Implications of macroinvertebrate taxonomic resolution for freshwater assessments using functional traits: The Paute River Basin (Ecuador) case

Gonzalo Sotomayor1,2 | Henrietta Hampel2 | Raúl F. Vázquez2,3 | Marie Anne Eurie Forio1 | Peter L. M. Goethals1

1Faculty of Bioscience Engineering, Department of Animal Sciences and Aquatic Ecology, Ghent University, Ghent, Belgium
2Facultad de Ciencias Químicas, Laboratorio de Ecología Acuática (LEA), Universidad de Cuenca, Cuenca, Ecuador
3Facultad de Ingeniería, Departamento de Ingeniería Civil, Universidad de Cuenca, Cuenca, Ecuador

Correspondence
Gonzalo Sotomayor, Faculty of Bioscience Engineering, Department of Animal Sciences and Aquatic Ecology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium. Email: gonzalo.sotomayor@ugent.be

Abstract
Aim: This study aimed at investigating the taxonomic resolutions (TRs) of benthic macroinvertebrates for freshwater assessments in the scope of the functional trait approach (FTA).

Location: Macroinvertebrate samples were collected in 22 locations within the Paute River Basin (PRB), Ecuador, over three years (2010, 2011 and 2012).

Methods: Biological traits were allocated as scores to the macroinvertebrate data (at genus level) through fuzzy coding, using published data. The scores of each genus were used to derive scores for the corresponding family. These two sets of scores were standardized and compared, they were similar in 82% of the cases. Functional diversity (FD) was described by the rRao index, which showed no significant differences between coarse (family level) and fine (genus level) TRs. Cluster analyses using the K-means algorithm were performed to determine similarities between both rRao data sets. The WQ cluster number (K) was varied between 2 and 5 to determine a threshold K value (Kth), after which a WQ assessment differed as a function of the TR being used.

Results: Kth was 3. Family-level identification in the framework of the FTA in the PRB was suitable in detecting changes of macroinvertebrate assemblages (until Kth = 3).

Main conclusions: The proposed methodology could be implemented in other basins where decision-makers could decide whether the level of functional trait data similarity is sufficient for WQ management purposes and whether the defined Kth is acceptable. The reliability of the key methodological steps was assessed using performance statistics that have rarely been applied to ecological studies. Despite related research performed in other regions, the present study is the first South American attempt to investigate the effects of TR of benthic macroinvertebrates on freshwater bioassessments using functional traits.

Keywords
Andean basin, cluster analysis, functional traits, rRao index, taxonomic resolution
Traditionally, taxonomic-based metrics have been used as indicators of aquatic ecosystem integrity and conservation status (Bellucci et al., 2013; Bück et al., 2010; Sotomayor et al., 2020). However, the functional trait approach (FTA) has emerged in the last decades as one of the most promising tools for bioassessment of freshwater ecosystems (Bruno et al., 2016). Functional traits are morphological, physiological, phenological or behavioural characteristics of organisms that influence their performance or fitness (Nock et al., 2016). Traits are usually divided into two categories (Menezes et al., 2010), that is biological (e.g. maximum body size, life span, feeding and reproductive strategies, mobility) and ecological (related to habitat preferences, such as pH and temperature tolerances, organic pollution tolerance). The ecological functions of species can be described by a multitude of biological traits that reflect their adaptation to environmental disturbances (Townsend & Hildrew, 1994). For example, the maximal size, body shape, life span, reproductive cycle and feeding strategies are typical traits reflecting the adaptation of species to more or less disturbed environments (Poff et al., 2006). As these traits are comparable across taxonomic groups, they provide a unified measure through communities differing in taxonomic composition. For bioassessment, this unified measure is relevant because the FTA implies that environmental filters select taxa with suitable traits to coexist under similar environmental conditions or types of disturbances (Poff, 1997; Storey et al., 1991).

Benthic macroinvertebrates play an important role in the stream ecosystem functions, that is, influence on nutrient cycles, primary productivity, decomposition and translocation of materials (Wallace & Webster, 1996). Despite the fact that several significant studies have been carried out on benthic macroinvertebrates and their functional traits (Bêche & Resh, 2007; Chará-Serna et al., 2012; Forio et al., 2018; Príncipe et al., 2010; Reynaga et al., 2020), there exist substantial debate among biologists on the most appropriate approaches for sampling design, field sampling, laboratory analyses and data assessments. A point of disagreement in freshwater bioassessment using benthic macroinvertebrates has been the level at which they should be identified, that is the relevant taxonomic resolution (Bailey et al., 2001; Dolédec et al., 2000; Gayraud et al., 2003; Jones, 2008; Mueller et al., 2013; Vieira et al., 2006). Taxonomic resolution (TR) varies depending on the goals of a study as the level of identification should match the need for information (Ferraro & Cole, 1995). Hence, under ideal conditions the default taxonomic level for bioassessments should be species; however, under real-world constraints (i.e. financial, logistical, lack of local taxonomic knowledge) detailed high TRs (i.e. at genus or species level) are not an option for many regions in the world.

