

**This article has been accepted for publication in *Hydrobiologia* following peer review. The final publication is available at Springer via <http://dx.doi.org/10.1007/s10750-017-3100-z>**

**Challenges using extrapolated family-level macroinvertebrate metrics in moderately disturbed tropical streams: a case-study from Belize**

Rachael Carrie<sup>1,3</sup>

Lancaster Environment Centre, Lancaster University, Lancashire, LA1 4YQ, UK

Michael Dobson<sup>2</sup>

The Freshwater Biological Association, Ambleside, Cumbria, LA22 0LP, UK

Jos Barlow

Lancaster Environment Centre, Lancaster University, Lancashire, LA1 4YQ, UK

<sup>1</sup>Present address: Institute of Science and the Environment, University of Worcester, Worcestershire, WR2 6AJ, UK

<sup>2</sup>Present address: APEM Limited, The Technopole Centre, Edinburgh Technopole, Milton Bridge Nr Penicuik, Midlothian, EH26 0PJ, UK

## **ACKNOWLEDGMENTS**

This research was funded by the Natural Environmental Research Council-Economic and Social Research Council grant ES/F013035/1, the Rufford Small Grants Foundation (grant 11376-2), and the Freshwater Biological Association Hugh Cary Gilson Memorial Award for 2012, and made possible by the Ya'axché Conservation Trust, particularly Devina Bol, Anignazio Makin, Octavio Cal, Pastor Ayala and Abelino Zuniga who provided assistance in the field and laboratory. We also thank Nabor Moya for clarifying the fuzzy-coding technique, Patrick Keenan and Phil Haygarth for assistance with nutrient analysis, and John Murphy for comments on an earlier draft. We gratefully acknowledge the Belize Fisheries Department, Belize Forest Department and Belize Agricultural Health

---

<sup>3</sup>Corresponding author: e-mail: [r.carrie@worc.ac.uk](mailto:r.carrie@worc.ac.uk) +44 (0)1905 542707

Authority for permitting the study and export of specimens and the Animal Health and Veterinary Laboratories

Agency for permitting the import of specimens into the UK, and anonymous reviewers for their comments on earlier versions of this manuscript.

## **Abstract**

Family-level biotic metrics were originally designed to rapidly assess gross organic pollution effects, but came to be regarded as general measures of stream degradation. Improvements in water quality in developed countries have reignited debate about the limitations of family-level taxonomy to detect subtle change, and is resulting in a shift back towards generic and species-level analysis to assess smaller effects. Although the scale of pollution characterizing past condition of streams in developed countries persists in many developing regions, some areas are still considered to be only moderately disturbed. We sampled streams in Belize to investigate the ability of family-level macroinvertebrate metrics to detect change in stream catchments where less than 30% of forest had been cleared. Where disturbance did not co-vary with natural gradients of change, and in areas characterized by low intensity activities, none of the metrics tested detected significant change, despite evidence of environmental impacts. We highlight the need for further research to clarify the response of metrics to disturbance over a broader study area that allows replication for confounding sources of natural variation. We also recommend research to develop more detailed understanding of the taxonomy and ecology of Neotropical macroinvertebrates to improve the robustness of metric use.

## **Key Words**

Bio-assessment; taxonomic resolution; tropical data gaps

## **Introduction**

Biotic metrics were first developed in Europe during the early 20<sup>th</sup> Century, as a means of using invertebrate assemblages to assess stream water quality, specifically organic pollution (e.g. Kolkwitz & Marsson, 1909, Woodiwiss, 1964). These metrics were subsequently modified to reflect family-level tolerances allowing rapid stream assessment (Armitage et al., 1983, Dickens & Graham, 2002). They have since been pivotal for monitoring organic stream pollution in many of the more developed regions, including the United States, parts of Europe and Australia, where they underpin sophisticated bio-assessment programs (Wright, 2000, Poquet et al., 2009), and have come to be regarded as general measures of stream condition rather than measures of stressor specific stream degradation (Wyżga et al., 2013).

Wide-scale abatement of gross organic pollution and improvements in stream water quality in developed countries have reignited discussions centred on the limitations of family-level tolerance metrics to detect changes in stream fauna (Taylor et al., 1997, Smith et al., 1999), to the extent that Friberg et al. (2011) consider they are becoming obsolete. In particular, the ability of family-level information to distinguish variation associated with low-level effects including diffuse sediment inputs and mild eutrophication has been questioned (Armitage et al., 1983, Hilsenhoff, 1988, Jones, 2008). Although subsequent investigations have yielded mixed results (Hawkins et al., 2000, Waite et al., 2004, Metzeling et al., 2006), there has been a shift in bio-assessment research back towards generic and species-level analysis to assess and diagnose subtle and often diffuse effects (e.g. Monk et al., 2012, Buendia et al., 2013, Murphy et al., 2013).

Despite these concerns, family-level metrics still present a valuable assessment tool in many developing regions because urbanization is rapidly increasing and sanitation and sewerage systems tend to be poor (Capps et al., 2016): estimates suggest that up to 90 percent of wastewater in developing countries is discharged untreated, and environmental guidelines and water governance are inadequate or weakly enforced (WWAP, 2015). Furthermore, implications for fauna in tropical streams may be greater than in temperate systems: higher temperatures reduce the capacity of stream water to hold dissolved oxygen, potentially leaving them more vulnerable to increased oxygen demand linked to organic pollution (Jacobsen, 1998, Lewis, 2008), although this was not found to be the case in the Australian Wet Tropics (Connolly et al., 2004).

In contrast to extensive knowledge about temperate streams, there exists relatively little information about the response of tropical stream ecosystems to human activities, or the identity and environmental characteristics of the species belonging to the macroinvertebrate families traditionally used to classify temperate streams (Jacobsen et al., 2008, Alonso-EguíaLis et al., 2014). Knowledge is increasing in some areas (e.g., Buss & Salles, 2007 (Brazil), Tomanova & Tedesco, 2007 (Bolivia), Rios-Touma et al., 2014 (Peru & Ecuador), yet relatively few developing tropical countries have developed biotic metrics based on knowledge of local fauna-environmental relationships (Jacobsen et al., 2008). Frequently, evaluation has been made by extrapolating from metrics based largely on the oxygen requirements of temperate stream macroinvertebrate families (see Chang et al., 2014 and references therein), which may not reflect the requirements of macroinvertebrates in tropical streams. While some question the validity of this approach because of species and genus-specific tolerances (Resh, 2007), family-level tolerance metrics like the Biological Monitoring Working Party (BMWP) score, first developed for use in the UK and often incorporating local modifications for application elsewhere, have been used alone and in combination with additional metrics to detect large-scale impacts in the developing Americas (Jacobsen, 1998, Knee & Encalada, 2013), the Caribbean (Naranjo López et al., 2005), Asia (Boonsoong et al., 2009), and Africa (Kobingi et al., 2009).

