Tackling discrepancies in freshwater invertebrate trait databases: Harmonising across continents and aggregating taxonomic resolution

Stefan Kunz[1], Ben J. Kefford[2], Leon Metzeling[3], Astrid Schmidt-Kloiber[4], Wolfram Graf[4], Philippe Usseglio-Polatera[5], Christoph D. Matthaei[6], Charles P. Hawkins[7], N. LeRoy Poff[8], Laura Twardochleb[9], and Ralf B. Schäfer[1]

[1]Institute for Environmental Sciences, University of Koblenz-Landau, Landau, Germany
[2]Centre for Applied Water Science, Institute for Applied Ecology, University of Canberra, Canberra, Australia
[3]Environment Protection Authority Victoria, Applied Sciences Division, Macleod, Australia
[4]Institute of Hydrobiology and Aquatic Ecosystem Management, University of Natural Resources and Life Sciences Vienna (BOKU), Vienna, Austria
[5]University of Lorraine, CNRS, LIEC, Metz, France
[6]Department of Zoology, University of Otago, Dunedin, New Zealand
[7]Department of Watershed Sciences and the Ecology Center, Utah State University, Logan, USA
[8]Department of Biology, Colorado State University, Fort Collins, USA
[9]Department of Fisheries and Wildlife, Michigan State University, East Lansing, USA

Abstract

1. Use of invertebrate traits rather than species composition may facilitate large-scale comparisons of community structure and responses to disturbance in freshwater ecology because the same traits can potentially occur everywhere. In recent years, comprehensive invertebrate trait databases have been established at different scales (e.g. regions, continents). The wide availability of invertebrate trait data supports trait-based studies, especially at large scales. However, a number of data-related issues complicate the use of invertebrate traits for ecological studies. For example, standardised definitions for for freshwater invertebrate traits are lacking, which impedes comparisons across regions. Moreover, it is uncertain how harmonising varying trait definitions among databases might influence identification of trait-environment relationships. In addition, taxonomic resolution between observational taxonomic data and trait data can differ, making aggregation of traits necessary. At present, it is unknown how different trait aggregation approaches compare with expert-assigned trait affinities. In this study we aimed to identify discrepancies in trait definitions across freshwater invertebrate databases and to compare trait aggregation methods to identify how taxonomic hierarchy structure and trait variability influence results from different aggregation approaches.

2. We describe discrepancies in the definitions of traits used to create freshwater invertebrate trait databases in Europe, North America, New Zealand, and Australia. Based on our comparisons of these trait databases, we established four novel trait datasets by harmonising trait definitions of commonly used traits. Next, we used two of these datasets to compare aggregated traits obtained by different aggregation methods with traits assigned by experts, both at the family level. The trait aggregation methods we compared used either the mean or the median and different weightings. We further explored the effects of harmonisation and trait aggregation by re-analysing data from a case study.

3. We found that among trait databases, trait definitions often differed because varying numbers of traits were used to describe groups of related traits (e.g., respiration traits) and the focus of descriptions for groups of related traits also varied (e.g., for feeding mode some databases focused on the food source, whereas others focused on mouthpart morphology). The coding to describe traits (binary, fuzzy) also varied among databases. Our comparison of different aggregation methods showed that family-level aggregated and expert-assigned traits were similar, especially from approaches using the median, and that trait coding method, uneven taxonomic hierarchical structure, and high trait variability can influence aggregated trait affinities. We further showed that harmonised and aggregated data performed similarly to observational data in identifying trait–environment relationships.

4. By outlining discrepancies in trait definitions we hope to motivate the development of standardised terminology for invertebrate traits. Our results also illustrate the usefulness of harmonised datasets for ecological study and provide guidance for the circumstances under which the choice of trait aggregation method is important.

Introduction

Explaining and predicting how communities are shaped by environmental factors is a primary main goal of ecology. Species traits are measurable properties of an organism (McGill et al., 2006), and comparing communities based on the traits possessed by species may facilitate testing of a range of ecological hypotheses (Statzner et al., 2001). Traits are assumed to be adaptations (e.g., physiological, behavioural) of organisms to their environment and to indicate direct or indirect linkages between the biological response of an organism or a population and its environment (Southwood, 1977; Verberk et al., 2013). In addition to providing local-scale mechanistic-based expressions of species-environment relationships, trait-based approaches might be suitable for large-scale analyses because trait responses are less constrained by biogeographic boundaries and taxon distributional areas than taxonomic responses (Baird et al., 2011; Bonada et al., 2007).

Invertebrate traits have been increasingly used in freshwater ecology, e.g. by relating macroinvertebrate trait composition to environmental factors (Bhowmik and Schäfer, 2015; Poff et al., 2010; Szöcs et al., 2014). In the last decades, freshwater ecologists have compiled comprehensive invertebrate trait databases, which typically include data for a single region or continent (Kefford et al., 2020; Philips and Smith, 2018; Schmidt-Kloiber and Hering, 2015; Tomanova et al., 2006; Usseglio-Polatera et al., 2000; Vieira et al., 2006). The availability of invertebrate trait data from different continents enables comparisons of regional trait variation and allows for testing the consistency of trait structure–environmental factor relationships across both small and large spatial extents. To date, such analyses have been carried out mostly within continents, using information from one or two trait databases. For example, Bonada et al. (2007) compared trait composition for Mediterranean and temperate regions in Europe based on traits from Usseglio-Polatera et al. (2000) (typically referred to as the Tachet database), Poff et al. (2010) characterized trait composition across sites in the Western USA based on traits from Poff et al. (2006), and Botwe et al. (2018) used trait definitions from Poff et al. (2006) and Schäfer et al. (2011) to test for effects of salinity on invertebrate traits across different sites in South Australia. Analyses that synthesize invertebrate trait information from more than two different continents are rare, but see Brown et al. (2018) and Statzner et al. (1997).

