Taxonomic resolution based on DNA barcoding affects environmental signal in metacommunity structure

Gillian K. Martin[1,2,3,4], Sarah J. Adamowicz[1,2,5], and Karl Cottenie[2,6]

[1]Biodiversity Institute of Ontario, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada N1G 2W1
[2]Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada N1G 2W1

Abstract: Invertebrate communities in freshwater streams form the basis of many biomonitoring protocols that rely on coarse taxonomic resolution. Coarse resolution may group together species with different environmental preferences, masking the relationship between taxonomic composition and environmental variables. Alternatively, closely related species often share similar traits, and therefore, refining taxonomic resolution will not affect, or even may decrease, the strength of relationships between taxa and environment. To test these competing hypotheses, we characterized the metacommunity patterns of 5 orders (Coleoptera, Diptera, Ephemeroptera, Plecoptera, Trichoptera) and 1 subclass (Oligochaeta) of invertebrates in 21 streams in Algonquin Provincial Park, Ontario, Canada. We determined community composition at family, genus, and species (DNA barcode cluster) levels. For each taxonomic level within each order, we computed the amount of variation explained by 20 local environmental and spatial variables. We also compared site diversity rankings, based on richness, Shannon index, and Simpson index values, between species–genus, genus–family, and species–family levels of resolution. We found evidence to support our 2nd hypothesis. Also, reducing taxonomic resolution decreased the consistency of site rankings for all 3 diversity indices. These results suggest the ecological interchangeability of species within genera, or even within families, given the environmental variables we measured. Furthermore, these results emphasize the importance of carefully considering taxonomic resolution for metacommunity work as well as biomonitoring.

Key words: metacommunity, taxonomic sufficiency, community assembly, macroinvertebrate, stream, DNA barcoding

Biomonitoring methods have been used extensively to monitor human effects on freshwater systems (Bonada et al. 2006). Understanding the effect of taxonomic resolution in indicator taxa is of particular interest because the use of rapid bioassessment protocols relying on only high-level identifications has increased (Jones 2008). Achieving fine-scale taxonomic resolution is limited by technical and financial issues, including large incremental increases in cost and time (Marshall et al. 2006), loss of taxonomists (Jones 2008), and difficulties identifying the large number of immature specimens collected (Lenat and Resh 2001). However, as costs for species-level identification based on DNA barcoding have decreased, investigators have advocated barcoding use in biomonitoring (Smith and Fisher 2009, Valentini et al. 2009, Hajibabaei et al. 2011, Sweeney et al. 2011).

Results of previous investigations suggest that taxonomic resolution should significantly affect our ability to characterize and understand communities, and a lack of biological knowledge at the species level may mask community differences (Lenat and Resh 2001). For example, significant differences have been detected in feeding patterns and habitat preferences among species in the trichopteran family Hydropsychidae (Gordon and Wallace 1975, Wallace 1975). Resh and Unzicker (1975) found that pollution tolerances varied among species in the trichopteran genus *Ceraclea*, which contained species that fell into each of 3 tolerance categories (intolerant, facultative, and tolerant). Despite the large number of papers on the effect of taxonomic resolution on the ability to detect differences in community composition, no consistent conclusion has been drawn across a wide variety of phyla (Jones 2008, Bevilacqua et al. 2012). For aquatic insects, genus- and even family-level data are often sufficient to differentiate between stream communities, but species-level data may perform marginally better (Wright 1995, Bowman and Bailey 1997, Hewlett 2000, Waite et al. 2004). When comparing molecular to morphological identifica-

E-mail addresses: [3]Present address: Département des sciences biologiques, Université du Québec à Montréal, 141 avenue du Président-Kennedy, Montréal, Québec, Canada H2X 1Y4; [4]gilliankmartin1@gmail.com; [5]sadamowi@uoguelph.ca; [6]cottenie@uoguelph.ca

DOI: 10.1086/686260. Received 15 May 2015; Accepted 12 October 2015; Published online 14 March 2016.
Freshwater Science. 2016. 35(2):701–711.

tions, Sweeney et al. (2011) found that finer taxonomic resolution, perhaps resulting from an increased ability to detect species, dramatically improved their ability to detect differences between 2 sites that were up- vs downstream of a stream segment experiencing significant nutrient input (sources included a golf course and various agricultural uses). Other researchers have found that as taxon richness within sites increased, similarity matrices became less consistent in their description of community variation as taxa were pooled into higher categories (Bowman and Bailey 1997). Here, we incorporate comprehensive species-level information, as provided through DNA barcoding, into our examination of taxonomic resolution to mitigate some of the difficulties associated with morphological identification (Jackson et al. 2014).

A relatively unexplored topic is how both spatial and environmental variation affect the role of taxonomic resolution on analysis of community composition. Many researchers have considered environmental data by grouping similar sites together (e.g., Hewlett 2000) or by comparing affected to less-affected sites (e.g., Waite et al. 2004, Sweeney et al. 2011). However, incorporation of spatial factors, such as distance between sites, has been limited. Increasing taxonomic resolution should allow for the identification of spatially localized distribution patterns that result from dispersal limitation, which is potentially independent from environmental heterogeneity. Heino (2013) suggested testing the effect of taxonomic resolution on analysis of metacommunity dynamics, but to our knowledge, no investigators have conducted such an analysis. Incorporating DNA barcoding data to test the effect of taxonomic resolution on community dynamics has also been limited (e.g., Sweeney et al. 2011, Gill et al. 2014, Jackson et al. 2014, Stein et al. 2014). These investigators focused primarily on the effect of improving species-level identification through DNA barcoding and effects of improved identification on biomonitoring. They did not address the effect of spatial structure on community assembly. DNA barcoding is a feasible method to address the 'taxonomic impediment' that faces large-scale aquatic invertebrate ecological studies (Pfenninger et al. 2007). Whether the level of taxonomic resolution used in metacommunity analyses affects the characterization of the mechanism behind community assembly, i.e., niche vs neutral dynamics, has remained an open area for investigation.

We hypothesize that coarse taxonomic resolution will group together species with potentially very different environmental preferences (e.g., Gordon and Wallace 1975, Wallace 1975), masking the relationship between community composition and environmental variables. Consequently, as species are grouped into higher taxonomic ranks the strength of the relationship between community composition and environmental variables should decline, negatively affecting the amount of variation explained by the metacommunity model. Alternatively, closely related species typically may differ little in their environmental preferences because they share similar traits (Buchwalter et al. 2008), and so grouping them will not affect, or perhaps will increase, detection of relationships between communities and environmental variables. In addition, if the species most often missed by morphological taxonomy are rare species, Siqueira et al. (2011) predict that coarser levels of taxonomic resolution will not affect metacommunity models because rare species have environmental responses similar to common species.