While the adaptation of these metrics can be valuable, it assumes the environments to be assessed are suffering gross organic or large-scale impacts. However, many developing tropical areas are still minimally or only moderately disturbed. For example, in eight Latin American countries over 25 percent of the national population is concentrated in the largest city (Cohen, 2006), and according to the 2015 Global Forest Resource Assessment, over half of Central America's land area is covered by forest or other woody vegetation. In less populated areas, particularly where stream headwaters are protected, stressors affecting streams can be mild and diffuse. Given concerns about the efficacy with which family-level metrics discriminate subtle impacts in the regions for which they were intended, there is a need to assess their ability to discriminate low-level effects elsewhere.

We investigated family-level macroinvertebrate metric response to environmental change in a tropical landscape that is only moderately disturbed. We selected four catchments in southern Belize where the population density was low ( $< 7$  people/km<sup>2</sup> in 2012, Statistical Institute of Belize, 2016), and less than 30 percent of the forest had been converted to other land-uses (Cherrington et al., 2010). We hypothesized that family metrics would vary in relation to decreasing forested condition along a gradient from subsistence/non mechanized farming to residential

activities to commercial agriculture. Our study catchments were environmentally heterogeneous (Carrie et al., 2015) and since environmental and subsequent biological change can be associated with natural as well as anthropogenic gradients (Neff & Jackson, 2011), we expected to observe a response to both gradients. Ability to detect anthropogenic environmental change can be confounded by natural variation, particularly where disturbance effects are subtle and natural variability large (Hawkins et al., 2000, Metzeling et al., 2006), so we controlled for season, stream order and altitude and located our treatment categories separately within streams of different geological types (i.e., treatment group was nested within categories reflecting upstream geological influence). Due to logistical constraints imposed by natural patterns of disturbance, we acknowledge potential issues associated with collinearity between disturbance and longitudinal position of the site on the stream, and attempted to quantify this influence as a confounding parameter. A secondary objective therefore, was to consider bio-assessment in circumstances where logistical constraints limit sample design.

## **Materials and Methods**

### **Study Area**

We selected 40 sites below 130 m asl across four 3<sup>rd</sup> to 5<sup>th</sup> order stream systems that drain 2600 km<sup>2</sup> of semi-deciduous upland and coastal-plain broadleaf rainforest and pine savannas (Fig. 1). The headwaters of all but one stream were contained within forested reserves and little infrastructural or urban development was present. Sites were evenly distributed across calcareous and silicate geologies. Swidden agriculture (hereafter termed 'milpa'), residential activities, cattle pastures and citrus plantations occurred across the study area. Banana cultivation and sand and gravel mining occurred only in silicate catchments. Five sites at least 1 km apart were selected a priori within each of four treatment categories: (1) forest; (2) milpa; (3) residential activities; and (4) commercial agriculture in each stream type using local knowledge, aerial photography and reconnaissance. Access limitations prevented sampling of one of the five forested calcareous sites, meaning we sampled a total of 39 sites.

To confirm allocation of sites to treatment categories, information about human activities was recorded at the 20 transects following Peck et al. (2006). Adaptations were made to human activity categories to reflect activities present in the study area. The presence of on-going or recent non-mechanized rotational farming in the riparian zone was summed for milpa; the presence of banana, citrus, pasture and/or cattle and associated equipment

(e.g., abstraction pipes, cattle fencing) was used as evidence of commercial agriculture; evidence of residential activities including bathing and laundry and community buildings was recorded as residential use. Forested reaches had little or no evidence adjacent to the stream of recent anthropogenic disturbance. Observations were weighted according to proximity to the stream channel, and summed to produce a site score describing the different categories of human activity following Kaufmann et al. (1999). Comparison of scores confirmed treatment groups to be dissimilar to the others ( $p < 0.05$ ), except commercial agriculture and residential activities in silicate stream sections (Tukey's test  $p = 0.472$ ). Furthermore, all silicate agricultural and residential sites were influenced by in-stream sand and gravel mining. Thus silicate stream treatment categories were revised to 'forest', 'milpa' and 'multiple stressors', where multiple stressors represented a combination of commercial agriculture, residential and mining activities.

#### Macroinvertebrates

Macroinvertebrates were sampled between March and May 2012 (end dry season) following Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates (AQEM) project methods (AQEM Consortium, 2002). Sampling was limited to one season to control for the potentially confounding effects of seasonality on macroinvertebrate assemblages observed in the study streams previously (Carrie et al., 2015), and that others have linked to temporal variation in metric scores in other lowland neotropical streams (Helson and Williams, 2013). Sampling was timed to occur during low or base-flow conditions because human impacts are likely to be more pronounced and because streams are more accessible. Twenty units (30 x 30 cm) were sampled in microhabitats proportional to their cover at each site, using a kick net (30 x 25 cm frame, 500  $\mu$ m mesh) and manual sampling. Microhabitats were defined as combinations of substrate and flow types with coverage of at least five percent in a sampling reach approximately 2x the average channel width. All reaches contained riffle-pool sections. Samples were fixed in situ using 10 percent neutral buffered formalin and specimens sorted from residue in the laboratory before preservation in 70 percent ethanol. In a departure from AQEM methodology, specimens were sorted from all residue, rather than from sub-samples. Regional keys were used (Springer et al., 2010), along with resources from North and South America (Merritt et al., 2008, Mugnai et al., 2010) to enumerate whole samples to family taxonomic level. Brachyura, Ostracoda, Oligochaeta, Polychaeta and Collembola were identified to order.

We selected four structural metrics commonly used to evaluate stream condition in tropical streams: (1) family richness; (2) the number of families in orders Ephemeroptera, Plecoptera, and Trichoptera (EPT richness); (3) a version of the BMWP score (Armitage et al. 1983) adapted for Costa Rican streams (BMWP-CR) (see Online Resources 1 for scoring families); and (4) the Average Score Per Taxon (ASPT-CR) that is derived from the BMWP-CR index by dividing BMWP values by the number of contributory taxa (Armitage et al., 1983). We expected scores to decrease in response to disturbance. We also selected seven functional metrics considered sensitive to change in the environmental characteristics most variable in our study streams (sedimentation and riparian land cover – see results) (e.g. Benstead et al., 2003, Sutherland et al., 2012, Buendia et al., 2013). The proportion of the total fauna (by abundance) within three habit (clinger (CL), burrower (BU), swimmer (SW)) and four feeding (collector-gatherers (CG), collector-filterers (CF), scrapers (SC), shredders (SH)) groups was calculated following Tomanova et al. (2008), Tomanova and Usseglio-Polatera (2007), Merritt et al. (2008) and our own observations for taxa not listed (Table Online Resources 2). Following Moya et al. (2011), allocations were weighted by the strength of affinity for more than one feeding or habit trait. We expected that CG and BU would increase in relative abundance in response to disturbance and that SC, SH, CF, SW and CL would decrease (Merritt et al., 2008, Tomanova et al., 2008), although where perturbation is low a non-linear response to nutrient and light subsidies may be evident in some functional and taxonomic metrics (Odum et al., 1979, Bojsen & Jacobsen, 2003).