The heterogeneity of information in freshwater invertebrate trait databases, besides the diversity of taxa across regions, is likely a major reason for the lack of studies across continents. There are several challenges presented by using invertebrate trait data from multiple established databases in ecological studies. One such challenge is inconsistency in terminology. In this study, we follow the terminology proposed by Schmera et al. (2015), where a grouping feature is defined as a general property (e.g. feeding mode) that comprises a "group of related traits (e.g., predator, shredder, etc.) that vary among species or among individuals within a species". Thus, we use the term grouping feature in place of the term trait, and the term trait instead of trait state, modality or trait category.

Another challenge to using data from multiple continents is that invertebrate trait databases include different grouping features and related traits. To harmonize grouping features from different regions, or to structure them in such a way as to facilitate cross-database comparison, commonly accepted and unambiguous trait definitions are required (Schneider et al., 2019). Ideally, the same traits and grouping features would be reported across databases or standardised terminology would exist to easily harmonise them; however, there is a lack of standardised terminology of trait definitions and poor metadata quality in many trait databases, making grouping-feature harmonisation difficult (Baird et al., 2011; Kissling et al., 2018). To our knowledge, only Brown et al. (2018) har-

monised grouping features from more than two geographically distant invertebrate trait databases, for a limited set of grouping features (8) and taxa (112), in a study on the influence of decreasing glacier cover on functional diversity and invertebrate assemblages.

A third challenge to comparing across invertebrate trait databases is inconsistent coding of trait data (Culp et al., 2011), or assigning affinities (aka affinity scores). Traits of individual freshwater invertebrates can be difficult to quantify because, unlike plant traits, they are often difficult to measure directly. For example, to describe feeding habits requires that we understand mouthpart morphology, feeding behaviour, and specific food sources (Moog and Hartmann, 1995). Some traits can be described as binary or categorical, which ignores uncertainty in how traits are expressed in any particular organism (e.g., adult terrestrial stage, presence of gills). Other traits are better suited to being described as continuous (e.g., tolerance of pollution, body size). One approach for dealing with uncertainty is the use of fuzzy coding, where traits are assigned probabilistic values. Fuzzy codes, which are usually converted to proportions, are used to account for plasticity in traits, variability in traits within taxonomic groups above the species level, and incomplete knowledge. A challenge is presented by the use of different types of coding for trait affinities in different databases. For example, in the study from Brown et al. (2018) the authors needed to reclassify trait affinities because the individual databases employed different coding approaches (i.e., European and New Zealand databases used fuzzy coding, North American database used binary coding).

Finally, discrepancies in the taxonomic resolutions (e.g., species, genus, or family) when comparing among trait databases, or when linking observational taxonomic data to trait databases, presents another challenge. When observations or data in one trait database are at a more precise taxonomic level than other data (e.g. observations at species-level and trait data at genus-level) trait information of the less precise taxonomic level is often assigned (e.g. Szöcs et al., 2014; Voß and Schäfer, 2017). Conversely, if trait information is only available at more precise taxonomic levels than observed taxa, traits are aggregated to a less precise taxonomic level (e.g. Aspin et al., 2019; Pilière et al., 2016; Poff et al., 2006; Szöcs et al., 2014). To date, studies have used different methods of trait aggregation, e.g. the mean (Magliozzi et al., 2019), median (Szöcs et al., 2014) or mode (Pilière et al., 2016), but studies on how and to what extent different trait aggregation methods influence trait-based analyses are missing.

These challenges when working with or synthesizing trait data motivated us to assess the characteristics and comparability of freshwater invertebrate trait data contained in disparate databases. We aimed to explore methods of improving the comparability of trait data across databases, despite differences in trait definitions and the taxonomic level of data in three ways: (1) by investigating and describing discrepancies in trait definitions between trait databases from Europe, North America, Australia, and New Zealand; (2) by harmonising trait definitions across databases to provide consistent trait grouping features for comparison across geographic locations; and (3) by comparing the influence of different trait aggregation methods and grouping-feature harmonisation on the identification of trait–environment relationships.

Methods

To address our study objectives, we took a multi-prong approach to describing, combining, and assessing multiple sources of freshwater invertebrate data that differed in grouping-feature definitions and taxonomic resolution. Identified trait definitions discrepancies served as a starting point to develop four novel invertebrate trait datasets by harmonising seven grouping features from existing trait databases. The harmonised trait datasets were used to evaluate the influence of different trait aggregation approaches on the identification of trait–environment relationships, our third study objective, in two ways by (1) comparing family-level aggregated trait affinities with family-level trait affinities assigned by experts for two of the established trait datasets where such data were available and (2) comparing the influence of taxonomic hierarchy and trait variability on the outcomes of the different trait aggregation approaches for simulated data. Finally, to further investigate how harmonising and aggregating trait data can modify the outcomes of trait–environment relationship characterisation, we used data from one of the newly established datasets with harmonised and aggregated traits to repeat an analysis of salinisation effects on biological traits (Szöcs et al., 2014) and compared our results with those of the original study.