METHODS

Study system

Benthic invertebrate communities in 21 wadeable streams situated in Algonquin Provincial Park, Ontario, Canada, were each sampled once between 21 June and 13 July 2011 (Fig. 1). We define Algonquin Provincial Park to represent 1 metacommunity. Characteristic of the Canadian Shield region, drainage patterns are deranged because numerous small lakes disrupt the connections between streams (Christopherson and Byrne 2006). Five insect orders (Coleoptera, Diptera, Ephemeroptera, Plecoptera, and Trichoptera) and 1 subclass of annelids (Oligochaeta) were assessed separately because several of these taxa are used as indicators for biomonitoring studies (Lenat and Resh 2001), and they had high abundance and diversity within the samples.

Field sampling

Standard biomonitoring protocols based on provincial standards were used (Jones et al. 2007). At each stream, a reach encompassing 3 transects (2 riffles, 1 pool) was selected. Working from down- to upstream, we collected two 2-min kick-net samples (1-m^2, 500-μm-mesh net) at each transect. To ensure equal sampling of all microhabitats along a transect, a nearshore and a more central location were sampled. In the field, 2–4 individuals spent ~3–4 h picking specimens at each site. Kick-net samples were subsampled into 250-mL portions, and each subsample was disturbed and then double-checked by the lead researcher (GKM) to ensure all live specimens >1 mm were retained to reach a minimum of 100 specimens/kick sample. Specimens were stored in 95% ethanol and transferred to a −20°C freezer upon return from the field.

Twenty-three environmental variables were measured at each site. Twenty exhibited variability and were included in the final analysis. At the middle transect (i.e., pool) of each site, in the middle of the water column, the following variables were measured: pH (HI98129; Hanna Instruments, Woonsocket, Rhode Island), temperature (°C) (Hanna), dissolved O_2 (mg/L) (YSI 6600 V2; Yellow Springs Instruments, Yellow Springs, Ohio), conductivity (μS/L) (YSI), canopy cover, and global positioning system (GPS) coordinates. For all sites, canopy cover was approximated by eye by a

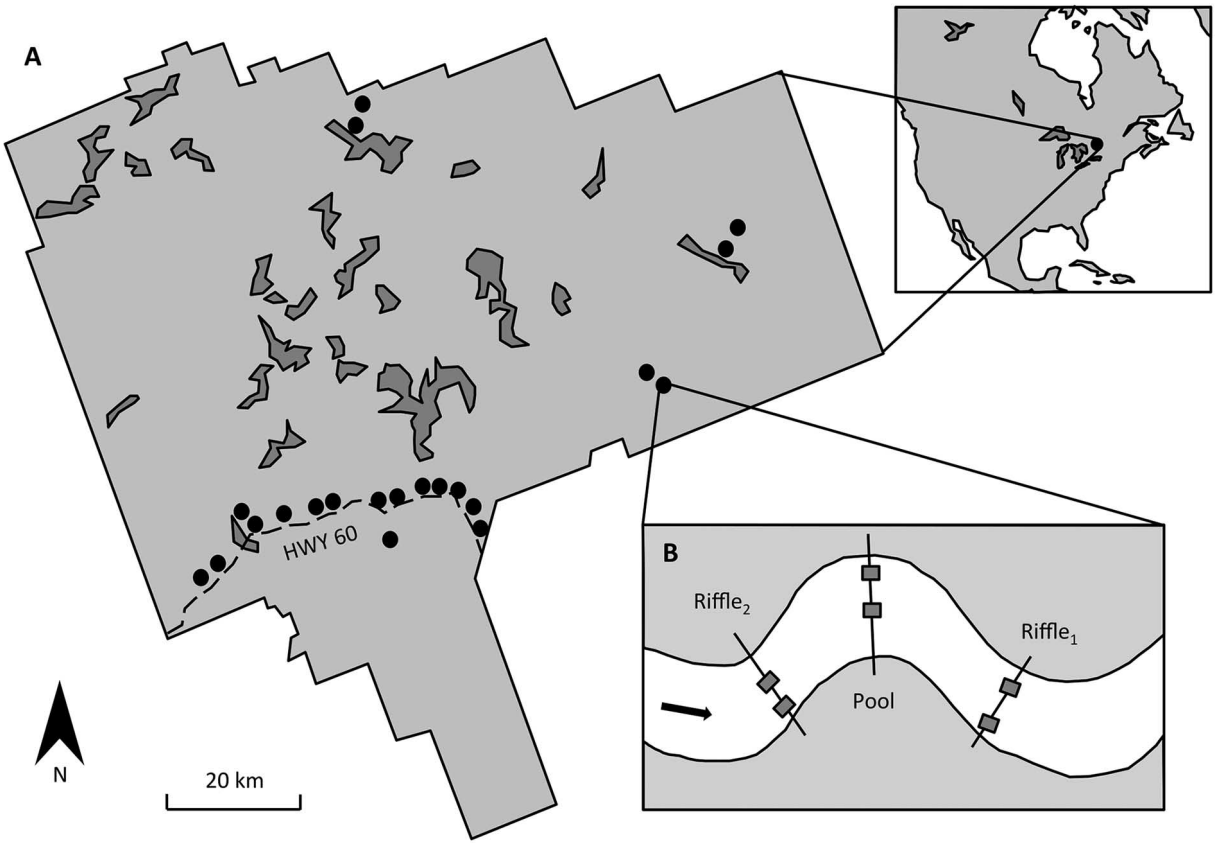

Figure 1. A.—Approximate locations of sampled stream reaches, marked by black dots, in 21 freshwater streams in Algonquin Provincial Park, Ontario, Canada (lat 45.8°N, long 78.7°W). B.—Illustration of sampling procedure at each stream reach. Gray boxes in stream represent a 1-m^2 kick-net sample along a transect. Arrow represents direction of stream flow. Transects were sampled from riffle 1 to riffle 2, moving upstream.

single investigator (GKM) and categorized as: 0–24, 25–49, 50–74, and 75–100%.

Habitat was characterized more specifically at each of the 3 transects. Dominant and 2nd-dominant substrates were measured on the basis of a modified Wentworth scale, and 7 categories were assigned: clay (hard pan), silt (gritty, <0.06 mm), sand (grainy, 0.06–2 mm), gravel (2–65 mm), cobble (65–250 mm), boulder (>250 mm), and bedrock. Woody debris, detritus, emergent macrophytes, rooted floating macrophytes, submergent macrophytes, free-floating macrophytes, floating algae, filamentous algae, attached algae, and slime or crust algae were recorded as abundant, present, or absent, again with scoring done by 1 researcher (GKM). Maximum depth (m), wetted width (m), bankfull width (m), maximum hydraulic head (mm), and sampling date also were recorded for each transect.