#### Environmental variables

Triplicate measurements of water temperature, dissolved oxygen (DO), pH, and electrical conductance (EC) were made using hand held probes (YSI 55, YSI, Yellow Springs, OH, U.S.A, Oakton PCSTestr 35, Eutech Instruments, Vernon Hills, IL, U.S.A). Analysis of nitrate ( $\text{NO}_3\text{N}$ ) and soluble reactive phosphate ( $\text{PO}_4^{3-}$ ) was made in situ in triplicate from filtered samples taken mid-stream (Whatman cellulose acetate membrane filters (0.45  $\mu\text{m}$ )) using the Hach DR890 colorimeter (Hach Company, Loveland, CO, U.S.A) and the cadmium reduction and ascorbic acid methods respectively.

Channel substrates, embeddedness, water depth and width, and riparian canopy were measured following the quantitative methods of Peck et al. (2006) along 20 transects placed evenly over a distance equal to 40x the mean stream width. Riparian forest width was estimated at the same 20 points. Stream condition was assessed using the Stream Visual Assessment Protocol (SVAP) (Natural Resources Conservation Service, 1998) adapted for Belize



(Esselman, 2001). This method qualitatively ranks features describing physical attributes (e.g. water appearance, bank stability etc.) from 1 to 10. The average of the scores assigned to each feature provides an overall site score ranging from degraded (1) to reference condition (10).

#### Covariation between environmental disturbance and natural variation

We controlled for confounding natural variation by stratifying our survey network by surficial geology and restricting sites to lowland streams of similar stream order. However, the most impacted areas in the studied silicate streams occurred only on two 5<sup>th</sup> order main stems. We were logistically unable to widen the study area, so impacted and unimpacted sites were selected on the same watercourses which limited our ability to replicate sites within treatment categories to reduce covariation between disturbance and longitudinal position on the stream (Fig. 1). To account for longitudinal position, the distance of each site from the coast was calculated from a digitized drainage map (1:50,000) using ArcGis (Version 9.2, ESRI, Redlands, California) and, following Ligeiro et al. (2013), a continuous stream size metric was calculated by multiplying the mean values for stream width and depth.

#### Statistical analyses

Uncorrelated composite variables explaining variation in the original dataset were extracted using principal components analysis (PCA) with varimax rotation on 14 environmental variables considered indicative of stream impairment (Tables 1 & 2) and two variables describing longitudinal position. The number of components to retain was determined following consideration of variance explained and scree plots. Variation among treatment groups in PCA site scores was tested using one-way ANOVA followed, when significant ( $p < 0.05$ ), by pairwise comparisons (Tukey's test) separately for calcareous and silicate streams. Welch's ANOVA and Games-Howell post-hoc tests were used where Levene's tests indicted heterogeneous variances.

To examine patterns of variation and test if assemblage composition differed among calcareous and silicate streams we used non-metric multidimensional scaling (nMDS) and permutational multivariate analysis of variance (PERMANOVA) (Anderson *et al.* 2008) on  $\log_{10}(x+1)$ -transformed data with Bray-Curtis dissimilarity as the distance measure (Bray & Curtis 1957). To test variability in biotic metrics among treatment groups we used one-way ANOVA, followed by pairwise comparisons (Tukey's test) when significant ( $p < 0.05$ ), separately for each stream type. Levene's tests for these comparisons produced a non-significant result ( $p > 0.05$ ), confirming variance

across groups was equivalent despite the small differences in sample size among some of the treatment groups. To investigate the possibility that model outputs were driven by spatial autocorrelation we visually assessed residuals of models that gave a significant result. Variation in taxonomic attributes among treatment groups was also examined using nMDS and PERMANOVA separately for each stream type, followed by pairwise comparisons when a significant effect was found. The indicator value method (INDVAL) (Duf r ne & Legendre, 1997) was used to identify taxon discriminating between treatment groups when a significant effect was found. The indicator value of a taxon varies from 0 to 100, and attains maximum value when all individuals of a taxon occur at all sites of a single treatment.

nMDS and PERMANOVA were performed using Primer-E (Clarke & Gorley, 2006) with the PERMANOVA+ (Anderson *et al.*, 2008) add-on. All other tests and plots were done using R Freeware (R Development Core Team, 2008). Two-dimensional nMDS solutions were retained because stress values were  $\leq 0.2$  (Clarke & Warwick, 2001) and *p*-values for the test statistic (pseudo *F*-value) in PERMANOVA were based on 9999 permutations. Data were checked for distribution assumptions and skewness, and percentage data were arcsine transformed prior to analysis.

## Results

### Environmental characteristics

Three principal components described 76.7 percent of variation in environmental characteristics in silicate streams. PC1 accounted for 46.1 percent (Table 1) and was loaded with variables describing covarying anthropogenic and natural gradients. It was positively correlated with the distance of the sampled site from the sea, DO, SVAP, pH and riparian width and negatively correlated with the proportion of fines, sands and embeddedness and water turbidity. PC1 site scores varied significantly among treatment groups ( $F=17.96$ ,  $p<0.001$ ), and highlighted that disturbance was confounded by the natural longitudinal gradient. Pairwise comparisons indicated that riparian forest alongside downstream sections in areas subject to multiple stressors was more degraded, that substrates were more embedded with sand and fines, and water more turbid and acidic than in upstream areas of forest ( $p<0.001$ ) and milpa land ( $p=0.008$ ). These characteristics also varied among milpa and forested sites ( $p=0.021$ ). There was no significant variation among treatment groups in the environmental characteristics described by PCs 2 and 3 (Table 1).

Five principal components described 78.7 percent of variation in environmental characteristics in calcareous streams and PC1 accounted for 27.7 percent of this (Table 2). It was positively related to SVAP,  $\text{NO}_3\text{-N}$ , riparian width and DO and negatively related to substratum embeddedness and  $\text{PO}_4^{3-}$ . In contrast to silicate streams, variables describing longitudinal position of the site were strongly loaded on PC2, rather than PC1. Thus PC1 was considered to represent a disturbance gradient less confounded by this potential source of natural variation. PC1 varied among treatment groups ( $F=7.89, p=0.002$ ). Pairwise comparisons indicated stream substrates were more embedded at sites in all treatment groups, riparian forest was more degraded, stream water  $\text{PO}_4^{3-}$  lower and DO higher than at sites in forested areas (forest:milpa  $p=0.038$ , forest:residential  $p=0.023$ , forest:agriculture  $p=0.001$ ). There was no significant variation among treatment groups in the environmental characteristics described by PCs 2 - 4 (Table 2). Mean values for environmental characteristics are provided for each treatment group (Table Online Resources 3 & 4).

#### Macroinvertebrates

The primary determinant of macroinvertebrate assemblage composition was stream type (PERMANOVA,  $F=14.47, p<0.001$ ) (Fig. 2) and, in silicate streams alone, treatment group (PERMANOVA,  $F=1.81, p=0.004$ ) (Fig. 3). In silicate streams, assemblage composition at sites influenced by multiple stressors was significantly different from upstream forested sites ( $t=1.42, p=0.013$ ) and milpa sites ( $t=1.38, p=0.017$ ) and they were characterized by Naucoridae and Thiaridae (Table 3). Compositional differences among forested and milpa sites were minimal ( $t=1.15, p=0.167$ ), although six taxa were significant indicators of forested sites, and nine of milpa sites. Three were common to both forest and milpa sites, consistent with their generally high compositional similarity. Perlidae was the most distinctive indicator of upstream forested sites but was also present, albeit in low abundances, in areas subject to multiple stressors. In contrast, Odontoceridae and Athericidae were entirely absent from the downstream sites. Athericidae also characterized milpa sites but Calamoceratidae was the most distinctive indicator and it was also absent from downstream sites subject to multiple stressors, as were Hydrobiosidae and Psychodidae. Hydroptilidae, Limoniidae and Coenagrionidae showed affinity for milpa sites, but were recorded with less frequency in downstream sections.

