; Altermatt, F. Utility of Environmental DNA for Monitoring Rare and Indicator Macroinvertebrate Species. *Freshw. Sci.* **2014**, *33*, 1174–1183. [CrossRef]
16. Carew, M.E.; Pettigrove, V.J.; Metzeling, L.; Hoffmann, A.A. Environmental Monitoring Using Next Generation Sequencing: Rapid Identification of Macroinvertebrate Bioindicator Species. *Front. Zool.* **2013**, *10*. [CrossRef]
17. Fernández, S.; Rodríguez, S.; Martínez, J.L.; Borrell, Y.J.; Ardura, A.; García-Vázquez, E. Evaluating Freshwater Macroinvertebrates from EDNA Metabarcoding: A River Nalón Case Study. *PLoS ONE* **2018**, *13*, 1–17. [CrossRef]

18. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; DeWaard, J.R. Biological Identifications through DNA Barcodes. *Proc. R. Soc. B Biol. Sci.* **2003**, *270*, 313–321. [CrossRef]

19. Jackson, J.K.; Battle, J.M.; White, B.P.; Pilgrim, E.M.; Stein, E.D.; Miller, P.E.; Sweeney, B.W. Cryptic Biodiversity in Streams: A Comparison of Macroinvertebrate Communities Based on Morphological and DNA Barcode Identifications. *Freshw. Sci.* **2014**, *33*, 312–324. [CrossRef]

20. Zizka, V.M.A.; Leese, F.; Peinert, B.; Geiger, M.F. DNA Metabarcoding from Sample Fixative as a Quick and Voucher-Preserving Biodiversity Assessment Method. *Genome* **2019**, *62*, 122–136. [CrossRef]

21. Taberlet, P.; Coissac, E.; Hajibabaei, M.; Rieseberg, L.H. Environmental DNA. *Mol. Ecol.* **2012**, *21*, 1789–1793. [CrossRef]

22. Deiner, K.; Fronhofer, E.A.; Mächler, E.; Walser, J.C.; Altermatt, F. Environmental DNA Reveals That Rivers Are Conveyer Belts of Biodiversity Information. *Nat. Commun.* **2016**, *7*. [CrossRef] [PubMed]

23. Barnes, M.A.; Turner, C.R. The Ecology of Environmental DNA and Implications for Conservation Genetics. *Conserv. Genet.* **2016**, *17*, 1–17. [CrossRef]

24. Hajibabaei, M.; Porter, T.M.; Robinson, C.V.; Baird, D.J.; Shokralla, S.; Wright, M.T.G. Watered-down Biodiversity? A Comparison of Metabarcoding Results from DNA Extracted from Matched Water and Bulk Tissue Biomonitoring Samples. *PLoS ONE* **2019**, *14*, 1–16. [CrossRef] [PubMed]

25. Gleason, J.E.; Elbrecht, V.; Braukmann, T.W.A.; Hanner, R.H.; Cottenie, K. Assessment of Stream Macroinvertebrate Communities with EDNA Is Not Congruent with Tissue-Based Metabarcoding. *Mol. Ecol.* **2020**. [CrossRef] [PubMed]

26. Hajibabaei, M.; Spall, J.L.; Shokralla, S.; van Konynenburg, S. Assessing Biodiversity of a Freshwater Benthic Macroinvertebrate Community through Non-Destructive Environmental Barcoding of DNA from Preservative Ethanol. *BMC Ecol.* **2012**, *12*, 1–10. [CrossRef] [PubMed]

27. Stein, E.D.; White, B.P.; Mazor, R.D.; Miller, P.E.; Pilgrim, E.M. Evaluating Ethanol-Based Sample Preservation to Facilitate Use of DNA Barcoding in Routine Freshwater Biomonitoring Programs Using Benthic Macroinvertebrates. *PLoS ONE* **2013**, *8*, 1–7. [CrossRef]

28. Stanfield, L. *Ontario Stream Assessment Protocol, 9.0*; Fisheries Policy Section; Ontario Ministry of Natural Resources: Peterborough, ON, Canada, 2017.

29. Milián-García, Y.; Young, R.; Madden, M.; Bullas-Appleton, E.; Hanner, R.H. Optimization and Validation of a Cost-Effective Protocol for Biosurveillance of Invasive Alien Species. *Ecol. Evol.* **2020**. [CrossRef]

30. Elbrecht, V.; Leese, F. Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Front. Environ. Sci.* **2017**, *5*, 1–11. [CrossRef]

31. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]

32. Ratnasingham, S.; Hebert, P.D.N. The Barcode of Life Data System. *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef]

33. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.

34. Oksanen, J.; Blanchet, G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. *Vegan:Community Ecology Package*; 2019. Available online: https://CRAN.R-project.org/package=vegan (accessed on 16 January 2021).