In areas where coarser TRs (i.e. at family level) are the only possibility, certain aspects have to be considered, including (a) how well finer TR data (e.g. species or genus composition) are represented by coarser TR data? and (b) how much loss in accuracy of freshwater quality assessments is produced by the availability of coarser TR data instead of finer TR data? Arguing that taxonomic composition provides little information on ecosystem functioning, many researchers have been working instead on the FTA, in which site-by-taxon data sets are re-defined, using functional or biological trait information, into site-by-trait data sets (Jones, 2008). In the FTA, the generated trait databases are defined either on finer or on coarser TR, as it is also the case in Latin America (Forio, 2017; Ramírez & Gutiérrez-Fonseca, 2014; Tomanova & Usseglio-Polatera, 2007).

The main goal of this research was to study the TR of benthic macroinvertebrates using functional traits for freshwater assessments. Hence, the specific goals were to (a) quantify the difference between families and genera in terms of functional description; (b) define and apply a functional diversity index for site characterization using families and genera; and (c) compare sampling site grouping on the basis of functional diversity as per TR (i.e. families or genera), using an appropriate cluster algorithm. While related research has been performed in other regions (e.g. Dolédec et al., 2000; Gayraud et al., 2003; Mueller et al., 2013), it has hardly considered study sites with pronounced elevational and climatic gradients as the ones that characterize the current study site and, definitively, has not used the methodology that has been implemented herein. Indeed, the present study is the first South American attempt to investigate the effects of taxonomic resolution of benthic macroinvertebrates on freshwater bioassessment using functional traits. Particularly, very few studies have considered how taxonomic resolution influences...
functional diversity measures (Schmera et al., 2017), which emphasizes the international relevance of this study.

2 | METHODS

2.1 | Study area

The Paute River Basin (PRB) is located in the south of Ecuador (Figure 1). It is of high importance owing to its significant hydroelectric potential (CONELEC, 2011; Salazar & Rudnick, 2008). The basin has an area of 6,442 km² including the eastern lower portion towards the Amazon plateau. Its elevational ranges between 500 and 4,250 m above sea level (a.s.l.), and slopes vary between 25% and 50%. The climatology of the basin is highly variable. Its average air temperature varies between 4.4°C and 18.6°C. This wide range of variation of the average air temperature depicts the high variability of this meteorological variable in response to the significant elevational gradient of the study site. The lower air temperatures correspond to the western Andes range with an average of 6°C (Páramo), while the warmest areas are situated in the valleys and subtropical zones (Amazonia region), with an interannual average fluctuating between 22°C and 26°C. Due to the wide elevational range, rainfall oscillates in intensity and duration, with maximum annual averages between 2,500 and 3,000 mm at higher elevations and minimum annual averages between 600 and 800 mm in the valleys (Celleri et al., 2007). Two major cities, namely Cuenca and Azogues, are located in the basin with approximately 500,000 and 33,850 inhabitants, respectively (2010 census). Pollution in the study basin comes from point and non-point sources, including domestic wastewaters, agricultural runoff, animal husbandry and industrial effluents (Da Ros, 1995). Degraded streams are characterized by high values of faecal coliforms, chlorides, total hardness, nitrate, total alkalinity, biochemical oxygen demand, turbidity, water temperature and low dissolved oxygen values. On the contrary, pristine sites are associated with good plans of land use (i.e. protected areas) and adequate habitat conditions (i.e. riparian vegetation, streambed heterogeneity) (Sotomayor, 2016; Sotomayor et al., 2018; Sotomayor et al., 2020).

2.2 | Sampling and analysis of benthic macroinvertebrates

Macroinvertebrate data were collected from 22 sampling sites located in the study basin within three years (2010, 2011 and 2012) repeatedly, by the former Ecuadorian National Secretary of Water (SENAGUA)—Santiago River Hydrographic Demarcation (DHS). At each sampling site, a 20-m-long reach was selected. Three evenly spaced transects were delineated throughout the 20-m-long reach (Von Ellenrieder, 2007). Macroinvertebrate samples were collected by a standard “kick-sampling” using a 25 × 25 cm² nylon handnet (mesh size: 0.5 mm) placed on the stream bottom, in front of which (upstream) the substratum was vigorously stirred by kicking it (Jacobsen et al., 1997). Sampling was carried out for three minutes, encompassing all existing microhabitats characterized by different depths, substrates and water velocities. The three samples from each transect were pooled together, and sampling continued by visually inspecting (for about 20 min) the substrate and aquatic vegetation to collect the tightly clinging taxa (e.g. Blephariceridae or Elmidae) that may have not been dislodged by kick-sampling (Roldán, 2003). Macroinvertebrate samples were preserved in 70% ethanol and sorted with the use of a stereomicroscope.

On average, the sampling sites were visited four times per year. Some were sampled more frequently, because they were located either at highly impacted sites (e.g. sites B2, B8 y B13 in Figure 1) or, on the contrary, at unaltered environmental (i.e. reference) locations (e.g. sites SB4, COL1, PIN5 in Figure 1). As a result, a benthic macroinvertebrate database was developed for nrep = 22 sampling sites, nrep = 230 sampling replicates that overall contain ngen = 79 genera, ngen = 44 families and nord = 17 superior taxonomic groups (in its great majority orders).