The only functional metric to vary significantly among treatment groups in silicate streams was the proportion of collector-gatherers ( $p=0.041$ ). CGs were less important at downstream multiple stressor sites relative

to forested ( $p=0.039$ ) and milpa sites ( $p=0.022$ ). Mean values for most structural metrics were highest at milpa sites, revealing a non-linear response (Fig. 4). Family richness, EPT family richness, BWMP-CR and ASPT-CR were all significantly higher at milpa than multiple stressor sites, while only BWMP-CR and ASPT-CR were significantly higher at forest than multiple stressor sites. Visual analysis of model residuals for BWMP-CR suggested sites may not have been wholly independent. However, clear patterns were not revealed in the spatial plots for most other model residuals (Online Resources 5).

Although classification boundaries from remote geographic locations cannot be reliably used to assess condition in the study area, structural metric scores were high enough at all sites to demonstrate no evidence of major impacts on the macroinvertebrate fauna. Nineteen of the 20 silicate sites sampled would have been classified as being of ‘excellent’ quality (BMWP-CR >120) using Costa Rican criteria, including those influenced by cattle, banana cultivation and in-stream mining (Online Resources 6).

In calcareous streams there was no evidence of consistent compositional variation or evidence of significant variation in structural or functional metrics among treatment groups. (Fig. 4). However, tolerance metrics and family richness tended to be lower at agricultural sites than others, although variation within groups reduced the significance of these differences. This is demonstrated by the family richness metric, for which the lowest and highest values (17 and 36) were recorded at agricultural sites (both cattle-dominated). Minimum EPT richness (7) and BMWP-CR (82) scores were calculated for samples taken from agricultural sites (cattle and citrus), while maximum scores were recorded at sites subject to residential activities (EPT richness: 15 and BMWP-CR: 198). The minimum ASPT-CR score (4.81) was recorded in an area of subsistence farming whilst the maximum score (5.91) was recorded in a forested reach. Sixteen of the 19 calcareous sites sampled achieved BMWP-CR scores of 120 or more, including those influenced by residential activities and cattle (Online Resources 6).

## **Discussion**

The efficacy with which family-level metrics can detect subtle environmental change is increasingly questioned by researchers in temperate streams (Hawkins et al., 2000, Lenat & Resh, 2001, Waite et al., 2004, Heino, 2014). Controlling, where possible, for environmental heterogeneity in a tropical landscape where levels of disturbance were moderately low, we found little evidence that family-level metrics responded consistently to anthropogenic

environmental change. Significant variation in metric scores was not observed among calcareous treatment groups where the disturbance gradient appeared lower than in silicate streams, and where it did not correspond closely with a gradient of longitudinal change. Although a significant response was observed in assemblage attributes measured at our most disturbed silicate stream sections, disturbance was strongly collinear with the natural upstream-downstream gradient, making it impossible to separate the two. Our results are limited to dry season data only. However, since it is probable that the influence of human activities is more discernible at this time because flows are lowest, our results highlight three key problems associated with assessing anthropogenic effects using family-level metrics in tropical landscapes that are only moderately disturbed. Below we examine these problems and consider implications associated with our findings.

#### Problems of collinearity

Natural gradients influence the suitability of land for agriculture and other uses making it problematic to attribute variation in aquatic biota to stressor gradients, because of covariation in natural and human environments (e.g. Richards et al., 1997, Yates & Bailey, 2010). Our ability to assess the efficacy with which family-level metrics could detect anthropogenic change in silicate streams was limited by covariation between patterns of disturbance in the study landscape and longitudinal position of the site on the stream. Like others (e.g. Richards et al., 1997, Kratzer et al., 2006, Yates & Bailey, 2010), we were unable to relate biological change solely to anthropogenic effects because environmental characteristics that can be important drivers of assemblage composition and that may reflect natural heterogeneity, including substratum, flow dynamics and macrophyte abundance, varied as silicate streams flowed downstream.

At downstream sites, riparian and depositional characteristics, particularly fine sediment accumulation, were significantly altered from those measured at sites upstream. Corresponding biological change was explained by the replacement of families considered intolerant of organic stress (e.g. Odontoceridae, Calamoceratidae, Athericidae) with taxa that have greater ability to recover from disturbance events (e.g. Naucoridae, Thiaridae). The compositional changes observed contributed to significant reductions in family and EPT richness and BMWP-CR and ASPT-CR scores at multiple stressor sites. This response was consistent with that shown in other tropical streams where the removal of taxa from sites subject to high-levels of anthropogenic stress resulted in a reduction of metric scores that were originally designed to respond to organic pollution (Thorne & Williams, 1997, Raburu et al.,

2009, Helson & Williams, 2013), and with studies in temperate streams which found that sedimentation significantly influenced families intolerant of organic stress (Sutherland et al., 2012). Given the nature of activities impacting downstream locations, which have been related to changes in macroinvertebrate assemblages elsewhere (Harding et al., 2000, Castillo et al., 2006, Schäfer et al., 2007), it would be surprising if anthropogenic biological alteration had not occurred. However, the extent to which biological change corresponds with natural and/or anthropogenic environmental change in the silicate streams remains unclear.

#### Importance of effect size

Our results support the notion that response of family-level information can be robust to strong environmental gradients but can yield ambiguous patterns for low-level effects (Jones, 2008). Consistent with previous conclusions (Carrie et al., 2015), compositional data indicated environmental variation associated with geology exerted stronger influence than that associated with human activities (Fig. 2). The strength of geological influence is not surprising given observations elsewhere (Neff & Jackson, 2011), especially since our data confirmed our study streams were only moderately disturbed: DO, pH and turbidity (Online Resources 3 & 4) were within the range stipulated in Columbian (Decree 1594 of 1984), Costa Rican (Decree 33903 of 2007) and Mexican (CE-CCA-001-89) legislation to safeguard aquatic life, and were similar to measurements made in minimally or moderately disturbed streams across a number of tropical regions (Thorne & Williams, 1997).

Nonetheless, significant variation in PC1 scores among all treatment categories indicated that human activity had altered stream condition. Yet, in calcareous streams there was little evidence that associated environmental change was strongly related to composition or biotic metrics. At agricultural sites, mean EPT and ASPT-CR scores were reduced, suggesting a response in at least some intolerant families, but taxa that were frequently sampled at forested and milpa sites (e.g. Athericidae, Odontoceridae, Ecnomidae, Perlidae) were not consistently removed. Variation within the agricultural group may reflect homogenization of pastoral and arable sites, although this is not supported by the range in family richness at cattle-dominated sites, nor would it explain metric score variability within the more homogenous milpa and residential groups. We consider it more likely reflects 'noise' resulting from natural biological variability masking any subtle anthropogenic effects detectable at family-level, and/or incomplete knowledge about the response of Neotropical macroinvertebrates to disturbance effects (e.g. Tomanova & Tedesco, 2007, Jacobsen et al., 2008, Alonso-EguíaLis et al., 2014).