Selection of traits and harmonisation of trait databases

To explore the characteristics and comparability of freshwater invertebrate trait data with differing grouping-feature definitions and taxonomic resolution, we obtained invertebrate trait data from six trait databases for four regions: Europe, North America, Australia, and New Zealand. For Europe we obtained trait information from the freshwaterecology.info database (Schmidt-Kloiber and Hering, 2015), which contains taxa at the species level, and the Tachet database (Usseglio-Polatera et al., 2000), which contains taxa at species, genus, and family levels. For North America we primarily obtained trait information from Twardochleb et al. (2021) (hereafter referred to as the CONUS database) and where taxa were missing from that database, we obtained data from Vieira et al. (2006) (hereafter referred to as the Vieira database). Both databases contain taxa at species, genus and family levels. Data on body form for European and North American taxa were based on expert knowledge (Usseglio-Polatera, 2020a). For Australia and New Zealand, we used trait databases from Kefford et al. (2020) (hereafter referred to as the Australian database) and Philips and Smith (2018) (hereafter referred to as the New Zealand database) respectively. These databases contain taxa at species, genus, family, and less precise taxonomic levels.

We established four datasets, one for each of the four geographic regions, by harmonising the traits of seven grouping features from the trait databases. We selected traits that were available in all databases, are commonly applied in trait-based ecological studies, and describe different parts of the biology of a species: life history (Voltinism), morphology (Respiration, Body form, Size), ecology (Locomotion, Feeding mode) and reproduction (Oviposition). We omitted ecological traits that describe habitat preferences (e.g. temperature preference) because these traits are missing in the New Zealand trait database. Grouping features were classified differently across the databases, so we harmonised them into 26 traits (Table 1). To do so we combined similar traits into one trait (e.g. crawlers and sprawlers into crawlers) and assigned the highest trait affinity score among the combined traits for each taxon.

Next, we selected taxa for inclusion in our datasets and normalised the trait affinity scores for each taxon. We first omitted all taxa with a lower taxonomic precision than family level and consolidated duplicate taxa that appeared within a database, by either applying the median for

fuzzy coded traits, or the maximum for binary traits. We then established our harmonised datasets by using either fuzzy coding or, when unavailable, binary coding. Fuzzy codes are reported with different ranges in the trait databases (e.g., freshwaterecology.info: 0–10, Tachet database: 0–3 or 0–5). We normalised these to a range between 0 and 1 and converted trait affinities to proportions. In cases where binary coding was needed, we converted categorical and continuous traits into binary traits with values of 1 and 0, which, thus, corresponded with the range of fuzzy coded traits. For example, in the freshwaterecology.info database, the classification of the grouping feature voltinism accounts for different faunistic regions and we substituted a value of 1 for entries such as "arctic" or "boreal". We assumed that values of 1 and 0 for binary traits corresponded to the highest and no affinity for a particular trait.

Table 1: Traits of harmonised grouping features from six invertebrate trait databases and four geographic regions. The last column indicates traits that were combined for harmonisation (no combining needed if empty).

Grouping feature	Trait	Combined traits
Voltinism	Semivoltine	<1 generation per year
	Univoltine	1 generation per year
	Bi/multivoltine	>1 generation per year
Body Form	Cylindrical	Cylindrical, tubular
	Flattenend	Flattenend, dorsoventrally flattened[†]
	Spherical	Spherical, round (humped)
	Streamlined	Streamlined, fusiform
Size	Small	<9 mm, <10 mm[‡]
	Medium	9 - 16 mm, 10 - 20 mm
	Large	>16 mm, >20 mm
Respiration	Gills	Tracheal gills, gills
	Plastron/Spiracle	Temporary air store, spiracular gills, atmospheric breathers, plant breathers, functional spiracles, air (plants), aerial, plastron/spiracle
	Tegument	Cutaneous, tegument
Locomotion	Burrower	Interstitial, boring, burrowing
	Crawler	Sprawler, walking, climber, clinger, crawler
	Sessile	Attached, sessile
	Swimmer	Skating, diving, planctonic, swimming
Feeding mode	Filterer	Active/passive filterer, absorber, filter-feeder, collector-filterer, filterer
	Gatherer	Deposit-feeder, collector-gatherer, detrivore, gatherer
	Herbivore	Grazer, scraper, piercer herbivore, herbivore, algal piercer, piercer (plants)[§]
	Parasite	
	Predator	Piercer (animals)[§], predator
	Shredder	Miner, xylophagus, shredder, shredder detrivore
Oviposition	Aquatic eggs	Eggs attached to substrate/plants/stones, free/fixed eggs/clutches
	Ovoviviparity	
	Terrestrial eggs	Terrestrial clutches, terrestrial

† The trait "bluff (blocky)" occurred in the Vieira database and was newly classified by expert knowledge into cylindrical and flattened (Usseglio-Polatera, 2020a).