Specimen selection for molecular analysis

In the laboratory, all preserved specimens were counted and sorted to family, based on keys compiled by Merritt et al. (2008). To eliminate size bias, R (version 2.15; R Foundation for Statistical Computing, Vienna, Austria) was used to select 20 specimens randomly, where possible, for DNA barcoding from each family from each target order from each stream. However, to ensure sequencing success with standardized protocols, specimens had to be $\geq\sim$1 mm in size. Nearly all specimens were >1 mm in size. Specimens were drawn randomly from the total pool of specimens across all 3 transects because the unit of analysis was the stream reach. Because of the difficulties associated with identifying Oligochaeta orders, Oligochaeta specimens were not sorted below subclass. Thus, 20 specimens per subclass per stream were selected for DNA barcoding. Preliminary site-based estimates (*EstimateS*, Colwell 2004; Appendix S1) indicated that >70% of the community expected to be present in these streams was captured for all groups except Diptera. We ran Chao 1 and Chao 2 analysis (Chao et al. 2009) on our initial data set based on barcoding 20 individuals/family in each stream (Table S1). The analysis indicated that adding 20 specimens/stream for the most diverse Diptera family, Chironomidae, would capture 95% of the Chironomidae community. Therefore, an additional 20 Chironomidae specimens/stream were selected randomly. Where <20 specimens in a given family were present from a given stream, all were selected for

barcoding. We photographed all selected specimens and removed a 1–2-mm length of leg or abdominal tissue from the insect specimens for molecular analysis. For specimens ~1–2 mm long, the whole specimen was used. Fragments of Oligochaeta could not be associated with particular specimens because of specimen damage/breakage and lack of morphological difference between body segments; therefore, 20 fragments/site were selected, where possible, for DNA barcoding.

Specimen identification based on DNA barcoding

We followed standard protocols developed by the Canadian Center for DNA Barcoding, University of Guelph (described in detail in Appendix S1) to sequence specimens for the animal barcode region (Hebert et al. 2004) of the cytochrome *c* oxidase subunit I (COI) gene. To test the effect of taxonomic resolution on metric performance and explanatory power of metacommunity models, the specimens were grouped according to family, genus, and a species-level proxy based on the barcode index number (BIN; Ratnasingham and Hebert 2013), available through Barcode of Life Data Systems (BOLD; Ratnasingham and Hebert 2007). We used BINs to avoid difficulties with morphological specimen identification to species and genus levels. We identified specimens to family morphologically, whereas we made genus-level determinations based on taxonomic information associated with each BIN web page on BOLD. We verified these genus assignments morphologically (Merritt et al. 2008). The BIN algorithm assigns specimens to molecular operational taxonomic units (MOTUs), which are similar to species units, based on sequence data alone. The algorithm uses a seed threshold of 2.2% divergence to separate intraspecific from interspecific groupings but also explores patterns of genetic continuity–discontinuity to refine the groupings (Ratnasingham and Hebert 2013). Other investigators of freshwater insects have found that patterns of barcode sequence variability reliably discriminate species entities recognized through classical study, including adults (Rivera and Currie 2009, Zhou et al. 2009, 2010, 2011, Ekrem et al. 2010, Pauls et al. 2010, Sweeney et al. 2011, Curry et al. 2012, Ruiter et al. 2013, Jackson et al. 2014). Moreover, at the individual specimen level, DNA barcoding can be significantly superior to morphological approaches for achieving species-level identifications in stream invertebrates because of the high proportion of juvenile or damaged specimens and morphologically similar species (Sweeney et al. 2011, Jackson et al. 2014). Last, because of incomplete taxonomic works and barcode reference libraries for some taxa included here, particularly the Oligochaeta, a molecular approach to the delineation of species units was necessary to enable the broad taxonomic scope of our study.

MOTU assignments (BINs) were possible for 97% (2819 of 2905) of the specimens, with <3% exhibiting sequence failure or poor-quality sequences. Genus-level identifications were reached for 68.7% of specimens, and family-level identifications were reached for 95.8% (see Table S2 for a breakdown by order). Specimens without sequence data were identified morphologically to genus level when possible (Merritt et al. 2008). Specimens without a BIN or genus assignment were removed from all analyses. Removal of the small proportion of specimens that failed to yield a sequence and corresponding BIN assignment was not expected to affect the following analyses because those specimens were morphologically similar to successfully sequenced specimens within the same stream; i.e., the specimens removed from analysis apparently did not represent singleton species. Moreover, in many cases only 1 BIN was found per genus at a site, suggesting that unaccounted cryptic variation generally would not be an issue when specimens could not be identified. All taxonomic information and specimen photographs are available on BOLD (www.boldsystems.org), and project names are listed in Appendix S1.

Species, environment, and spatial matrices

Taxon matrix Taxon abundance matrices were created for each order at a BIN, genus, and family level (Table S3). Total abundances for each taxon were based on estimates from DNA barcoding (see Appendix S1 for a detailed explanation of abundance calculations). Taxon data sets were transformed with the Hellinger transformation (Rao 1995) before ordination analysis (Legendre and Gallagher 2001). Presence–absence data were used for the Oligochaeta because fragmentation made it impossible to distinguish individual specimens. In addition, corresponding taxon matrices were not created at the genus level for Diptera and at the genus and family level for Oligochaeta because of poor identification success for those levels.

Environment matrix Free-floating macrophytes and floating algae were removed from the analysis because these variables showed no variation across sites. Mean values were used whenever variables were recorded at multiple transects at a site. Seven variables (temperature, dissolved O_2, conductivity, maximum depth, wetted width, bankfull width, and hydraulic head) were $\log_{10}(x)$-transformed after averaging. Sampling date was included to account for the possible effect of adult emergence. Principal component analysis (PCA) was used to reduce the number of explanatory factors, and the first 4 component axes, based on an eigenvalue cutoff of 1, were used to create a site × environment matrix. Axes 1 through 4 explained 66% of the total variation in the environmental matrix (27.8, 16.8, 11.5, and 9.6%, respectively).

Space matrix GPS coordinates were converted to 2-dimensional Cartesian coordinates and transformed through principal coordinates of neighbor matrices analysis to de-

scribe variable degrees of possible spatial structure in the data set and the effect of spatial structure on other measured variables (Legendre and Legendre 2012). We used the 6 positive eigenvalues to create the site × space matrix.