## Knowledge gaps

Non-significant trends evident in structural metric response, including decreases in ASPT-CR and EPT at calcareous agricultural sites and patterns indicative of a subsidy-stress response in most structural metrics at silicate milpa relative to forested sites, suggest more finely resolved taxonomic information might enable better discrimination of low-level effects. Furthermore, we failed to detect significant or predictable change in functional metrics, even where effects were presumed to be greatest. This may reflect our incomplete knowledge of taxon identity and ecology, since patterns contrast with response to change in riparian and depositional characteristics reported elsewhere (Benstead et al., 2003, Buendia et al., 2013).

Using generic level information, Lorion and Kennedy (2009) demonstrated considerable variation in assemblage composition and EPT richness among forested and pasture reaches in Costa Rican streams, even when pasture was buffered by riparian forest. Others have illustrated the high discriminatory power that generic- and species-level information can provide (Dudgeon, 2012, Monk et al., 2012), even where the magnitude of disturbance is small (Hawkins et al., 2000, Benstead et al., 2003, Waite et al., 2004, Buendia et al., 2013). Some traits are considered relatively stable among genera of many families (Dolédec et al., 2000). Yet, the consolidation of trait information to family-level is questionable because of within-family variation and behavioural plasticity, particularly in the feeding habits of Neotropical taxa (Tomanova et al., 2008). We attempted to reflect behavioural plasticity by assigning fuzzy-coded traits (Moya et al., 2011). Ultimately however, this was done using family-level data and traits aggregated from knowledge of taxa inhabiting North and South American streams: equivalent information is not available for Central America. Congeneric species with differing geographical distribution may possess similar traits and show similar responses to environmental change (Dolédec et al., 2000). However, ambiguities do arise at the family-level. In Kenyan streams for example, Dobson et al. (2002) noted that some Baetidae, generally allocated to the grazer-scraper feeding group, appeared to adopt a shredding role. This example, and others (e.g. Tomanova et al., 2008), underscore that the traits we assigned may not correspond closely with those of taxa inhabiting the study streams.

## Conclusions

Overall, our findings correspond with others that have shown family-level metrics can lack the sensitivity required to distinguish subtle effects (Hilsenhoff, 1988, Hawkins et al., 2000, Waite et al., 2004). Importantly our study also highlighted a number of problems associated with the use of family-level metrics in contexts where the implementation of ideal sampling strategies is constrained. Given these findings we conclude with recommendations for the progression of stream bio-assessment under similar circumstances.

### Replication

Our findings imply family-level structural metrics, particularly the ASPT-CR and EPT scores, may have utility for monitoring coarse-scale effects in the study streams. However, analysis of environmental and biological information over a larger study area that allows replication for longitudinal position is recommended to separate natural and anthropogenic effects, and to clarify the nature of metric response. Expanding the study area would allow for replication of sites in independent catchments and would enable consideration of a wider range of human activities and land-use. In addition, a larger sample size would allow use of statistical techniques like mixed-effect modelling to help account for spatial auto-correlation in sample design (Zuur et al., 2009). Such approaches rely on sample sizes that can accommodate fixed and random effects and model intercepts, and in our study sampling effort would have had to have been doubled to achieve potentially meaningful model outputs (A. Zuur, pers. comm.).

### Reducing the taxonomic and ecological knowledge gap

Our results imply a need to resolve family-level data to a finer taxonomic resolution and learn more about how ecological and biological traits vary within families before we can be rigorous in attempts to describe low-level environmental change in the study streams. We were unable to evaluate the similarity of family-level information with analysis of the same data at finer taxonomic resolution (e.g. Buss & Vitorino, 2010) because consistent identification beyond family is impossible given knowledge about the Belizean stream macroinvertebrate fauna (Carrie & Kay, 2014). These uncertainties will be difficult to address until knowledge about the taxonomy, ecological requirements and behaviours of species and genera present in the region is expanded. Taxonomic work in progress for Belize (Carrie & Kay, 2014), recent advances in knowledge and continuing efforts in the wider region (e.g. Buss & Salles, 2007, Tomanova & Tedesco, 2007, Tomanova & Usseglio-Polatera, 2007, Roque et al.,



2008, Brown et al., 2009, Springer et al., 2010, Rios-Touma et al., 2014, Alonso-EguíaLis et al., 2014), and DNA sequence based approaches (Hajibabaei et al., 2011, Stein et al., 2014) should greatly facilitate this endeavour.

### **Compliance with ethical standards**

**Conflict of interest** None of the authors has conflict interests, financial or otherwise.

### **References**

- Alonso-EguíaLis, P., Mora, J. M., Campbell, B. y M. Springer (editores) 2014. Diversidad, conservación y uso de los macroinvertebrados dulceacuícolas de México, Centroamérica, Colombia, Cuba y Puerto Rico. Instituto Mexicano de Tecnología del Agua, Jiutepec, México. 444 p.
- Anderson, M. J., R. N. Gorley & K. R. Clarke, 2008. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. PRIMER-E Ltd., Plymouth, UK.
- AQEM Consortium, 2002. Manual for the application of the AQEM method. A comprehensive method to assess European streams using benthic macroinvertebrates, developed for the purpose of the Water Framework Directive. Version 1.0 <http://www.aqem.de/mains/products.php> Accessed 24th July 2013.
- Armitage, P. D., D. Moss, J. F. Wright & M. T. Furse, 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research* 17: 333-347.
- Benstead, J. P., M. M. Douglas & C. M. Pringle, 2003. Relationships of stream invertebrate communities to deforestation in eastern Madagascar. *Ecological Applications* 13: 1473-1490.
- Bojsen, B. H. & D. Jacobsen, 2003. Effects of deforestation on macroinvertebrate diversity and assemblage structure in Ecuadorian Amazon streams. *Archiv Fur Hydrobiologie* 158: 317-342.
- Boonsoong, B., N. Sangpradub & M. T. Barbour, 2009. Development of rapid bioassessment approaches using benthic macroinvertebrates for Thai streams. *Environmental Monitoring and Assessment* 155: 129-147.
- Bray, J. R. & J. T. Curtis, 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27: 325-349.
- Brown, B. V., A. Borkent, J. M. Cumming, D. M. Wood, N. E. Woodley & M. A. Zumbado, 2009. Manual of Central American Diptera. Volume 1, vol 1. NRC Research Press, Ottawa, Ontario.
- Buendia, C., C. N. Gibbins, D. Vericat, R. J. Batalla & A. Douglas, 2013. Detecting the structural and functional impacts of fine sediment on stream invertebrates. *Ecological Indicators* 25: 184-196.
- Buss, D. F. & F. F. Salles, 2007. Using Baetidae species as biological indicators of environmental degradation in a Brazilian river basin. *Environmental Monitoring and Assessment* 130: 365-372.
- Buss, D. F. & A. S. Vitorino, 2010. Rapid bioassessment protocols using benthic macroinvertebrates in Brazil: evaluation of taxonomic sufficiency. *Journal of the North American Benthological Society* 29: 562-571.