‡ Reflects the different size classifications by the Vieira and CONUS databases from the other trait databases.

§ The trait piercer was defined in the Tachet database for piercing plants and animals, in contrast to the other databases (Usseglio-Polatera et al., 2000). Taxa exhibiting this trait have been assigned to predators or herbivores based on expert knowledge (Usseglio-Polatera, 2020b).

Comparing trait aggregation methods

For our study aim of comparing the effects of different aggregation methods on trait–environment relationship inference, we aggregated the traits of two of the harmonised datasets to the family level. We used three different aggregation approaches: (1) direct aggregation of taxa to family level with either the median or mean, giving equal weight to every taxon. We denote these aggregation methods as $direct\_agg_{median}$ and $direct\_agg_{mean}$, respectively; (2) stepwise aggregation, first to the genus level and subsequently to the family level with either the median or mean. This approach gives equal weights to each genus. Hereafter, we denote this aggregation type as $stepwise\_agg_{median}$ or $stepwise\_agg_{mean}$, respectively; and (3) aggregation using a weighted mean approach, denoted as $weighted\_agg$. This method weights each genus according to its number of species in the trait datasets, regardless of whether information was available for every used grouping feature (Figure 1). When we refer to $direct\_agg_{mean}$, $stepwise\_agg_{mean}$, and $weighted\_agg$ together we denote these methods as mean aggregation methods and when we refer to $direct\_agg_{median}$ and $stepwise\_agg_{median}$ together, we denote these methods as median aggregation methods.

Comparison of family-level aggregated traits with family-level assigned traits

To evaluate the effect of the five trait aggregation methods ($direct\_agg_{median}$, $direct\_agg_{mean}$, $stepwise\_agg_{median}$, $stepwise\_agg_{mean}$, $weighted\_agg$) on the assignment of trait affinity scores, we compared our aggregated affinities with trait affinities assigned at family level by experts. Expert-assigned trait data were available for a subset of grouping features and taxa for Australia (Chessman, 2018) and North America (CONUS database), so we conducted this comparison only for these two datasets. For the Australian dataset, we used the grouping features feeding mode (all traits listed in Table 1 except parasite) and size (220 families ∗ 8 family–trait combinations = 1760 comparisons). For the North American dataset, we used the grouping features feeding mode, respiration, size, voltinism, and locomotion (94 families ∗ 17 family–trait combinations = 1598 comparisons). Assigned traits from North America were on a categorical scale, and we converted them to binary prior to comparing them with aggregated trait affinities. Size was assigned as continuous variable for Australia, thus, we converted size into three traits (small, medium, and large) with a binary coding. Because trait affinities ranged from 0 to 1, the maximum difference possible in trait affinity was 1 to –1, either which corresponds to 100 %. For convenience and to improve interpretation, we calculated absolute trait differences.

8

Effect of taxonomic hierarchy and trait variability on aggregation outcomes

To examine the influence of taxonomic hierarchy and trait variability on the outcomes of the different trait aggregation methods, we created three hypothetical scenarios. We simulated three different families, each composed of 25 total species but with different taxonomic hierarchical structures. The three families consisted of (1) a family with an equal number of genera and species (five genera each with five species each), denoted as *sim_base*; (2) a family in which one genus had a much larger number of species than the other four genera (1 genus with 13 species, 4 genera with 3 species each), denoted as *sim_extreme*; (3) a family in which all genera had a different number of species (8, 2, 7, 3, 5), denoted as *sim_variation*. We assigned a hypothetical grouping feature with three traits (T1, T2, and T3) to each scenario. We then simulated the 25 affinities, one per species, for each trait by sampling from a truncated normal distribution bound by 0 and 1 and with a mean value of 0.5. To simulate different levels of trait variability, we repeated the sampling 100 times for each of the 5 levels of standard deviations (0.2, 0.4, 0.6, 0.8, and 1), resulting in 12,500 simulated trait affinities for each simulated trait (25 species per family $*$ 5 levels of trait variability $*$ 100 replicates). We converted simulated trait affinities to proportions, as was done during trait database processing, and assigned the 12,500 simulated affinities per trait to each of the three family scenarios. We then applied the five trait aggregation methods described above to each simulated dataset. We compared the resulting ranges of aggregated trait affinities between levels of trait variability and taxonomic hierarchical scenario as well as the differences in trait affinities obtained by each aggregation method.