2.3 | Allocation of biological traits

Seven biological trait categories were selected (Table 1). They are well documented as good indicators for capturing compositional changes due to ecohydrological stress (Forio, 2017; Reynaga et al., 2020). Information on individual traits was gathered from literature about research applied on Neotropical zones, that is Tomanova et al. (2006, 2007); Chará-Serna et al. (2012); Ramírez and Gutiérrez-Fonseca (2014) and Longo and Blanco (2014). Biological trait databases at the genus (g) level (Dbg-trait) were constructed (Figure 2) on the basis of the affinity of each taxon for trait categories using affinity scores within a fuzzy coding procedure (Chevenet et al., 1994) upon literature information. A score value of 0 indicates that there is no affinity of the taxon for the trait category, 1 low affinity, 2 medium affinity and 3 strong affinity.

2.4 | Trait data processing

Two levels of TRs, namely, genus (g) and family (f), were used in the scope of the FTA (Figure 2). The following protocol, adapted from Gayraud et al. (2003), was used for data preprocessing to provide the same weight to each taxon and every biological trait, prior to further analyses:

1. For all nrep, the affinity scores of each genus from Dbg-trait were standardized so that the sum of affinity scores to a given trait equalled 1 (zg-trait).
2. The standardized affinity scores of each genus from zg-trait were used to derive corresponding scores at family level by averaging the affinity scores of all genera belonging to a given family (Dbf-trait). The latter were then rescaled so that their sum, for a
TABLE 1  Traits and categories used in the study

| Trait                  | Category                                      |
|------------------------|-----------------------------------------------|
| Feeding habits         | Collector–filterer (C-Ft)                    |
|                        | Collector–gatherer (CG)                      |
|                        | Piercers (Pc)                                |
|                        | Predators (Pr)                               |
|                        | Scrapers (Sc)                                |
|                        | Shredders (Sh)                               |
|                        | Parasite (PA)                                |
| Respiration            | Tegument (Teg)                               |
|                        | Gill (Gill)                                  |
|                        | Plastron (Pla)                               |
|                        | Spiracle (Spi)                               |
| Body form              | Streamlined                                  |
|                        | Flattened                                    |
|                        | Cylindrical                                  |
|                        | Spherical                                    |
| Maximum body size (mm) | <2.5                                         |
|                        | 2.5–5                                        |
|                        | 5–10                                         |
|                        | 10–20                                        |
|                        | 20–40                                        |
|                        | 40–80                                        |
| Body flexibility       | None (<10°)                                  |
|                        | Low (>10°–45°)                               |
|                        | High (>45°)                                  |
| Mobility and attachment to substratum (locomotion) | Flier (Fli)                                   |
|                        | Surface swimmer (SS)                         |
|                        | Full water swimmer (FWS)                    |
|                        | Crawler (Cra)                                |
|                        | Burrower (Bur)                               |
|                        | Temporarily attached (TA)                   |
| Reproduction           | Asexual (As)                                 |
|                        | Clutches and cemented (CC)                   |
|                        | Clutches and free (CF)                       |
|                        | Clutches in vegetation (CV)                  |
|                        | Clutches and terrestrial (CT)                |
|                        | Isolated eggs and clutches (IEC)             |
|                        | Isolated eggs and free (IEF)                |
|                        | Oovoviviparity (Ovi)                         |

| Trait category proportions were calculated by multiplying the standardized affinity scores (either \(z_{g-\text{trait}}\) or \(z_{f-\text{trait}}\)) of each taxon with its correspondent log(x + 1)-transformed abundance (Latli et al., 2017; Schmera et al., 2014). A log(x + 1) transformation was used instead of a ln(x) transformation owing to the fact that for several taxa, an abundance of 1, leading to a ln value equal to 0, was observed. The resulting trait proportions were rescaled (rsc) so that their sum equalled 1 for a given trait and river reach. As a result, two databases (\(DB_{g-\text{rsc}}\) and \(DB_{f-\text{rsc}}\)) with standardized information of both TRs were derived.

2.5 | Comparison of taxonomic resolutions

To quantify the difference between genera (\(DB_{g-\text{rsc}}\)) and families (\(DB_{f-\text{rsc}}\)) in terms of their functional descriptions (i.e. using data of trait standardized abundance at genus and family levels), the differences in trait structure (DTS) were characterized through the Nash and Sutcliffe (1970) coefficient (NSE) and the Pearson-type correlation coefficient (\(r^2\)). The NSE is defined as:

\[
NSE = 1 - \frac{\sum_{i=1}^{n} \left( Txg_i - Txf_i \right)^2}{\sum_{i=1}^{n} \left( Txg_i - Txg \right)^2},
\]

where \(Txg_i\) is the rescaled \(i\)-th functional description of a genus from \(DB_{g-\text{rsc}}\), \(Txf_i\) is the rescaled \(i\)-th functional description of the linked family (simulated from the genera that it encompasses), \(Txg\) is the rescaled mean of the functional description of the corresponding genus, and \(n\) is the total number of observations. NSE is dimensionless and varies in the interval \(-\infty < NSE \leq 1\). NSE = 1 indicates a perfect coincidence of the two data sets being compared, while NSE \(\leq 0\) characterizes completely dissimilar data sets. The higher the NSE value in the interval [0, 1], the more similar are the two data sets (Ritter & Muñoz-Carpena, 2013; Vázquez et al., 2002; Wu et al., 2006). The \(r^2\) index can range from 0 (no correlation) to 1 (perfect correlation). Higher values indicate higher correlation between the two data sets being compared. Despite some pitfalls of \(r^2\) for characterizing the similarity of two data sets (Legates & McCabe, 1999; Vázquez et al., 2002), it has been used in this study as a complement to the information provided by the NSE, particularly on the correlation between both data sets. Depending on the context of its use, typically, values greater than 0.7 are considered to reflect an acceptable correspondence between two data sets (Santhi et al., 2001; Schober & Schwarte, 2018; Vázquez, 2003).