35. Macher, J.N.; Vivancos, A.; Piggott, J.J.; Centeno, F.C.; Matthaei, C.D.; Leese, F. Comparison of Environmental DNA and Bulk-Sample Metabarcoding Using Highly Degenerate Cytochrome c Oxidase I Primers. *Mol. Ecol. Resour.* **2018**, *18*, 1456–1468. [CrossRef] [PubMed]

36. Martins, F.M.S.; Galhardo, M.; Filipe, A.F.; Teixeira, A.; Pinheiro, P.; Paupério, J.; Alves, P.C.; Beja, P. Have the Cake and Eat It: Optimizing Nondestructive DNA Metabarcoding of Macroinvertebrate Samples for Freshwater Biomonitoring. *Mol. Ecol. Resour.* **2019**, *19*, 863–876. [CrossRef] [PubMed]

37. Martins, F.M.S.; Porto, M.; Feio, M.J.; Egeter, B.; Bonin, A.; Serra, S.R.Q.; Taberlet, P.; Beja, P. Modelling Technical and Biological Biases in Macroinvertebrate Community Assessment from Bulk Preservative Using Multiple Metabarcoding Markers. *Mol. Ecol.* **2020**, 1–18. [CrossRef]

38. Erdozain, M.; Thompson, D.G.; Porter, T.M.; Kidd, K.A.; Kreutzweiser, D.P.; Sibley, P.K.; Swystun, T.; Chartrand, D.; Hajibabaei, M. Metabarcoding of Storage Ethanol vs. Conventional Morphometric Identification in Relation to the Use of Stream Macroinvertebrates as Ecological Indicators in Forest Management. *Ecol. Indic.* **2019**, *101*, 173–184. [CrossRef]

39. Shokralla, S.; Singer, G.A.C.; Hajibabaei, M. Direct PCR Amplification and Sequencing of Specimens' DNA from Preservative Ethanol. *Biotechniques* **2010**, *48*, 305–306. [CrossRef] [PubMed]

40. Majaneva, M.; Diserud, O.H.; Eagle, S.H.C.; Hajibabaei, M.; Ekrem, T. Choice of DNA Extraction Method Affects DNA Metabarcoding of Unsorted Invertebrate Bulk Samples. *Metabarcoding Metagenom.* **2018**, *2*, 1–12. [CrossRef]

41. Ficetola, G.F.; Boyer, F.; Valentini, A.; Bonin, A.; Meyer, A.; Dejean, T.; Gaboriaud, C.; Usseglio-Polatera, P.; Taberlet, P. Comparison of Markers for the Monitoring of Freshwater Benthic Biodiversity through DNA Metabarcoding. *Mol. Ecol.* **2020**, 1–14. [CrossRef]

42. Leese, F.; Sander, M.; Buchner, D.; Elbrecht, V.; Haase, P.; Zizka, V. Improved Freshwater Macroinvertebrate Detection from EDNA through Minimized Non-Target Amplification. *bioRxiv* **2020**. [CrossRef]

43. Genito, D.; Gburek, W.J.; Sharpley, A.N. Response of Stream Macroinvertebrates to Agricultural Land Cover in a Small Watershed. *J. Freshw. Ecol.* **2002**, *17*, 109–119. [CrossRef]

44. Hepp, L.U.; Milesi, S.V.; Biasi, C.; Restello, R.M. Effects of Agricultural and Urban Impacts on Macroinvertebrates Assemblages in Streams (Rio Grande Do Sul, Brazil). *Zoologia* **2010**, *27*, 106–113. [CrossRef]
45. Yates, A.G.; Bailey, R.C. Covarying Patterns of Macroinvertebrate and Fish Assemblages along Natural and Human Activity Gradients: Implications for Bioassessment. *Hydrobiologia* **2010**, *637*, 87–100. [CrossRef]
46. Yates, A.G.; Bailey, R.C. Effects of Taxonomic Group, Spatial Scale and Descriptor on the Relationship between Human Activity and Stream Biota. *Ecol. Indic.* **2011**, *11*, 759–771. [CrossRef]
47. Barton, D.R. The Use of Percent Model Affinity to Assess the Effects of Agriculture on Benthic Invertebrate Communities in Headwater Streams of Southern Ontario, Canada. *Freshw. Biol.* **1996**, *36*, 397–410. [CrossRef]
48. McKnight, U.S.; Rasmussen, J.J.; Kronvang, B.; Binning, P.J.; Bjerg, P.L. Sources, Occurrence and Predicted Aquatic Impact of Legacy and Contemporary Pesticides in Streams. *Environ. Pollut.* **2015**, *200*, 64–76. [CrossRef] [PubMed]