Data preprocessing

Consolidating duplicate taxa in EU, NOA, AUS & NZ databases → AUS traits standardised to [0 - 1] because multiple codings used throughout this database → Omitted taxa less precise than family-level

Harmonisation → Traits converted to proportions and normalised to [0 - 1] → Trait aggregation

Family	Genus	Species	Trait A_h$_1$	Trait A_h$_2$	Trait A_h$_3$
F_1	G_1	S_1	0.25	0.75	0
F_1	G_1	S_2	0.5	0.5	0
F_1	G_1	S_3	0.5	0.5	0
F_1	G_2	S_4	0	1.0	0
F_1	G_2	*NA*	0.25	0.75	0
F_1	*NA*	*NA*	0.25	0.75	0

Direct aggregation

Family	Trait	Method	Value
F_1	A_h$_1$	mean	0.29
F_1	A_h$_2$	mean	0.71
F_1	A_h$_3$	mean	0
F_1	A_h$_1$	median	0.25
F_1	A_h$_2$	median	0.75
F_1	A_h$_3$	median	0

Stepwise aggregation

Family	Genus	Trait	Method	Value
F_1	G_1	A_h$_1$	mean	0.4167
F_1	G_2	A_h$_1$	mean	0.125
F_1	NA	A_h$_1$	-	0.25

$$A\_h_1 = mean(0.4167, 0.125, 0.25) = 0.26$$

Weighted aggregation

Weights based on species per genera: 3 for G_1, 1 for G_2, weight for F_1 set to 1

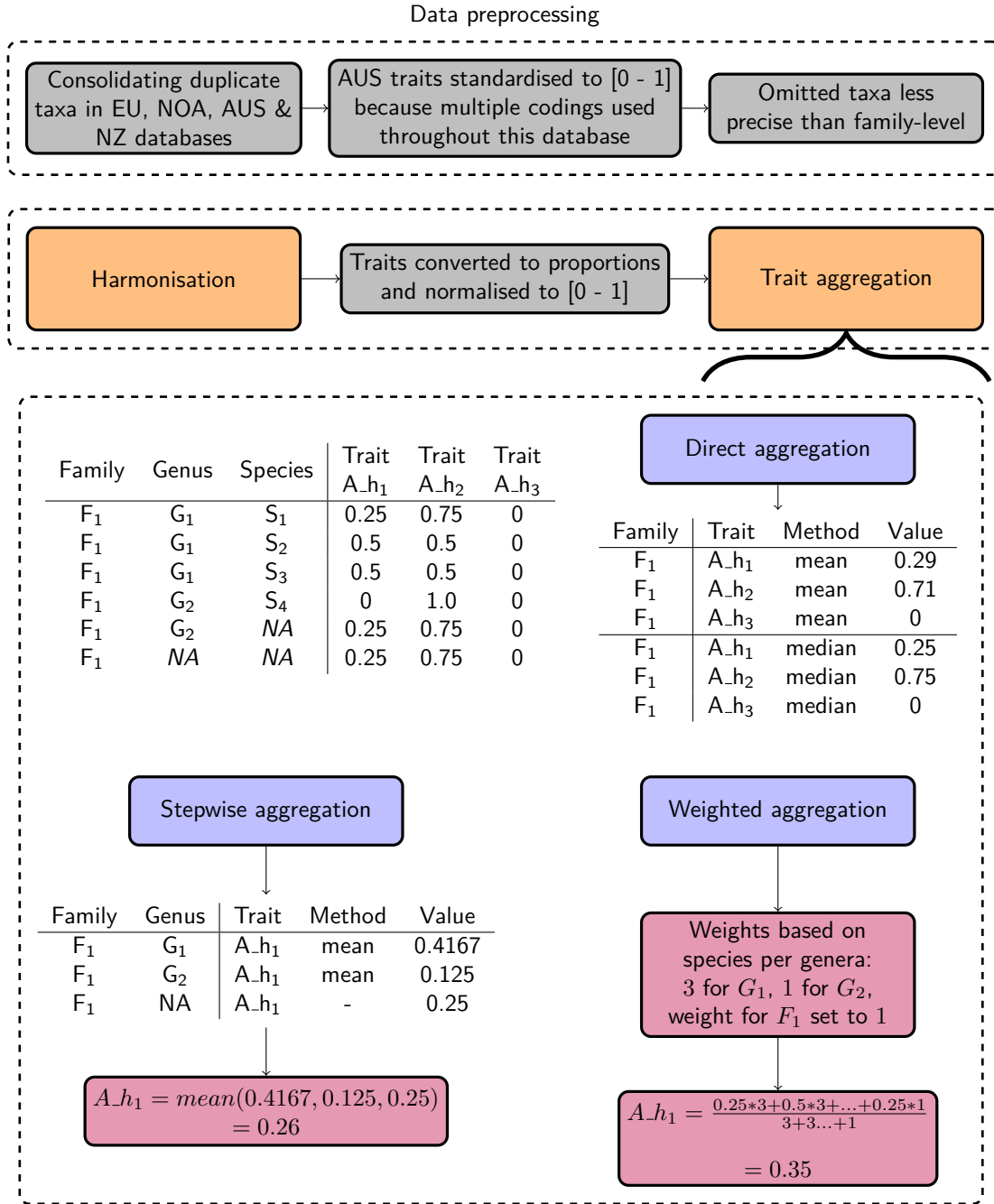$$A\_h_1 = \frac{0.25*3+0.5*3+...+0.25*1}{3+3...+1} = 0.35$$

Figure 1: Data processing steps of the selected traits. Intermediate (grey) and main (orange) steps of data preparation are depicted. The dashed bottom box illustrates the different trait aggregation methods (hypothetical data in the upper left corner). The aggregation methods (purple) and intermediate steps of the aggregation methods (pink) are displayed. Abbreviations: EU, Europe; NOA, North America; AUS, Australia; NZ, New Zealand.