In the context of (numerical) modelling, where observed and simulated data sets are compared with each other, some authors have adopted scales of agreement such as (Moriasi et al., 2007; Parajuli et al., 2009): excellent (\(r^2\) and/or NSE > 0.90), very good (\(r^2\) and/or NSE in the range 0.75–0.89), good (\(r^2\) and/or NSE in the range 0.50–0.74), fair (\(r^2\) and/or NSE in the range 0.25–0.49) and poor (\(r^2\) and/or NSE in the range 0–0.24). Hence, a threshold value of 0.75, for either index, has been adopted herein to depict an acceptable similarity between two data sets. The empirical exceedance probability distribution of both indices was calculated (Figure 3a) on the basis of the Weibull plotting position method (Makkonen, 2006; Vázquez, 2003).
Calculating functional diversity

Functional diversity (FD) can be summarized using indices based on trait values and taxa dominance in the community, that is abundance (Pla et al., 2012). Rao’s quadratic diversity index (Q) (Rao, 1982) is one of the most important FD metrics (Mouchet et al., 2010) that has been widely used to measure the effects of anthropogenic activities on the FD of aquatic organisms. In general, the observed pattern is a reduction in Q values for impacted areas (Desrosiers et al., 2019; Magliozzi et al., 2020; Meng et al., 2021). Nevertheless, Q suffers from some pitfalls (Chen et al., 2018; Liu & Rao, 1995; Pons & Petit, 1996), among which the fact that its range of values depends on the scale of the dissimilarities between the compared data sets is one of the most relevant. Consequently, the rRao variation of Q that overcomes this problem (Pavoine et al., 2005) by standardizing Q with respect to its maximum value (Champely & Chessel, 2002) was used. rRao is determined from the computation of the Euclidean distance ($d_{ij}$), between taxa $i$ and $j$ in the trait space, as a measure of dissimilarity between $i$ and $j$ (Chen et al., 2002):

$$rRao = \frac{\sum_{i=1}^{N-1} \sum_{j=1}^{N} w_i w_j \left(\delta_{ij}\right)^2}{2}$$

where $N$ is the number of taxa, $w_i$ is the relative frequency of taxa $i$, $w_j$ is the relative frequency of taxa $j$, and $\delta_{ij} = \sqrt{2d_{ij}}$. Herein, the suggested rRao was calculated for all of the replicates considering both data sets, $D_{B-grep}$ and $D_{B-frep}$, resulting respectively in data sets $rRao_{g-rep}$ and $rRao_{f-rep}$. rRao varies in the interval $[0, 1]$, with 0 representing a total loss of FD, that is, a heavily impacted stream ecosystem. Being dimensionless, rRao enables a direct comparison between data of both TRs.

Defining the WQ class number after which genus- and family-based assessments are dissimilar

2.7.1 Data preprocessing

Firstly, the Kolmogorov–Smirnov test was used to compare the distributions of the two TRs, that is, $rRao_{g-rep}$ and $rRao_{f-rep}$, by computing the maximum distance between the cumulative distributions of both TRs (Thas, 2010). Secondly, the Shapiro–Wilk test (Shapiro & Wilk, 1965) was used to evaluate whether the distributions of $rRao_{g-rep}$ and $rRao_{f-rep}$ data sets were normal by considering all of the available data. This test showed with a 95% confidence level that neither of the two data sets is normally distributed ($p = 0.0001 < \alpha = 0.05$), exhibiting
in both cases a positive skew. Therefore, it was decided (Sotomayor et al., 2018) to aggregate the \( r_{Rao} \) and \( r_{Rao_{rep}} \) data for every one of the sampling sites (i.e. using central tendency values). Thus, individual Shapiro-Wilk tests per site were run for either TR considering all of the data available per site. For a particular site, if the test indicated normality, the average \( r_{Rao} \) value was used for that site, for the respective TR being analysed; otherwise, the median central tendency value was assigned (Anderson & Finn, 1996; Helsel et al., 2020). As a result, two sets of aggregated data were obtained, \( r_{Rao_{g-samp}} \) and \( r_{Rao_{f-samp}} \) each of them containing 22 data in correspondence with the 22 sampling sites.

2.7.2 | Cluster analysis

Both data sets \( r_{Rao_{g-samp}} \) and \( r_{Rao_{f-samp}} \) were used in the context of cluster analysis (CA) through the \( K \)-means algorithm (MacQueen, 1967), which is a non-hierarchical and simple partitional clustering algorithm that attempts to find \( K \) non-overlapping clusters. Thus, \( K \) is the number of clusters, that is, in the current research, the number of FD levels for both TRs, and is defined in advance. Every point in the data is then assigned to the closest cluster centroid, forming a cluster. The centroid of each cluster is then updated based on the points assigned to that particular cluster. This process is repeated until no centroid changes are needed (Wu, 2012). The optimal \( K \)-means clustering in one dimension via dynamic programming (Wang & Song, 2011) was used herein. \( K \) was varied between 2 and 5 for each \( K \)-means run on both data sets. Lower \( K \) values are associated with higher \( r_{Rao} \) values and vice versa. This \( K \) range was fixed in correspondence with what is commonly reported as the number of surface water quality classes (Howladar et al., 2018; Moyel, 2014; Shakhari et al., 2019; Sotomayor et al., 2018).