- Capps, K. A., C. N. Bentsen, & A. Ramírez, 2016. Poverty, urbanization, and environmental degradation: urban streams in the developing world. *Freshwater Science* 35: 429-435.
- Carrie, R. H., M. Dobson & J. Barlow, 2015. Effects of geology and season on macroinvertebrates in lowland Belizean streams: implications for tropical bio-assessment. *Freshwater Sciences* 4: 648–662.
- Carrie, R. H. & E. Kay, 2014. Belize. In Alonso-EguíaLis, P., J. M. Mora, B. Campbell & M. Springer (eds) *Diversidad, conservación y uso de los macroinvertebrados dulceacuícolas de México, Centroamérica, Colombia, Cuba y Puerto Rico*. Instituto Mexicano de Tecnología del Agua, Jiutepec, México, 33-62.
- Castillo, L. E., E., Martinez, C. Ruepert, C. Savage, M. Gilek, M. Pinnock & E. Solis, 2006. Water quality and macroinvertebrate community response following pesticide applications in a banana plantation, Limon, Costa Rica. *Science of the Total Environment* 367: 418-432.
- Chang, F., J. E. Lawrence, B. Rios-Touma & V. H. Resh. 2014. Tolerance values of benthic macroinvertebrates for stream biomonitoring: assessment of assumptions underlying scoring systems worldwide. *Environmental monitoring and assessment* 186: 2135-2149.
- Cherrington, E. A., E. Ek, P. Cho, B. F. Howell, B. E. Hernandez, E. R. Anderson, A. I. Flores, B. C. Garcia, A. Sempris & D. E. Irwin, 2010. *Forest Cover and Deforestation in Belize: 1980-2010*. Water Center for the Humid Tropics of Latin America and the Caribbean (Cathalac), Panama, Panama City.
- Clarke, K. R. & R. N. Gorley, 2006. *PRIMER v6: User Manual/Tutorial*. PRIMER-E Ltd, Plymouth, UK.
- Clarke, K. R. & R. M. Warwick, 2001. *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*, 2nd edn edn. Plymouth Marine Laboratory, Plymouth, UK.
- Cohen, B., 2006. Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technology in Society* 28: 63-80.
- Connolly, N. M., M. R. Crossland & R. G. Pearson, 2004. Effect of low dissolved oxygen on survival, emergence, and drift of tropical stream macroinvertebrates. *Journal of the North American Benthological Society* 23: 251-270.
- Dickens, C. W. S. & P. M. Graham, 2002. The South African Scoring System (SASS) Version 5 Rapid Bioassessment Method for Rivers. *African Journal of Aquatic Science* 27: 1-10.
- Dobson, M., A. Magana, J. M. Mathooko & F. K. Ndegwa, 2002. Detritivores in Kenyan highland streams: more evidence for the paucity of shredders in the tropics? *Freshwater Biology* 47: 909-919.
- Dolédec, S., J. Olivier & B. Statzner, 2000. Accurate description of the abundance of taxa and their biological traits in stream invertebrate communities: effects of taxonomic and spatial resolution. *Archiv Fur Hydrobiologie* 148: 25-43.
- Dudgeon, D., 2012. Responses of benthic macroinvertebrate communities to altitude and geology in tributaries of the Sepik River (Papua New Guinea): the influence of taxonomic resolution on the detection of environmental gradients. *Freshwater Biology* 57: 1794-1812.
- Dufrêne, M. & P. Legendre, 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345-366.

- Esselman, P. C., 2001. An Ecological Baseline Study of Physical Habitat, Water Chemistry and Fish Assemblages of the Monkey River Watershed, Belize. MSc. Thesis, University of Georgia.
- Friberg, N., N. Bonada, D. C. Bradley, M. J. Dunbar, F. K. Edwards, J. Grey, R. B. Hayes, A. G. Hildrew, N. Lamouroux, M. Trimmer & G. Woodward, 2011. Biomonitoring of human impacts in freshwater ecosystems: the good, the bad and the ugly. In Woodward, G. (ed) *Ecosystems in a Human-Modified Landscape: A European Perspective*: 44 (Advances in Ecological Research). Academic Press, London, 1-68.
- Hajibabaei, M., S. Shokralla, X. Zhou, G. A. C. Singer & D. J. Baird, 2011. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLoS one* 6: e17497.
- Harding, J., J. M. Quinn & C. W. Hickey, 2000. Effects of mining and production forestry. In Collier, K. J. & W. M.J. (eds) *New Zealand stream invertebrates: ecology and implications for management*. New Zealand Limnological Society, Christchurch, 230-259.
- Hawkins, C. P., R. H. Norris, J. N. Hogue & J. W. Feminella, 2000. Development and evaluation of predictive models for measuring the biological integrity of streams *Ecological Applications* 10: 1456-1477.
- Heino, J., 2014. Taxonomic surrogacy, numerical resolution and responses of stream macroinvertebrate communities to ecological gradients: Are the inferences transferable among regions? *Ecological Indicators* 36: 186-194.
- Helson, J. E., & Williams, D. D. 2013. Development of a macroinvertebrate multimetric index for the assessment of low-land streams in the neotropics. *Ecological Indicators* 29: 167-178.
- Hilsenhoff, W. L., 1988. Rapid field assessment of organic pollution with a family-level biotic index. *Journal of the North American Benthological Society* 7: 65-68.
- Jacobsen, D., 1998. The effect of organic pollution on the macroinvertebrate fauna of Ecuadorian highland streams. *Archiv Fur Hydrobiologie* 143: 179-195.
- Jacobsen, D., C. Cressa, J. M. Mathooko & D. Dudgeon, 2008. Macroinvertebrates: Composition, Life Histories and Production. In Dudgeon, D. (ed) *Tropical stream ecology*. Elsevier Academic Press, London, 66-96.
- Jones, F. C., 2008. Taxonomic sufficiency: The influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. *Environmental Reviews* 16: 45-69.
- Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger & D. V. Peck, 1999. Quantifying Physical Habitat in Wadeable Streams. Environmental Protection Agency, , Washington, D.C., 102.
- Knee, K. L. & A. C. Encalada, 2013. Land Use and Water Quality in a Rural Cloud Forest Region (Intag, Ecuador) *River Research and Applications* 30: 385-401.
- Kobingi, N., P. O. Raburu, F. O. Masese & J. Gichuki, 2009. Assessment of pollution impacts on the ecological integrity of the Kisian and Kisat rivers in Lake Victoria drainage basin, Kenya. *African Journal of Environmental Science and Technology* 3: 097-107.
- Kolkwitz, R. & M. Marsson, 1909. Ökologie der tierischen Saprobien. Beiträge zur Lehre von der biologischen Gewässerbeurteilung. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* 2: 126-152.