Diversity index analysis

Three diversity values were calculated based on taxon richness, the Shannon index, and the Simpson index (Hill 1973) for each site at each level of taxonomic resolution for all 5 insect orders. Oligochaeta were not included because only 1 level of taxonomic resolution was possible. These 3 indices express diversity as an effective number of taxa present but with differing emphasis on diversity and evenness. Kendall rank correlation analysis was used to compare site rankings between BINs and genus, genus and family, and BINs and family. This analysis was conducted to determine whether site rankings were affected by a loss of taxonomic resolution. An increasing rank correlation coefficient (Kendall's τ) indicates increased agreement in the order of site rankings between each comparison of taxonomic resolution.

Metacommunity analysis

Redundancy analysis, coupled with forward selection to test the significance and relative explanatory strength of the various environmental and spatial variables, was used to parse out the relative importance of environment and space on community composition (Siqueira et al. 2011). Results obtained when including both adult and larval Coleoptera were similar to those obtained upon including only larvae. Therefore, only the Coleoptera larval results are presented. All analyses were conducted in R, with the packages *ade4* (Dray and Dufour 2007), *vegan* (Oksanen et al. 2013), and *packfor* (Dray et al. 2011).

To test whether the level of taxonomic resolution affected metacommunity dynamics, a multiple linear regression was used to assess the relationship between % variation in community composition explained by environment and level of taxonomic resolution. The same taxonomic rank could have different meanings across the higher taxa in terms of biological similarity among lower taxa, so taxonomic resolution was quantified based upon the average % sequence divergence within each level. Genetic distance summary tools available through BOLD (www.boldsystem.org; Ratnasingham and Hebert 2007) were used to estimate average divergence values among individuals within BINs, among BINs within genera, and among genera within families to facilitate comparisons across groups. The distance model Kimura-2 Parameter was used with pairwise deletion of missing data (Kimura 1980).

Our results were compared to results from a study by Heino and Mykrä (2008). Their community characterization was based on morphological identification to species for 90% of Ephemeroptera, 95% of Plecoptera, 82% of Trichoptera, and 74% of Chironomidae specimens, and the remaining specimens were identified to species group or genus. For plotting purposes, their data points are considered closest to our genus-level results because work suggests that even expert identifiers miss ~40% of the species of freshwater invertebrates when using morphological methods (Sweeney et al. 2011). We used the average % divergence within genera value obtained in our study to assign reasonable divergence values for their data. However, their data points may fall somewhere between our species and genus points because of mixed resolution of their identifications.

RESULTS

In total, 37 families, 65 genera, and 414 BINs were found among the 2905 individuals selected for sequencing from the 21 streams sampled for all 5 orders and 1 subclass (Table 1).

Diversity rankings of streams change with taxonomic resolution

Site rankings based on richness, Shannon, and Simpson indices changed significantly for all sites when BIN data were compared to genus data, when genus data were compared with family data, and when BIN data were compared with family data for Coleoptera, Ephemeroptera, Plecoptera, and Trichoptera (Table 2). For Diptera, site rankings based on richness changed significantly when BIN data were compared with family data but did not change significantly when based on Shannon and Simpson indices

Table 1. Number of barcode index numbers (BINs), genera, and families found for 5 orders and 1 subclass of aquatic invertebrates from 21 freshwater streams in Algonquin Provincial Park. Dashes indicate that either genus- or family-level identifications could not be reached for most of the indicated taxon. As a result, subsequent analyses were not conducted at those levels.

Resolution level	Trichoptera	Coleoptera	Diptera	Ephemeroptera	Plecoptera	Oligochaeta
BIN	47	13	263	44	11	36
Genus	26	11	–	22	6	–
Family	11	4	11	7	4	–

Table 2. Comparison of richness, Shannon index, and Simpson index values across all 3 levels of taxonomic resolution in all insect orders (Kendall's τ, p in parentheses). Only barcode index numbers (BINs) identifications were available for Oligochaeta. Genus-level identifications were unavailable for Diptera.

Metric	Comparison	Order				
		Coleoptera	Diptera	Ephemeroptera	Plecoptera	Trichoptera
Richness	BIN–genus	0.980 (<0.001)	–	0.904 (<0.001)	0.932 (<0.001)	0.819 (<0.001)
	Genus–family	0.734 (<0.001)	–	0.899 (<0.001)	0.910 (<0.001)	0.793 (<0.001)
	BIN–family	0.704 (<0.001)	0.470 (0.005)	0.838 (<0.001)	0.868 (<0.001)	0.675 (<0.001)
Shannon index	BIN–genus	0.905 (<0.001)	–	0.776 (<0.001)	0.803 (<0.001)	0.644 (<0.001)
	Genus–family	0.693 (<0.001)	–	0.784 (<0.001)	0.836 (<0.001)	0.633 (<0.001)
	BIN–family	0.623 (<0.001)	0.148 (0.349)	0.639 (<0.001)	0.799 (<0.001)	0.338 (0.036)
Simpson index	BIN–genus	0.897 (<0.001)	–	0.827 (<0.001)	0.886 (<0.001)	0.635 (<0.001)
	Genus–family	0.702 (<0.001)	–	0.804 (<0.001)	0.857 (<0.001)	0.560 (<0.001)
	BIN–family	0.672 (<0.001)	−0.081 (0.608)	0.691 (<0.001)	0.834 (<0.001)	0.333 (0.038)

(Table 2). Diversity index values for sites varied based on the index used (Fig. 2).

Explanatory power of metacommunity models increases with coarse taxonomic resolution

Environmental variables played a greater role than spatial variables in explaining community composition, and the amount of variation explained varied among the levels of taxonomic resolution (Appendix S2). As taxonomic resolution decreased, the amount of variation explained by the environment increased by up to a difference of 26.5% ($F_{1,4}$ = 10.4, p = 0.03). Environmental variables explained 7.8 to 13.5% of the variation in community composition at the BIN level, 12.9 to 20.1% at the genus level, and 10.8 to 37% at the family level. Environmental variables did not significantly affect Coleoptera family-level community assemblages (p = 0.102). We had previously examined the effects of rarer species on our results by repeating our analyses with only those species that had ≥5% of the total abundance. Results (not shown) were similar to our findings reported here.