Once all of the \( K \)-means runs with \( K \) varying in the [2, 5] interval were carried out, a comparative analysis was performed on the two outputs associated with every \( K \) value by using the Youden index (\( \psi \); Youden, 1950), the non-error rate (NER; Ballabio et al., 2018) and a measure of concordance (MoC; Pfitzner et al., 2009) to characterize their similarity. \( \psi \) is a commonly used measure to compare diagnostic abilities of two tests (Biggerstaff, 2000). It is a measure related to single classes; thus, the average (\( \psi \)) for all of the classes was calculated to characterize the whole clustering output. \( \psi = S_n - (1 - S_p) \), with \( S_n \) being the sensitivity and \( S_p \) being the specificity that, in turn, are classification parameters of the h-th class and are related to the confusion matrix (CM). \( S_n \) represents the ability of a given classifier to correctly identify the samples of the h-th class. The \( S_p \) represents the capability of a given classifier to reject samples of the other classes. It is calculated as the ratio of samples not belonging to the h-th class, which were not classified in the h-th class over the total number of samples not belonging to the h-th class (Sokolova et al., 2006). The NER is the average of \( S_n \) and provides an overall evaluation of the classification quality. \( \psi \) and NER have values between 0 (no class discrimination) and 1 (perfect class discrimination) (Frank & Todeschini, 1994).

Finally, the MoC was used for characterizing the extent to which two cluster outputs agree with each other. Hence, given that there are I clusters in \( K \)-means\(_{g-K} \) and J clusters in \( K \)-means\(_{f-K} \), and that each individual cluster in \( K \)-means\(_{g-K} \) is referred to as \( g_i \) and each cluster in \( K \)-means\(_{f-K} \) as \( f_j \), for \( i \in \{1, 2, ..., I\} \) and \( j \in \{1, 2, ..., J\} \), then any cluster \( g_i \) from \( K \)-means\(_{g-K} \) can be subdivided into smaller subclusters or fragments. A fragment \( m_{ij} \) consists of those elements of \( g_i \) that have also been allocated to a single cluster \( f_j \); \( m_{ij} \) is therefore the intersection between \( g_i \) and \( f_j \). If cluster \( g_i \) contains the fragment \( m_{ij} \), then cluster \( f_j \) also contains the same fragment. The term \( m_{ij}^2 / g_i f_j \) provides a symmetric measure of mutual concordance between the two clusters \( g_i \) and \( f_j \). The maximum value \( m_{ij}^2 / g_i f_j = 1 \) is attained only when \( g_i = f_j \). The MoC is calculated as follows and varies between 0 (no concordance) and 1 (perfect concordance).

\[
\text{MoC} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{m_{ij}^2}{g_i f_j} - 1}{\sqrt{IJ} - 1}
\]

At last, the threshold cluster number (\( K_m \)) was chosen as the \( K \) value after which genus- and family-based bioassessments are dissimilar as suggested by the \( \psi \), NER and MoC similarity indices. The \( r_{Rao} \) was calculated with the aid of the *ade4* R package (Dray & Dufour, 2007), while \( \psi \) and NER were calculated using the Classification toolbox version 5.0 (Ballabio & Consonni, 2013) in MATLAB® (Araghinejad, 2014). The \( K \)-means was implemented with the “Ckmeans.1d.dp” R package (Wang & Song, 2011), and the MoC index was calculated with Microsoft Excel® Spreadsheet.

3 | RESULTS

3.1 | Sampling and analysis of benthic macroinvertebrates

A total of 31,588 benthic macroinvertebrate individuals were collected: 27,292 individuals were identified to the genus level. The taxonomic analysis resulted in 79 genera, 44 families and 17 super- groups (in its great majority orders). Individuals belonging to the Oligochaeta could be identified only to order level. Baetodes were the most dominant genus, accounting for 50.3% of all individuals, followed by Leptohyphes (5.6%), Gigantodax (5.3%), Camelobaetidius (4.2%), Helicopsyche (3.3%) and Limnicola (3.2%). The rest of the genera accounted for 28.1% of the total abundance.

3.2 | Comparison of taxonomic resolutions (TRs)

The analysis suggested an exceedance probability of about 82% for having higher values than 0.75, for both NSE and \( r^2 \) (Figure 3a). The evolutions of the exceedance probability curves associated with NSE and \( r^2 \) were steadily similar (Figure 3a). However, not all taxa...
presented the same high NSE and $r^2$. Elmidae had a scatter plot for high-agreement (Figure 3b), Baetidae for medium-agreement (Figure 3c) and Leptophlebiidae for low-agreement (Figure 3d) of mean NSE and $r^2$ values. The fact that in about 82% of the cases, there is an acceptable similarity (“very good” according to the scale suggested by Moriasi et al., 2007; Parajuli et al., 2009) between the two rescaled functional trait databases $\text{Db}_{g-rsc}$ and $\text{Db}_{f-rsc}$, encourages using either TRs in freshwater quality bioassessments that include, for instance, functional diversity indices.