- Kratzer, E. B., J. K. Jackson, D. B. Arscott, A. K. Aufdenkampe, C. L. Dow, L. A. Kaplan, J. D. Newbold & B. W. Sweeney, 2006. Macroinvertebrate distribution in relation to land use and water chemistry in New York City drinking-water-supply watersheds. *Journal of the North American Benthological Society* 25: 954-976.
- Lenat, D. R. & 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.
- Lewis, W. M. J., 2008. Physical and chemical features of tropical flowing waters. In Dudgeon, D. (ed) *Tropical stream ecology*. Academic Press, USA, 1-22.
- Ligeiro, R., R. M. Hughes, P. R. Kaufmann, D. R. Macedo, K. R. Firmiano, W. R. Ferreira, D. Oliveira, A. S. Melo & M. Callisto, 2013. Defining quantitative stream disturbance gradients and the additive role of habitat variation to explain macroinvertebrate taxa richness. *Ecological Indicators* 25: 45-57.
- Lorion, C. M. & B. P. Kennedy, 2009. Relationships between deforestation, riparian forest buffers and benthic macroinvertebrates in neotropical headwater streams. *Freshwater Biology* 54: 165-180.
- Merritt, R. W., K. W. Cummins & M. B. Berg, 2008. *An Introduction to the Aquatic Insects of North America* 4th edn. Kendall Hunt Publishing Company, Iowa, USA.
- Metzeling, L., S. Perriss & D. Robinson, 2006. Can the detection of salinity and habitat simplification gradients using rapid bioassessment of benthic invertebrates be improved through finer taxonomic resolution or alternative indices? *Hydrobiologia* 572: 235-252.
- Monk, W. A., P. J. Wood, D. M. Hannah, C. A. Extence, R. P. Chadd & M. J. Dunbar, 2012. How does macroinvertebrate taxonomic resolution influence ecohydrological relationships in riverine ecosystems. *Ecohydrology* 5: 36-45.
- Moya, N., R. M. Hughes, E. Domínguez, F.-M. Gibon, E. Goitia & T. Oberdorff, 2011. Macroinvertebrate-based multimetric predictive models for evaluating the human impact on biotic condition of Bolivian streams. *Ecological Indicators* 11: 840-847.
- Mugnai, R., J. L. Nessimian & D. F. Baptista, 2010. *Manual de Identificacao de Macroinvertebrados Aquaticos*. Technical Books Editora, Rio de Janeiro.
- Murphy, J. F., J. Davy-Bowker, B. McFarland & S. J. Ormerod, 2013. A diagnostic biotic index for assessing acidity in sensitive streams in Britain. *Ecological Indicators* 24: 562-572.
- Naranjo López, C., D. D. González Lazo, A. L. Blandimarte, S. Munoz Riveaux & Y. Musle Cordero, 2005. Una metodología rápida y de fácil aplicación para la evaluación de la calidad del agua utilizando el índice BMWP-Cub para ríos cubanos. *Revista Tecnura* 9: 65-76.
- Natural Resources Conservation Service, 1998. *Stream Visual Assessment Protocol*, Washington, D.C.
- Neff, M. R. & D. A. Jackson, 2011. Effects of broad-scale geological changes on patterns in macroinvertebrate assemblages. *Journal of the North American Benthological Society* 30: 459-473.
- Odum, E. P., J. T. Finn & E. H. Franz, 1979. Perturbation theory and the subsidy-stress gradient. *Bioscience* 29: 349-352.
- Peck, D. V., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kauffman, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, T. Magee & M. Cappaert, 2006. *Environmental Monitoring and*

- Assessment Program-Surface Waters Western Pilot Study: Field Operations Manual for Wadeable Streams. U.S. Environmental Protection Agency, Washington, D.C.
- Poquet, J. M., J. Alba-Tercedor, T. Punti, M. D. Sanchez-Montoya, S. Robles, M. Alvarez, C. Zamora-Munoz, C. E. Sainz-Cantero, M. R. Vidal-Abarca, M. L. Suarez, M. Toro, A. M. Pujante, M. Rieradevall & N. Prat, 2009. The MEDiterranean Prediction And Classification System (MEDPACS): an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities. *Hydrobiologia* 623: 153-171.
- R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raburu, P. O., J. B. Okeyo-Owuor & F. O. Masese, 2009. Macroinvertebrate-based Index of biotic integrity (M-IBI) for monitoring the Nyando River, Lake Victoria Basin, Kenya. *Scientific Research and Essays* 4: 1468-1477.
- Resh, V. H., 2007. Multinational, freshwater biomonitoring programs in the developing world: Lessons learned from African and Southeast Asian river surveys. *Environmental Management* 39: 737-748.
- Richards, C., R. Haro, L. Johnson & G. Host, 1997. Catchment and reach-scale properties as indicators of macroinvertebrate species traits. *Freshwater Biology* 37: 219-230.
- Ríos-Touma, B., Acosta, R., & Prat, N. 2014. The Andean Biotic Index (ABI): revised tolerance to pollution values for macroinvertebrate families and index performance evaluation. *Revista de Biología Tropical*, 62: 249-273.
- Roque, F. O., L. S. Lecci, T. Siqueira & C. G. Froehlich, 2008. Using environmental and spatial filters to explain stonefly occurrences in southeastern Brazilian streams: implications for biomonitoring. *Acta Limnol. Brasiliensia* 20: 35-44.
- Schäfer, R. B., T. Caquetc, K. Siimesd, R. Muellere, L. Lagadicc & M. Liessa, 2007. Effects of pesticides on community structure and ecosystem functions in agricultural streams of three biogeographical regions in Europe. *Science of the Total Environment* 382: 272-285.
- Smith, M. J., W. R. Kay, D. H. D. Edward, P. J. Papas, K. S. J. Richardson, J. C. Simpson, A. M. Pinder, D. J. Cale, P. H. J. Horwitz, J. A. Davis, F. H. Yung, R. H. Norris & S. A. Halse, 1999. AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. *Freshwater Biology* 41: 269-282.
- Springer, M., A. Ramirez & P. Hanson, 2010. Macroinvertebrados de Agua Dulce de Costa Rica I, vol 58 (Supl. 4). *Revista de Biología Tropical*, San Jose, Costa Rica.
- Sutherland, A. B., J. M. Culp & G. A. Benoy, 2012. Evaluation of deposited sediment and macroinvertebrate metrics used to quantify biological response to excessive sedimentation in agricultural streams. *Environmental Management* 50: 50-63.
- Statistical Institute of Belize, 2016. Population estimates by Major Administrative Areas and Sex, 2008-2016. <http://www.sib.org.bz/statistics/population>. Accessed 08/11/16.