Effects of harmonisation and trait aggregation on trait-environment relationship inferences

To further evaluate the effects of grouping-feature harmonisation and trait aggregation on characterisation of trait–environment relationships, we repeated the analysis of Szöcs et al. (2014), who studied the effects of anthropogenic salinisation on invertebrates in the River Werra in Germany. The River Werra has been subject to effluents from the potash industry since the mid- 20th century, which makes it a useful system for studying the responses of invertebrates and their trait compositions to salinisation (Bäthe and Coring, 2011). Szöcs et al. (2014) used redundancy analysis (RDA) to compare trait compositions among sites downstream, upstream, and close to salt discharge (transition). Further details about the original study can be found in Szöcs et al. (2014).

We re-analysed the Szöcs et al. (2014) study by performing a RDA with harmonised grouping features from the dataset for Europe and with traits that we aggregated to family level with the five aggregation methods described above. We used all 21 grouping features from the original study, but replaced 6 with harmonised grouping features from our European trait dataset, namely: size, feeding mode, locomotion, oviposition, respiration, and voltinism. We compared the results of the original study's RDA with the results of the RDA from the harmonised grouping features. Specifically, the trait composition, expressed as community weighted mean (CWM) traits, was ordinated along an electric conductivity gradient. We compared the species scores obtained from the RDA (i.e., the coordinates of the tips of the vectors representing the CWM traits in the bi- or triplots). Following the original study, we used the Mahalanobis distance measure and identified traits associated with higher or lower salinity based on their distance to the ordination axis median. We considered CWM traits with a Mahalanobis distance greater than the 97.5 %-quantile of the Chi-square distribution (5.02) to have responded to either lower or higher salinity. Additionally, for testing the effect of different aggregation methods on inference about CWM trait–salinisation relationships, we repeated the RDA five times, each with family-level aggregated trait values assigned to corresponding taxa from Szöcs et al. (2014).

Data analysis

All data processing and RDA analyses were carried out using R (version 3.6.1; R Project for Statistical Computing, Vienna, Austria, 2020). RDA was computed using the vegan package (Oksanen et al., 2019). Raw data and the R code for data processing and grouping feature harmonisation are located in the Github repository: https://github.com/KunzstLD/Invertebrate_traits. Scripts and data to reproduce the trait aggregation, comparison of aggregated and assigned traits, and the RDA analysis are located in the Github repository https://github.com/KunzstLD/Trait-aggregation.

Results

Harmonised trait datasets

Taxonomic coverage of the harmonised trait datasets

The harmonised European, North American, Australian, and New Zealand trait datasets differed in terms of their taxonomic coverage. The European trait dataset had the largest taxon pool with 4601 taxa (Table 2) followed by the North American dataset (3753 taxa), the Australian dataset (1402 taxa), and the New Zealand dataset with the smallest taxon pool (478 taxa). The European, New Zealand, and North American datasets contained most taxa at the species level, whereas the Australian dataset comprised a similar number of taxa at species and genus level.

Table 2: Number (Nr.) of taxa per harmonised dataset and per taxonomic level. Numbers in parenthesis show rounded relative frequencies in percent.

Dataset	Taxa (Nr.)	Aquatic insects (Nr.)	Species	Genus	Family
EU	4601	3942 (86)	3739 (81)	704 (15)	158 (3)
NOA	3753	3305 (88)	2414 (64)	1163 (31)	176 (5)
AUS	1402	1016 (72)	564 (40)	578 (41)	260 (19)
NZ	478	443 (93)	404 (85)	47 (10)	27 (6)

Abbreviations: EU, European; NOA, North American; AUS, Australian; NZ, New Zealand.

Completeness of trait information

The numbers of entries with available information for the selected grouping features varied considerably within the harmonised European, North American, and Australian datasets (Table 3). By contrast, the New Zealand dataset contained complete trait information for most of the investigated grouping features (94–100%).

The amount of entries with available information for the selected grouping features varied strongly for the developed European, North American, and Australian datasets (Table 3). By contrast, the New Zealand dataset contained complete trait information for most of the investigated grouping features (between 94 % and 100 %).

Table 3: Rounded percentage of entries that include information for the individual grouping features shown per trait dataset.

Dataset	Body form	Oviposition	Voltinism	Locomotion	Size	Respiration	Feeding mode
EU	8	15	23	36	11	57	76
NOA	28	13	47	52	73	44	63
AUS	4	46	49	39	75	68	99
NZ	100	94	100	99	100	100	99

Abbreviations: EU, European; NOA, North American; AUS, Australian; NZ, New Zealand.

Discrepancies in invertebrate trait definitions across databases

Definitions of grouping features and traits varied in their level of detail in the investigated trait databases. The freshwaterecology.info, Tachet, and CONUS databases provided more detailed descriptions of trait information than the Vieira and New Zealand databases provided. The Australian trait database (Kefford et al., 2020) was a collection of seven specific trait datasets; thus, grouping features occurred multiple times with varying differentiation into traits. Depending on the dataset, trait information was described with more or less detail.

The definitions of grouping features also varied across databases in their differentiation into traits and in their focus, both which can lead to discrepancies in trait definitions. We provide a summary of discrepancies in trait definitions in the appendix (Table S1). Varying levels of differentiation among the trait databases were present in all investigated grouping features (Table 4 and Table S1).