3.3 | Calculating functional diversity

The main statistical descriptors of the rRao distributions for both TRs are almost identical (Figure 4a). Hence, no significant differences between rRao from both TRs were observed (Figure 4b). This trend was emphasized by the Kolmogorov–Smirnov test; that is, there was not a statistically significant difference between distributions of $r\text{Rao}_{g-rsc}$ and $r\text{Rao}_{f-rsc}$ at the 95.0% confidence level ($p$ value $= .08 > \alpha = 0.05$).

3.4 | Defining the WQ class number after which genus- and family-based assessments are dissimilar

3.4.1 | Data preprocessing

The aggregation process applied on the $r\text{Rao}_{g-samp}$ and $r\text{Rao}_{f-samp}$ data sets resulted in the assignment of median values at sampling sites (Figure 1) COL1, MAG1 and MAG2 for $r\text{Rao}_{g-samp}$ and B8, COL1 and MAG1 for $r\text{Rao}_{f-samp}$. For the rest of the sites, the respective mean values were assigned.

3.4.2 | Cluster analysis

Significant differences between the outputs of the cluster analysis were noticeable from $K = 4$ onwards (Table 2) as a function of TR. $K$-means outputs for $K = 2$ and $K = 3$ were very much similar in terms of both TRs, as the only difference in both cases occurred for the allocation of the MAG1 and PB2 sites for $K = 2$ and $K = 3$, respectively. This was confirmed by the values adopted by the $\psi$, NER and MoC similarity indices (Figure 5) that dropped significantly after $K_{th}=3$, characterizing as such dissimilar genus and family cluster outputs, which would imply dissimilar genus- and family-based bioassessments. Although a $K = 6$ value is generally not used in bioassessments and, as such, was not included in Figure 5, it was numerically inspected to cross-check on a consistent decrease in the similarity index values, which was indeed the case, emphasizing as such that in the current case, $K_{th} = 3$. It should be emphasized that one of the main aims of the current research was quantifying the difference between families and genera in terms of functional description. Thus, the goal was not to assess the ecological pertinence of the outcomes of the clustering algorithm used.

FIGURE 3 (a) Empirical distribution of exceedance probability of Nash and Sutcliffe coefficient (NSE) ($P(\text{NSE} \geq \text{nse})$) and $r^2$ ($P(r^2 \geq r^2)$) for the whole set of traits of family and genus; and scatter plots of mean NSE and $r^2$ calculated for the standardized traits of family and genus ($\text{Db}_{g-rsc}$ and $\text{Db}_{f-rsc}$) at every sampling site for representative (b) high-agreement, (c) medium-agreement and (d) low-agreement conditions. $X$ is either NSE or $r^2$.

FIGURE 4 (a) Main statistical properties of Rao index relative to maximum ($r\text{Rao}$) for all of the genera and families; and (b) $r\text{Rao}$ frequency histograms for genera and families.
One of the specific objectives of this study was defining and applying a functional diversity (FD) index for site characterization using families and genera. To find an accurate estimation of FD, the primary inputs (i.e. the trait information) are crucial. Some studies (Chará-Serna et al., 2012; Reynaga et al., 2020; Tomanova et al., 2006, 2007) assign trait scores at family and/or genus levels, and generate site-specific information through gut content analyses, mouthpart observations, size measurements or skeleton hardness measurements, among others. However, in this study, the generation of this site-specific data was not an option due to time and resource limitations. The alternative, considered in this research, was using published trait information. Considering the first law of geography—near things are more similar than distant ones (Tobler, 1970)—the use of trait data from distant areas may lead to considerable errors. The trait data that were used herein were obtained from reliable Latin American DBs (Chará-Serna et al., 2012; Longo & Blanco, 2014; Ramírez & Gutiérrez-Fonseca, 2014; Tomanova & Usseglio-Polatera, 2007). However, for some of our genera (12.7%), no trait data were available in these DBs. Thus, the missing data were completed with macroinvertebrate trait data from non-Neotropical areas but that also are situated in mountainous regions, in an attempt to reduce the use of non-representative trait information.

Many FD indices have been proposed aiming at measuring functional—richness, evenness and divergence—aspects that some
authors argue are the components of FD (Villégé et al., 2008) in the context of a multifaceted framework. It is also argued that this multifaceted framework of FD helps understanding how biodiversity interacts with the environment. Hence, not all of them provide the same information because not all of them are measuring the same facet; as such, there is a lack of consensus about which the most appropriate one is (Mouchet et al., 2010). Further, other authors put a lot of criticism over the FD multifaceted framework (i.e. Laliberté & Legendre, 2010). Rao’s quadratic entropy (Q) is the most frequently used index to quantify the FD of freshwater macroinvertebrate assemblages (Schmera et al., 2017) and in general functional ecology. One possible reason for that is due to the fact that Q represents a mixture between functional aspects—richness and divergence (Mouchet et al., 2010). Nonetheless, when computing Q for comparisons between data sets (i.e. under a comparative framework) with very different numbers of taxa, it is statistically biased (Casanoves et al., 2011; Chen et al., 2018) because its final value depends, among other aspects, on the scale of the dissimilarities between the compared data sets. Consequently, under these conditions of comparison, the use of Q is not adequate (Pavoine et al., 2005). Hereafter, for carrying out the current comparison between family and genus TRs, it was used the modification of the Rao index by Pavoine et al. (2005), that is Rao relative to a maximum (rRao), that is Q standardized with respect to its maximum as a statistical simple way of overcoming the referred bias pitfall. Further, the latter is in line with what has been done in other related ecohydrological studies (Koperski, 2019; Teresa et al., 2015).