- Stein, E. D., B. P. White, R. D. Mazon, J. K. Jackson, J. M. Battle, P. E. Miller, E. M. Pilgrim, & B. W. Sweeney, 2014. Does DNA Barcoding Improve Performance of Traditional Stream Bioassessment Metrics? *Freshwater Science* 33: 302-311.
- Taylor, B. R., 1997. Rapid assessment procedures: Radical re-invention or just sloppy science? *Human and Ecological Risk Assessment: An International Journal* 3: 1005-1016.
- Thorne, R. S. & W. P. Williams, 1997. The response of benthic macroinvertebrates to pollution in developing countries: A multimetric system of bioassessment. *Freshwater Biology* 37: 671-686.
- Tomanova, S., N. Moya & T. Oberdorff, 2008. Using macroinvertebrate biological traits for assessing biotic integrity of neotropical streams. *River Research and Applications* 24: 1230-1239.
- Tomanova, S. & P. A. Tedesco, 2007. Body size, ecological tolerance and potential for water quality bioindication in the genus *Anacroneuria* (Plecoptera: Perlidae) from South America. *Revista De Biologia Tropical* 55: 67-81.
- Tomanova, S. & P. Usseglio-Polatera, 2007. Patterns of benthic community traits in neotropical streams: relationship to mesoscale spatial variability. *Fundamental and Applied Limnology* 170: 243-255.
- Waite, I. R., A. T. Herlihy, D. P. Larsen, N. S. Urquhart & D. J. Klemm, 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, USA. *Freshwater Biology* 49: 474-489.
- Woodiwiss, F. S., 1964. The biological system of stream classification used by the River Trent Board. *Chemistry and Industry* 14: 443-447.
- WWAP (United Nations World Water Assessment Programme). 2015. The United Nations World Water Development Report 2015: Water for a Sustainable World. Paris, UNESCO
- Wright, J. F., 2000. An introduction to RIVPACS. Freshwater Biological Association.
- Wyżga, B., P. Oglęcki, H. Hajdukiewicz, J. Zawiejska, A. Radecki-Pawlik, T. Skalski & P. Mikuś, 2013. Interpretation of the invertebrate-based BMWP-PL index in a gravel-bed river: insight from the Polish Carpathians. *Hydrobiologia* 712: 71-88.
- Yates, A. & R. Bailey, 2010. Covarying patterns of macroinvertebrate and fish assemblages along natural and human activity gradients: implications for bioassessment. *Hydrobiologia* 637: 87-100.
- Zuur, A., E. N. Ieno, N. Walker, A. A. Saveliev & G. M. Smith, 2009. Mixed effects models and extensions in ecology with R. Springer, New York.

## Tables

Table 1. Summary of PCA information for disturbance in silicate streams. The highest loading for each environmental variable is given in bold. Variables describing longitudinal position are shaded in grey.

	PC1	PC2	PC3
Eigenvalue	7.38	3.38	1.51
Cumulative % variation	46.1	67.3	76.7
%Fines	<b>-0.956</b>	-0.038	-0.033
Distance Sea	<b>0.898</b>	-0.068	0.349
%Embeddedness	<b>-0.881</b>	0.353	-0.043
Dissolved Oxygen	<b>0.864</b>	-0.028	-0.085
SVAP <sup>a</sup>	<b>0.826</b>	0.158	0.413
Turbidity	<b>-0.823</b>	0.028	0.018
pH	<b>0.757</b>	0.129	0.421
Riparian Width	<b>0.747</b>	0.348	0.323
%Sand	<b>-0.720</b>	-0.015	-0.234
Conductivity	-0.336	<b>-0.835</b>	-0.159
Stream Size	0.519	<b>-0.814</b>	-0.091
Temperature	0.161	<b>0.779</b>	0.093
%Gravel	-0.084	<b>0.742</b>	-0.484
PO <sub>4</sub> <sup>3-</sup>	0.102	<b>-0.635</b>	-0.240
Riparian Canopy	0.043	0.218	<b>0.830</b>
NO <sub>3</sub> <sup>-</sup> N	0.368	0.044	<b>0.728</b>

<sup>a</sup>Stream Visual Assessment Protocol Score



Table 2. Summary of PCA information for disturbance in calcareous streams. The highest loading for each environmental variable is given in bold. Variables describing longitudinal position are shaded in grey.

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	4.43	3.29	2.06	1.71	1.10
Cumulative % variation	27.9	48.3	61.1	71.8	78.7
SVAP <sup>a</sup>	<b>0.901</b>	-0.111	-0.070	-0.067	-0.012
NO <sub>3</sub> <sup>-</sup> N	<b>0.766</b>	0.343	0.060	0.140	0.065
%Embeddedness	<b>-0.733</b>	-0.046	0.399	0.114	0.314
Riparian Width	<b>0.635</b>	-0.326	-0.464	-0.218	0.360
Dissolved Oxygen	<b>0.533</b>	0.505	0.020	0.221	-0.460
PO <sub>4</sub> <sup>3-</sup>	<b>-0.518</b>	0.162	0.292	-0.383	-0.078
Stream Size	-0.239	<b>0.902</b>	0.063	0.013	0.041
Distance Sea	0.247	<b>0.808</b>	0.164	0.180	-0.306
pH	-0.052	<b>-0.790</b>	0.194	0.202	-0.130
Turbidity	-0.132	0.057	<b>0.926</b>	0.133	-0.133
%Fines	-0.577	0.037	<b>0.607</b>	-0.160	0.175
Temperature	-0.241	-0.445	<b>0.594</b>	-0.399	-0.026
%Sand	-0.354	0.458	<b>-0.466</b>	0.391	0.291
Conductivity	-0.123	0.156	0.174	<b>0.883</b>	0.230
%Gravel	0.092	-0.375	-0.272	<b>0.645</b>	-0.353
Riparian Canopy	0.012	0.012	-0.082	0.117	<b>0.780</b>

<sup>a</sup>Stream Visual Assessment Protocol Score

Table 3. Summary of indicator analysis for treatment group comparison in silicate streams. Shown are the families with highest indicator values calculated based on the relative abundance and frequency of occurrence of each species among the a priori treatment categories.

Family	Indicator Value		Indicator Value	
	Multiple Stressors	Forest	Multiple Stressors	Milpa
Thiaridae	63.8*	36.1		
Naucoridae	61.5*	38.5		
Perlidae	7.7	82.7**		
Odontoceridae	0	66.7*		
Athericidae	0	66.7**	0	60.0*
Psephenidae	11.0	62.7*	9.7	78.2*
Elmidae	40.6	59.4*	38.0	62.0*
Gerridae	0.01	58.8*		
Calamoceratidae			0	80.0**
Hydroptilidae			13.6	75.4*
Limoniidae			19.0	71.4*
Coenagrionidae			24.6	68.4*
Hydrobiosidae			0	60.0*
Psychodidae			0	60.0*

\*\* significant at  $p < 0.01$ , \* significant at  $p < 0.05$

## Figure Captions

**Fig 1** Location of the study area and sites sampled in southern Belize. Sites sampled in the Rio Grande, Golden Stream and Deep River systems are considered to be calcareous, and the remaining sites silicate.

**Fig 2** Non-metric multidimensional scaling (nMDS) biplot showing Bray-Curtis similarity in community composition of macroinvertebrate assemblages at silicate and calcareous sites

**Fig 3** Non-metric multidimensional scaling (nMDS) biplot showing similarity of community composition at sites within each treatment group in (a) silicate and (b) calcareous streams

**Fig 4** Variation in (a,b) Family Richness, (c,d) EPT Richness, (e,f) BMWP Score and (g,h) ASPT Score among treatment groups in silicate and calcareous streams. F = Forest, M = Milpa, M St = Multiple Stressor, R = Residential and A = Agriculture. Boxes are interquartile ranges, centre lines are medians and error bars indicate 95<sup>th</sup> and 5<sup>th</sup> percentiles