Our environment and metacommunity results and those of Heino and Mykrä (2008) are plotted in Fig. 3, where we also graphically compare our results. Their data points are similar to ours. Spatial variables were related to the assemblage structure for Diptera and Oligochaeta communities at only a BIN level, explaining 3.08 (p = 0.043) and 7.43% (p = 0.044) of the variability in community composition, respectively. Spatial variables were marginally significant in structuring BIN-level assemblages for Plecoptera, explaining 15.2% of the variability in community composition (p = 0.054; Fig. S1). For the other 3 orders, Coleoptera, Ephemeroptera, and Trichoptera, spatial variables were not significantly related to the assemblage structure at any taxonomic resolution.

DISCUSSION

Effect of taxonomic resolution on relative diversity rankings of streams

Our results demonstrate a significant effect of taxonomic resolution on community characterization when using relative rankings of diversity indices. Aggregation to genus level resulted in a change in site rankings in all orders; aggregation from genus to family level resulted in a more substantial loss of information compared with the BIN to genus grouping. Together, these results suggest that genera are proportionately more different from one another in their

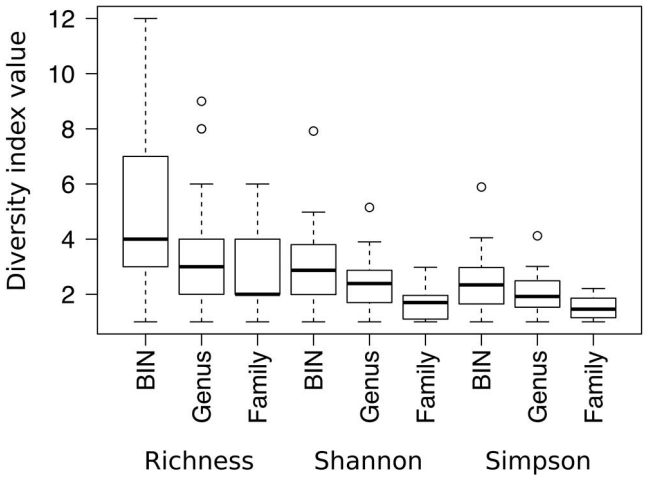

Figure 2. Variation in 3 diversity metrics based on barcode index number (BIN; species surrogates), genus, and family identifications for Trichoptera collected at 21 streams in Algonquin Provincial Park, Canada. Richness, Shannon index, and Simpson index for each site were calculated based on absolute abundance data. Results for other orders showed the same pattern. Lines in boxes are medians, box ends are quartiles, whiskers are 10th and 90th percentiles, and open circles are outliers.

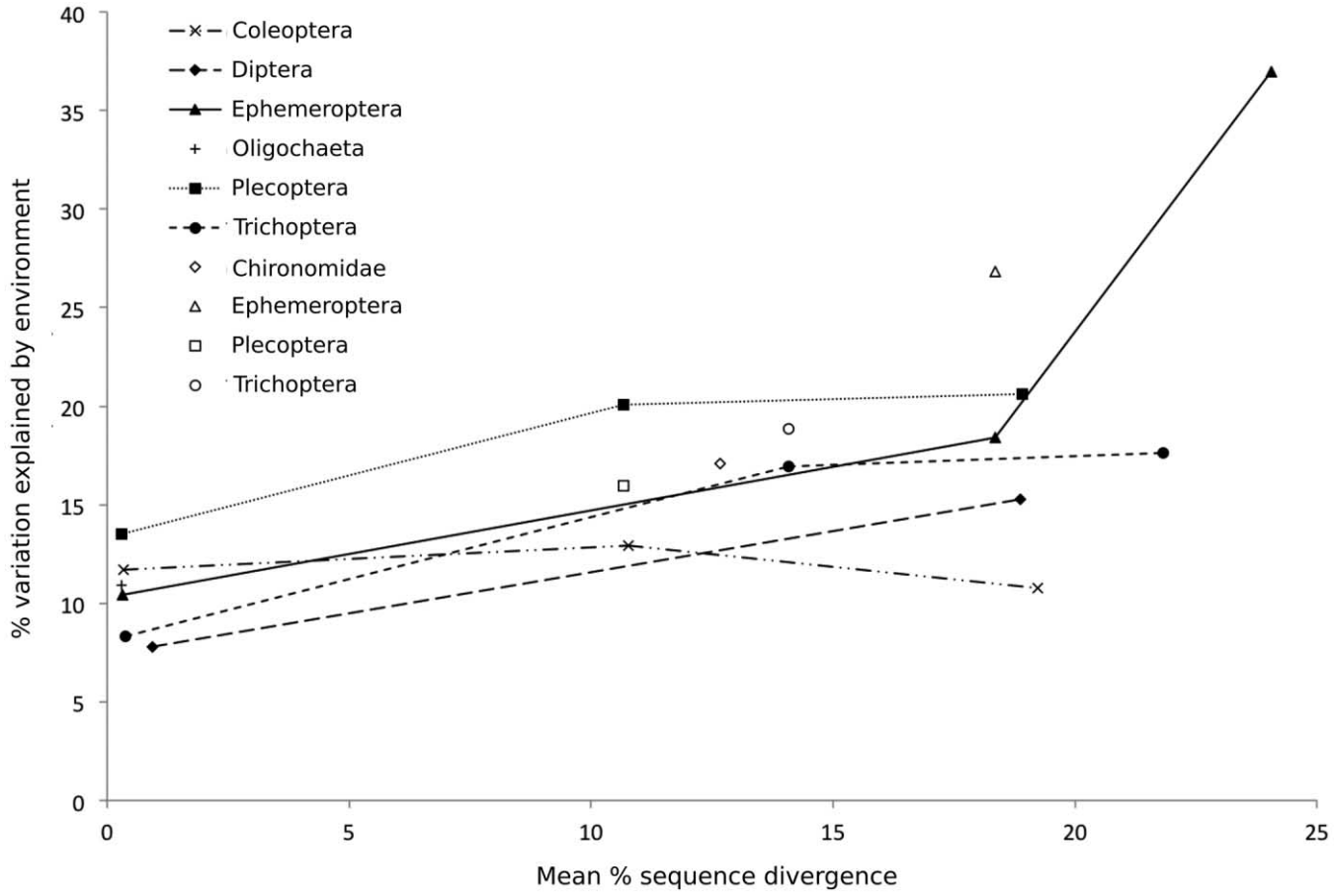

Figure 3. The variation in community structure (n = 21 streams) of 5 orders and 1 subclass of aquatic invertebrates explained by environment. Separate metacommunity analyses were conducted for each level of taxonomic resolution: barcode index number (BIN; species surrogates), genus, and family. We used average % divergence within each taxonomic level for each taxon as a quantification of taxonomic resolution. The 4 open symbols represent findings from a similar 3-y study conducted by Heino and Mykrä (2008).

responses to environmental variability than are species within genera. These results also support previous findings that diversity indices respond inconsistently to changes in taxonomic resolution (Bowman and Bailey 1997). However, our study examined only how individual site characterization changed when each stream was treated as its own stream type. These results could change if more streams were sampled and if distinct stream types, e.g., with a wider range of physical characteristics or distinct pristine vs degraded designations, were sampled.