For example, for the grouping feature feeding mode, discrepancies arose because traits were assigned in different ways. The Tachet database defined predators as carvers, engulfers, and swallowers. By contrast, the CONUS database defined predators as engulfers and carnivorous piercers. In turn, the Tachet database defined piercers as a separate trait encompassing both herbivorous and carnivorous piercers. Furthermore, the focus in the freshwaterecology.info database for feeding mode was primarily on the food source of a species (except for filterers), whereas the other databases focused on the strategies of food acquisition. For example, the freshwaterecology.info database defined predator as "eating from prey", and the other databases used the mouthpart morphology as basis for their definitions. The Tachet database captured the food source in an additional grouping feature. Locomotion definitions also differed in focus among databases. The New Zealand and freshwaterecology.info databases described locomotion as an organism's way of movement, Tachet as substrate relation and locomotion, Vieira as how organisms deal with flow, Australia as attachment, and CONUS included not only the method of movement but also the location of movement. Similarly, databases differed in their focus for reproduction traits. In the freshwaterecology.info and Tachet databases, reproduction was captured in one grouping feature and was defined as location of oviposit clutches and mode of reproduction. The Vieira database provided information on the oviposition location but not on reproductive behaviour. The Australian database reported grouping features for reproductive behaviour but also on oviposition location separately. The New Zealand database distinguished three grouping features related to reproduction: reproductive technique, oviposition location (e.g. water surface, terrestrial), and egg/egg mass (e.g. free, cemented).

Trait affinity scores also varied among the databases and were described by varying coding methods (e.g., binary, fuzzy, continuous). In the freshwaterecology.info and Australian databases a combination of coding methods is used, whereas in the Tachet and New Zealand databases exclusively fuzzy coding is used. The Vieira and CONUS databases contained categorical grouping features that can be converted into binary trait affinities (Table 4).

Table 4: Number of traits per grouping feature and type of coding of the traits for the grouping features used in this study per database. Oviposition location was used for the New Zealand database.

Grouping feature	freshwater-ecology.info	Tachet	CONUS	Vieira	Australia	New Zealand
Feeding Mode	10 traits; 10 point assginment system	7 traits; fuzzy [0 - 3]	6 traits; binary	8 traits; binary	16 traits†; binary, proportional [0 - 1], fuzzy [0 - 3]	6 traits; fuzzy [0 - 3]
Voltinism	6 traits; single category assignment system	3 traits; fuzzy [0 - 3]	3 traits; binary	3 traits; binary	7 traits; binary, proportional [0 - 1], fuzzy [0 - 3]	3 traits; fuzzy [0 - 3]
Locomotion	6 traits; 10 point assignment system	8 traits; fuzzy [0 - 5]	10 traits; binary	9 traits; binary	9 traits; binary, fuzzy [0 - 3]	4 traits; fuzzy [0 - 3]
Respiration	7 traits; binary	5 traits; fuzzy [0 - 3]	3 traits; binary	8 traits; binary	10 traits; binary, proportional [0 - 1], fuzzy [0 - 3]	4 traits; fuzzy [0 - 3]
Reproduction/ Oviposition	9 traits; binary	8 traits; fuzzy [0 - 3]	10 traits; binary	10 traits; binary	13 traits‡; binary	4 traits; fuzzy [0 - 3]
Size	-	7 traits; fuzzy [0 - 3]	3 traits; binary	3 traits; binary	9 traits; binary, continuous, fuzzy [0 - 3]	5 traits; fuzzy [0 - 3]
Body Form	-	-	-	4 traits; binary	4 traits; fuzzy [0 - 3]	4 traits; fuzzy [0 - 3]

† Some of the feeding mode traits used in the Australian database were similar (e.g. trait *Shredder*, *Shredder*, *Detrivore*, and *Collector*, *Shredder*).
‡ Not all traits were considered because trait information was partly presented as comments to describe other traits or due to incomplete information.

Comparing aggregation methods

Comparison of family-level aggregated traits with family-level assigned traits

The comparison of family-level aggregated trait affinities with those assigned by experts for the Australian and North American databases showed that aggregation method (median or mean) had a greater effect than the overall approach (direct, stepwise, or weighted by the number of species) in whether trait affinities differed from those assigned by experts.

The percentage of differing cases of aggregated versus assigned trait affinities varied by method between 16.2 and 22.9 % for the Australian dataset and between 15.3 and 47 % for the North American dataset (Table 5). In general, median aggregation methods yielded fewer cases with differences compared with mean aggregation methods. However, median aggregation methods produced, on average, greater absolute differences for both datasets. Standard deviations of absolute differences were similar for all tested aggregation methods. Maximum differences of 1 (i.e., the maximum possible difference) occurred for all investigated grouping features in both datasets (Figure 2).

Table 5: Percentage of differing cases, minimum, maximum, mean, and standard deviation of absolute differences between trait affinities assigned at family level by experts and aggregated trait affinities from five different aggregation methods.