The study showed that both differences in trait structure (DTS), characterized by the NSE and the $r^2$ indices, and FD, measured through the rRao index, exhibited no significant differences as a function of the TR. However, some authors argue that species-level identification is generally required when using diversity indices (Legendre & Legendre, 2012), such as the rRao index, for example when assessing freshwater pollution (Dalú et al., 2017; Guerold, 2000; Lenat & Resh, 2001). Others have defended, however, the use of diversity indices at family level (Bo et al., 2020; Corbi & Trivinho-Strixino, 2006; Ferraro & Cole, 1995; Mueller et al., 2013), which is in line with the results of this study (Figure 4). Taxa should be identified to a sufficient level to permit the detection of changes in stressed macroinvertebrate assemblages (Terlizzi et al., 2003). Current results suggested that family-level identification was detailed enough to obtain similar rRao values for either TR. Furthermore, when using cluster analysis upon the rRao values, for a number of water quality (WQ) classes $K = 2$ and $K = 3$, nearly all of the same sampling sites are in the same clusters for both TRs, emphasizing that family-level identification could be sufficient until $K = 3$.

The different K-means outputs are consistently grouping the same sites for either the best or the worse WQ conditions, independently of the considered K value and TR. This is congruent with previous findings in the Paute River Basin where sites belonging to the Burgay sub-basin (i.e. B2, B8 and B13; see Table 2) were associated with deteriorated WQ. On the contrary, the Collay sub-basin (i.e. COL1; see Table 2) was considered as a clean system (Sotomayor, 2016; Sotomayor et al., 2018, 2020).

In the study, the trait information of a given family was derived through the use of the trait data of all of the genera belonging to that family (through the averaging of $z_{Q-trao}$ for obtaining the family scores). This approach was aimed at, firstly, achieving standardized information for both TRs that would enable a direct comparison between genus-based and family-based results; and, secondly, capturing the genus, as well as feasible, variability occurring within a given family. Particularly, the latter aimed at mitigating as much as possible the limitation of using coarse taxonomic resolutions (when no finer information is available beforehand), which would normally imply larger errors if this in-family genus variability is neglected (Dalú et al., 2017; Wu, 1982) or roughly approximated by using information upon more common or dominant genera.

Dissimilarities between the K-means outputs as a function of the TRs started from $K = 4$ onwards, which was assessed through the $\psi$, NER and MoC indices (Figure 5). $\psi$ and NER are based on the confusion matrix (CM) assessment, considering the same K for either TR. This principle of using the same K for either cluster analysis is the so-called principle of distributional equivalence (Govaert & Nadif, 2018). An alternative technique for comparing the outputs of K-means as a function of the TRs is based on the mutual information (MI) approach, represented herein by the MoC index, which expresses the amount of information that the cluster outputs have in common (Pfitzner et al., 2009). Although the same K value for either TR was used in the current research, the MoC index enables the comparison of cluster outputs that are resulting from different K values. Further, the MoC index enables comparing the outputs of two different cluster algorithms applied on the same input data or, as done in the current study, comparing the outputs of the same clustering algorithm applied on different input data (i.e. different TR). The joint application of the CM and MI indices led to a congruent assessment of the $K$ value. To the best of our knowledge, it has rarely been used a combination of CM and MI indices in the context of a multiobjective bioassessment study (Ficko & Bončina, 2013).

Dealing with different TRs raises the question on the precision of correct identification of macroinvertebrates. Identification accuracy and precision are better for coarse taxonomic resolutions than for finer ones. Identification to genus or species level requires a significant investment in quality assurance, namely, a comprehensive library of macroinvertebrate taxonomic keys, a reference collection and the frequent consultation of taxonomic specialists (Jones, 2008). For example, Sondermann (2013) reported that for the Colombian elmid fauna, around 42% of the reported genera and 15% of the overall genus records turned out to be presumably erroneous. Similar findings were reported by Moulton et al. (2000) and Ochieng et al. (2019). Hence, working on family identification level assures more reliable databases, which can serve as input data to further bioassessment analyses.

This study focused on developing an alternative in the framework of FTA, which is normally applied on genus/species level, so that it could be implemented on a family level, with the potential of...
cost-reduced long-term monitoring for WQ management purposes. This research recommends the establishment of an initial period of monitoring and detailed taxonomic identification, to a species/genus level, upon which the proposed methodology could be applied to define, as in the current case, a similarity level of functional traits for finer and coarser TRs. The methodology also identifies a threshold level of WQ class discretization, until which working with coarser TRs is acceptable. By doing these, the proposed methodology aims at using in the future coarser TRs, with the consequent reduction in time and monetary expenses, as well as the need of high-level taxonomic expertise. The length of this initial period when a detailed TR is used would depend on several factors, among which the objectives of the study beforehand, the size of the study site and the frequency of monitoring are very influential. These factors are likely to affect as well the similarity level of functional traits for finer and coarser TRs that could be reached in the local application of the proposed methodology. Further, it is up to the decision-makers in charge of WQ management to decide on whether the reached level of functional trait data similarity is sufficient for WQ management purposes and whether the defined threshold level of WQ class discretization is acceptable.