Implications of our findings for metacommunity theory

The increase in the explanatory power of environmental variables with decreasing taxonomic resolution could be caused by several mechanisms. As alternatively hypothesized in our introduction, closely related species may share more similar traits (Buchwalter et al. 2008). Therefore, grouping species into genera or families could result in a stronger relationship between community composition and environment. This notion that neutral (ecologically largely interchangeable) species within a genus decrease the detectability of the environmental signal has received limited attention to date. Some indirect evidence for this notion exists based on similarities in niche requirements for neutral species among aquatic insects. The high species diversity found within aquatic insect orders suggests that functional redundancy is possible (Covich et al. 2004). However, studies on functional redundancy may be misleading as traits used to assign function may function differently despite their similar appearance (Vaughn 2010). Múrria et al. (2012) found that closely related species in the genus *Hydropsyche* filled similar and overlapping ecological niches. Thompson and Townsend (2006) found that predatory Trichoptera species showed considerable dietary overlap and were closely related. In contrast, in the terrestrial realm, work on herbivorous Lepidoptera larvae by Hebert et al. (2004) and parasitoid wasps by Smith

et al. (2008) suggested that many closely related species have very specific host relationships, particularly in these 2 studies, which showed intensive biotic interactions among species. Our results suggest that biologically cryptic species could be present in the groups observed in our study that are not affected by the explanatory variables measured because of either neutrality (Scheffer and van Nes 2006) or environmental factors that affect the terrestrial adults only. As barcode reference libraries improve—increasing the ability to assign Linnaean species names to specimens—and as work on identifying species-specific functional traits increases, incorporation of functional trait data will help identify the mechanisms behind the patterns observed.

Another important point to consider in the context of niche breadth is environmental quality. Active logging occurs in Algonquin Provincial Park, but the streams sampled are relatively free from anthropogenic effects compared with those investigated in many biomonitoring studies. Consequently, different results may be observed in more degraded communities where physiological differences (such as pollution tolerance thresholds) might be revealed, even among close relatives (e.g., Resh and Unzicker 1975) and potentially could result in a partial reversal of the trend detected here. Determining whether the patterns observed in our study occur in a variety of habitats under varying amounts of anthropogenic stress is an important next step.

Other researchers have attributed neutral-like dynamics in metapopulations to the formation of source–sink relationships with the environment (Bell 2001). In high-disturbance habitats, such as streams, frequent disturbance could further mask relationships with the environment and result in the appearance of neutrality (Lake 2000, Heino and Mykrä 2008). Stream community composition is highly variable. Over a 17-y period, Lenat and Resh (2001) found that only 5 to 10% of species were present during all years, and 30% appeared only once in a single stream. Last, unless an investigator explicitly incorporates dispersal ability into the analysis, species sorting with efficient dispersal, species sorting with high dispersal dynamics, or species sorting with high dispersal and neutral dynamics can be challenging to differentiate (Logue et al. 2011).

The strength of the environment–community relationship differed significantly among orders. The weakest relationship, observed for the Coleoptera, may be attributable to a near 1:1 ratio of BINs to genera and a very small number of families. As a result, the likelihood of detecting any relationship between environment and community structure is significantly smaller than if diversity were very high. However, the most diverse group, the Diptera, had only a moderately strong relationship, whereas the Ephemeroptera, with moderate richness, had the strongest relationship. This result suggests that richness alone does not affect the strength of the relationship. However, molecular evolutionary rates could differ among taxa, thereby affecting the observed relationship between the explanatory power of the metacommunity model and taxonomic resolution as quantified here. Further work should be done to incorporate ecological traits into a similar analysis, which would improve our understanding of the different relationships between environment and community structure observed in this study. This analysis could be accomplished by applying similar methods as used by Pandit et al. (2009) to compare the effects of taxonomic resolution between different functional groups or by expanding the environmental matrices to include ecological trait data, coupled with forward selection, to pinpoint the strength of contribution of these traits. Our findings point to the importance of incorporating ecological differences among taxa because these traits also can significantly affect the patterns observed.

Spatial structuring was significant for only 3 of the 5 orders and 1 subclass in our study. The relationship between environment and community composition was most commonly detected at coarser levels of taxonomic resolution (genus or family), whereas spatial relationships were detected at only the BIN level. This pattern may have occurred because genera and families typically have wider geographic distributions than BINs (Heino and Soininen 2007). However, our sites displayed some spatial clustering. The effect of space (dispersal limitation) probably would be higher if the spatial extent of sampling were greater and sites were more evenly spatially distributed. However, Brown and Swan (2010) found that dispersal is not a significant driver of aquatic invertebrate metacommunities in headwater streams. Assessing the effect of smaller-scale spatial variation also may yield different patterns because species might respond to more fine-grained resolution of the measured environmental variables than was studied here.

A number of previous researchers have investigated taxonomic composition at multiple levels of taxonomic resolution (Jones 2008), but few have investigated metacommunity patterns; i.e., examined the strength of space and environment in explaining composition with such a wide taxonomic breadth. To our knowledge, only Verleyen et al. (2009) examined the effect of taxonomic resolution on the ability to detect metacommunity patterns. Verleyen et al. (2009) used an analysis method similar to that used here to compare the explanatory power of environmental and spatial variables based on morphospecies and genus-level data for 15 regional diatom data sets, each containing data from ≥20 lakes. They found that taxonomic resolution did not significantly affect the results at a regional scale. However, the scale of their study was very different from the scale of our study (maximum distance between sites within a region was >2000 km in some cases). The difference in the effect of taxonomic resolution may have been a consequence of the facts that diatoms face greater

environmental filtering and are more efficient dispersers than are aquatic invertebrates (Astorga et al. 2012).

Sample size and effort can affect estimates of community composition and the ability to differentiate between sites (Cao et al. 1998). With the development of lower-cost sequencing methods, such as next-generation sequencing, conducting a similar experiment but identifying a greater proportion of the specimens collected will be important to determine whether the same pattern is observed. The number of individuals sequenced per site per family was relatively small in our study, but the most common species (≥5% abundance) should have been captured. Many community ecology methods include a recommendation to remove rare species from analyses (Cao et al. 1998). Our findings are applicable to rapid bioassessment methods that may not capture rare species because of smaller sample sizes, which highlights the importance of careful consideration of the level of taxonomic resolution when using similar methods.

Conclusion

Detection of a strong relationship between metacommunity structure and taxonomic resolution for multiple orders of aquatic invertebrates has implications for metacommunity research and for other research based on metacommunity theory. Our results are consistent with the use of rapid bioassessment protocols that currently focus on coarser levels of taxonomic resolution, but further studies similar to ours that explicitly incorporate stressor gradients are needed to support this practice. However, as costs of molecular analysis continue to fall in comparison with traditional methods of specimen identification, obtaining data at multiple levels of taxonomic resolution will undoubtedly contribute novel insights into community structuring and will enable more thorough investigation of relative niche breadth across taxa. The similarity between our results and those of Heino and Mykrä (2008) suggests that the patterns detected in our study are not unique to Algonquin Park. However, further study should be conducted to see if the pattern detected in our study holds true empirically for other systems. Incorporation of biological traits, such as dispersal ability and life-history traits, would improve our understanding of the relationship observed. For example, dispersal ability would inform whether one or multiple metacommunities are included within the sampling region, and life-history traits would help inform whether biological interactions or neutral dynamics play a role.

ACKNOWLEDGEMENTS

We thank the staff of the Canadian Centre for DNA Barcoding at the Biodiversity Institute of Ontario, University of Guelph, for molecular analysis. This work was supported by the Ontario Ministry of Research and Innovation through the Global Leadership Round in Genomics and Life Sciences program; the Discovery Grants program of the Natural Sciences and Engineering Research Council of Canada (SJA and KC); a grant from the Government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project led by Paul Hebert; and a Queen Elizabeth II Graduate Scholarship from the Government of Ontario (GKM). We thank 4 anonymous referees for their insightful comments on the manuscript.

LITERATURE CITED

Astorga, A., J. Oksanen, M. Luoto, J. Soininen, R. Virtanen, and T. Muotka. 2012. Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? Global Ecology and Biogeography 21:365–375.

Bell, G. 2001. Neutral macroecology. Science 293:2413–2418.

Bevilacqua, S., A. Terlizzi, J. Claudet, S. Fraschetti, and F. Boero. 2012. Taxonomic relatedness does not matter for species surrogacy in the assessment of community responses to environmental drivers. Journal of Applied Ecology 49:357–366.

Bonada, N., N. Prat, V. H. Resh, and B. Statzner. 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. Annual Review of Entomology 51:495–523.

Bowman, M. F., and R. C. Bailey. 1997. Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? Canadian Journal of Fisheries and Aquatic Sciences 54:1802–1807.

Brown, B. L., and C. M. Swan. 2010. Dendritic network structure constrains metacommunity properties in riverine ecosystems. Journal of Animal Ecology 79:571–580.

Buchwalter, D. B., D. J. Cain, C. A. Martin, L. Xie, S. N. Luoma, and T. Garland. 2008. Aquatic insect ecophysiological traits reveal phylogenetically based differences in dissolved cadmium susceptibility. Proceedings of the National Academy of Sciences of the United States of America 105:8321–8326.

Cao, Y., D. D. Williams, and N. E. Williams. 1998. How important are rare species in aquatic community ecology and bioassessment? Limnology and Oceanography 43:1403–1409.

Chao, A., R. K. Colwell, C. W. Lin, and N. J. Gotelli. 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology 90:1125–1133.

Christopherson, R. W., and M. Byrne. 2006. Geosystems: an introduction to physical geography. Canadian edition. Pearson–Prentice–Hall, Toronto, Canada.

Colwell, R. K. 2004. EstimateS: statistical estimation of species richness and shared species from samples. (Available from: http://viceroy.eeb.uconn.edu/estimates/)

Covich, A. P., M. C. Austen, F. Bärlocher, E. Chauvet, B. J. Cardinale, C. L. Biles, P. Inchausti, O. Dangles, M. Solan, M. O. Gessner, B. Statzner, and B. Moss. 2004. The role of biodiversity in the functioning of freshwater and marine benthic ecosystems. BioScience 54:767–775.

Curry, C. J., X. Zhou, and D. J. Baird. 2012. Congruence of biodiversity measures among larval dragonflies and caddisflies from three Canadian rivers. Freshwater Biology 57:628–639.

Dray, S., and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22:1–20.

Dray, S., P. Legendre, and F. G. Blanchet. 2011. packfor: forward selection with permutation. R version 0.0-8/r100. R Project for Statistical Computing, Vienna, Austria. (Available from: http://R-Forge.R-project.org/projects/sedar/)

Ekrem, T., E. Stur, and P. D. N. Hebert. 2010. Females do count: documenting Chironomidae (Diptera) species diversity using DNA barcoding. Organisms, Diversity, and Evolution 10:397–408.

Gill, B. A., R. A. Harrington, B. C. Kondratieff, K. R. Zamudio, N. L. Poff, and W. C. Funk. 2014. Morphological taxonomy, DNA barcoding, and species diversity in southern Rocky Mountain headwater streams. Freshwater Science 33:288–301.

Gordon, A. E., and J. B. Wallace. 1975. Distribution of the family Hydropsychidae (Trichoptera) in the Savannah River Basin of North Carolina, South Carolina and Georgia. Hydrobiologia 46:405–423.

Hajibabaei, M., S. Shokralla, X. Zhou, G. A. C. Singer, and D. J. Baird. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6:e17497.

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the United States of America 101:14812–14817.

Heino, J. 2013. The importance of metacommunity ecology for environmental assessment research in the freshwater realm. Biological Reviews of the Cambridge Philosophical Society 88:166–178.

Heino, J., and H. Mykrä. 2008. Control of stream insect assemblages: roles of spatial configuration and local environmental factors. Ecological Entomology 33:614–622.

Heino, J., and J. Soininen. 2007. Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms? Biological Conservation 137:78–89.

Hewlett, R. 2000. Geographic scale implications of taxonomic resolution and sample habitat for stream classification at a broad geographic scale. Journal of the North American Benthological Society 19:352–361.

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54:427–432.

Jackson, J. K., J. M. Battle, B. P. White, E. M. Pilgrim, E. D. Stein, P. E. Miller, and B. W. Sweeney. 2014. Cryptic biodiversity in streams: a comparison of macroinvertebrate communities based on morphological and DNA barcode identifications. Freshwater Science 33:312–324.

Jones, F. C. 2008. Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. Environmental Reviews 16:45–69.

Jones, F. C., K. M. Somers, B. Craig, and T. B. Reynoldson. 2007. Ontario benthos biomonitoring network: protocol manual. Analysis. Ontario Ministry of the Environment, Toronto, Canada. (Available from: http://desc.ca/programs/OBBN/)

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111–120.