5 | CONCLUSIONS

The functional trait approach (FTA) was applied as a way of studying the taxonomic resolution (TR) of benthic macroinvertebrates for freshwater assessments. For achieving this purpose, the differences between families and genera were quantified in terms of functional description, which led to the conclusion that in the study site a functional trait data similarity of about 82% exists for both TRs. Site characterization was carried out by using a functional diversity index (rRao) for families and genera with no significant differences between both TRs. Sampling site clustering on the basis of functional diversity was compared using the K-means algorithm for both TRs, based on two different frameworks, namely, the confusion matrix (CM) and the mutual information (MI) assessments. These frameworks led to a congruent assessment of the K value (K10 = 3 classes) until which family-level identification in the study basin was suitable in detecting changes of macroinvertebrate assemblages. The study proposes an alternative FTA methodology for WQ management purposes so that coarser TRs could be used in the future with the consequent reduction in monitoring time and costs and the need of high-level taxonomic expertise.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the former Ecuadorian National Secretariat of Water (SENAGUA)—DHS for making the raw data accessible to the current study and to Pablo E. Gutiérrez-Fonseca, Fernando Casanoves and Julio Di Rienzo for their helpful suggestions. This research was carried out in the context of the research Project “Efectos del estrés hidrónico sobre la biodiversidad en ríos abastecedores de la ciudad de Cuenca,” financed by the Research Directorate of the Universidad de Cuenca (DIUC).

CONFLICT OF INTEREST

The authors declare that they have not known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The database belongs to the Ecuadorian government, concretely to the former National Secretary of Water (SENAGUA), which was recently absorbed by the Environmental, Water and Ecological Transition Ministry (MAATE). To use this database, the first author subscribed with the Ecuadorian government a specific agreement that entitles him using the data for his PhD research. This signed agreement explicitly forbids him sharing the information with a third party without the specific authorisation of the government, which definitively might be almost impossible to get in a reasonable mid-term. Thus, the co-authors would prefer not to publicly share this database.

ORCID

Gonzalo Sotomayor https://orcid.org/0000-0003-3743-4526
Henrietta Hampel https://orcid.org/0000-0003-2249-5369
Raúl F. Vázquez https://orcid.org/0000-0003-2581-5372
Marie Anne Euri Forio https://orcid.org/0000-0001-6675-4751
Peter L. M. Goethals https://orcid.org/0000-0003-1168-6776

REFERENCES

Anderson, T., & Finn, J. (1996). The new statistical analysis of data. Springer-Verlag New York Inc.
Araghinejad, S. (2014). Data-driven modeling: Using MATLAB® in water resources and environmental engineering (Vol. 67). Springer.
Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2001). Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. Journal of the North American Benthological Society, 20(2), 280–286. https://doi.org/10.2307/1468322
Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. Analytical Methods, 5(16), 3790–3798.
Ballabio, D., Grisoni, F., & Todeschini, R. (2018). Multivariate comparison of classification performance measures. Chemometrics and Intelligent Laboratory Systems, 174, 33–44. https://doi.org/10.1016/j.chemolab.2017.12.004
Bêche, L. A., & Resh, V. H. (2007). Biological traits of benthic macroinvertebrates in California Mediterranean-climate streams: Long-term annual variability and trait diversity patterns. Fundamental and Applied Limnology, 169(1), 1–23. https://doi.org/10.1127/1863-9135/2007/0169-0001
Bellucci, C. J., Becker, M. E., Beauchene, M., & Dunbar, L. (2013). Classifying the health of connecticut streams using benthic macroinvertebrates with implications for water management. Environmental Management, 51(6), 1274–1283. https://doi.org/10.1007/s00267-013-0033-9
Biggerstaff, B. J. (2000). Comparing diagnostic tests: A simple graphic using likelihood ratios. Statistics in Medicine, 19(5), 649–663.
Bo, T., Doretto, A., Levrino, M., & Fenoglio, S. (2020). Contribution of beta diversity in shaping stream macroinvertebrate communities among hydro-ecoregions. Aquatic Ecology, 54(4), 957–971. https://doi.org/10.1007/s10452-020-09786-6
Tomanova, S., & Usseglio-Polatera, P. (2007). Patterns of benthic community traits in neotropical streams: Relationship to mesoscale spatial variability. *Fundamental and Applied Limnology, 170*(3), 243–255. https://doi.org/10.1127/1863-9135/2007/0170-0243

Tomanova, S., Goitia, E., & Helešic, J. (2006). Trophic levels and functional feeding groups of macroinvertebrates in neotropical streams. *Hydrobiologia, 556*(1), 251–264. https://doi.org/10.1007/s10750-005-1255-5

Tomanova, S., & Usseglio-Polatera, P. (2007). Patterns of benthic community traits in neotropical streams: Relationship to mesoscale spatial variability. *Fundamental and Applied Limnology, 170*(3), 243–255. https://doi.org/10.1127/1863-9135/2007/0170-0243