Lake, P. S. 2000. Disturbance, patchiness, and diversity in streams. Journal of the North American Benthological Society 19:573–592.

Legendre, P., and E. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia (Berlin) 129:271–280.

Legendre, P., and L. Legendre. 2012. Numerical ecology: developments in environmental monitoring. 3rd edition. Elsevier Science, Amsterdam, The Netherlands.

Lenat, D. R., and V. H. Resh. 2001. Taxonomy and stream ecology: the benefits of genus- and species-level identifications. Journal of the North American Benthological Society 20:287–298.

Logue, J. B., N. Mouquet, H. Peter, and H. Hillebrand. 2011. Empirical approaches to metacommunities: a review and comparison with theory. Trends in Ecology and Evolution 26: 482–491.

Marshall, J. C., A. L. Steward, and B. D. Harch. 2006. Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: the benefits and costs of using species abundance data. Hydrobiologia 572:171–194.

Merritt, R. W., K. W. Cumming, and M. B. Berg (editors). 2008. An introduction to the aquatic insects of North America. 4th edition. Kendall/Hunt, Dubuque, Iowa.

Múrria, C., N. Bonada, M. A. Arnedo, C. Zamora-Muñoz, N. Prat, and A. P. Vogler. 2012. Phylogenetic and ecological structure of Mediterranean caddisfly communities at various spatio-temporal scales. Journal of Biogeography 39:1621–1632.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, and H. Wagner. 2013. vegan: community ecology package. R Project for Statistical Computing, Vienna, Austria. (Available from: https://cran.r-project.org)

Pandit, S. N., J. Kolasa, and K. Cottenie. 2009. Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. Ecology 90: 2253–2262.

Pauls, S. U., R. J. Blahnik, X. Zhou, C. T. Wardwell, and R. W. Holzenthal. 2010. DNA barcode data confirm new species and reveal cryptic diversity in Chilean *Smicridea (Smicridea)* (Trichoptera:Hydropsychidae). Journal of the North American Benthological Society 29:1058–1074.

Pfenninger, M., C. Nowak, C. Kley, D. Steinke, and B. Streit. 2007. Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic *Chironomus* (Diptera) species. Molecular Ecology 16:1957–1968.

Rao, C. 1995. A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. Questiió 19:23–63.

Ratnasingham, S., and P. D. N. Hebert. 2007. BOLD: the barcode of life data system. Molecular Ecology Notes 7:355–364.

Ratnasingham, S., and P. D. N. Hebert. 2013. A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS ONE 8:e66213.

Resh, V. H., and J. D. Unzicker. 1975. Water quality monitoring and aquatic organisms: the importance of species identification. Water Pollution Control Federation 47:9–19.

Rivera, J., and D. C. Currie. 2009. Identification of Nearctic black flies using DNA barcodes (Diptera: Simuliidae). Molecular Ecology Resources 9:224–236.

Ruiter, D. E., E. E. Boyle, and X. Zhou. 2013. DNA barcoding facilitates associations and diagnoses for Trichoptera larvae of the Churchill (Manitoba, Canada) area. BMC Ecology 13:5.

Scheffer, M., and E. H. van Nes. 2006. Self-organized similarity, the evolutionary emergence of groups of similar species. Proceedings of the National Academy of Sciences of the United States of America 103:6230–6235.

Siqueira, T., L. M. Bini, F. O. Roque, S. R. Marques Couceiro, S. Trivinho-Strixino, and K. Cottenie. 2011. Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. Ecography 35:183–192.

Smith, M. A., and B. L. Fisher. 2009. Invasions, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. Frontiers in Zoology 6:1–12.

Smith, M. A., J. J. Rodriguez, J. B. Whitfield, A. R. Deans, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proceedings of the National Academy of Sciences of the United States of America 105:12359–12364.

Stein, E. D., B. P. White, R. D. Mazor, J. K. Jackson, J. M. Battle, P. E. Miller, E. M. Pilgrim, and B. W. Sweeney. 2014. Does DNA barcoding improve performance of traditional stream bioassessment metrics? Freshwater Science 33:302–311.

Sweeney, B. W., J. M. Battle, T. Dapkey, and J. K. Jackson. 2011. Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? Journal of the North American Benthological Society 30:195–216.

Thompson, R., and C. Townsend. 2006. A truce with neutral theory: local deterministic factors, species traits and dispersal limitation together determine patterns of diversity in stream invertebrates. Journal of Animal Ecology 75:476–484.

Valentini, A., F. Pompanon, and P. Taberlet. 2009. DNA barcoding for ecologists. Trends in Ecology and Evolution 24: 110–117.

Vaughn, C. C. 2010. Biodiversity losses and ecosystem function in freshwaters: emerging conclusions and research directions. BioScience 60:25–35.

Verleyen, E., W. Vyverman, M. Sterken, D. A. Hodgson, A. De Wever, S. Juggins, B. Van de Vijver, V. J. Jones, P. Vanormelingen, D. Roberts, R. Flower, C. Kilroy, C. Souffreau, and K. Sabbe. 2009. The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. Oikos 118:1239–1249.

Waite, I. R., A. T. Herlihy, D. P. Larsen, N. S. Urquhart, and D. J. Klemm. 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, U.S.A. Freshwater Biology 49:474–489.

Wallace, J. B. 1975. Food partitioning in net-spinning Trichoptera larvae: *Hydropsyche venularis*, *Cheumatopsyche etrona*, and *Macronema zebratum* (Hydropsychidae). Annals of the Entomological Society of America 68:463–472.

Wright, I. 1995. The impact of sewage effluent on the macroinvertebrate community of an upland stream: the effect of different levels of taxonomic resolution and quantification. Australian Journal of Ecology 20:142–149.

Zhou, X., S. J. Adamowicz, L. M. Jacobus, R. E. DeWalt, and P. D. N. Hebert. 2009. Towards a comprehensive barcode library for arctic life—Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. Frontiers in Zoology 6:1–9.

Zhou, X., L. M. Jacobus, R. E. DeWalt, S. J. Adamowicz, and P. D. N. Hebert. 2010. Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. Journal of the North American Benthological Society 29:814–837.

Zhou, X., J. L. Robinson, C. J. Geraci, C. R. Parker, O. S. Flint, A. Etnier, D. Ruiter, R. E. DeWalt, L. M. Jacobus, and P. D. N. Hebert. 2011. Accelerated construction of a regional DNA-barcode reference library: caddisflies (Trichoptera) in the Great Smokey Mountains National Park. Journal of the North American Benthological Society 30:131–162.