Database	Comparison to traits at family level	Differing cases [%]	Min. differences	Max. differences	Mean abs. differences	SD abs. differences
Australia	*direct_agg (median)*	16.53	0.01	1.00	0.45	0.27
	direct_agg (mean)	23.24	< 0.01	0.99	0.34	0.23
	stepwise_agg (median)	17.90	0.01	1.00	0.42	0.26
	stepwise_agg (mean)	23.24	< 0.01	0.99	0.33	0.22
	weighted_agg	23.24	< 0.01	1.00	0.34	0.24
North America	*direct_agg (median)*	15.33	0.17	1.00	0.70	0.26
	direct_agg (mean)	47.00	< 0.01	1.00	0.30	0.26
	stepwise_agg (median)	18.00	0.08	1.00	0.63	0.28
	stepwise_agg (mean)	47.00	< 0.01	1.00	0.30	0.27
	weighted_agg	47.00	< 0.01	1.00	0.31	0.28

Abbreviations: Min., Minimum; Max., Maximum; abs., absolute; SD, Standard deviation.
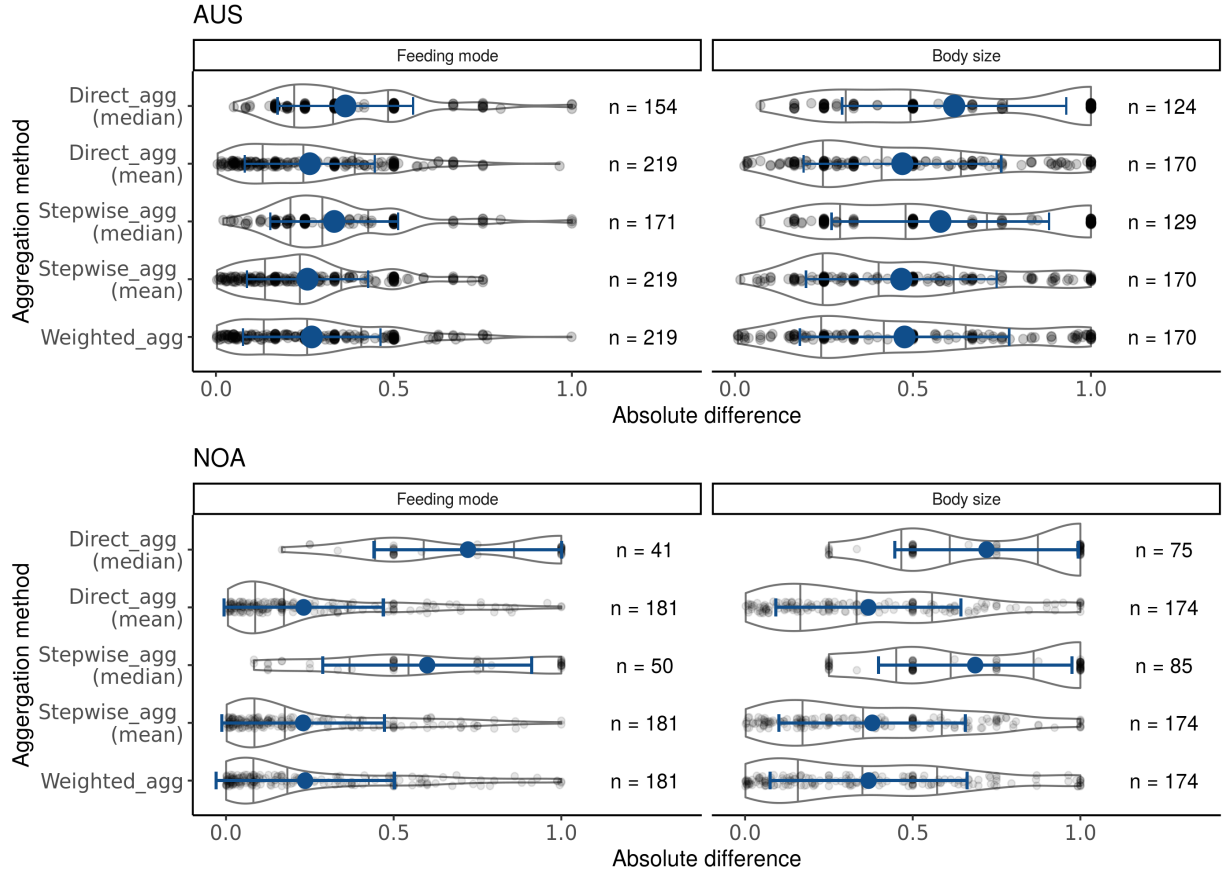
Figure 2: Cases (factor combination of investigated families and traits) where differences occurred between aggregated traits and expert assigned traits at family level for the Australian and North American dataset. Violin plots - mirrored density plots - show the density of the absolute trait affinity differences for the grouping features feeding mode and body size. Results for the North American dataset for the grouping features locomotion, respiration, and voltinism are shown in the supporting information (Figure S1). Absolute differences in trait affinities are depicted as grey dots. n denotes the number of cases per comparison where differences occurred. The blue dot indicates the mean of absolute differences, and the error bars show the standard deviation. The grey vertical lines show the 25th, 50th and 75th quantiles of the density estimate. Abbreviations: AUS, Australia; NOA, North America.

Comparison of aggregation methods with varying taxonomic hierarchies and trait variability

Our simulation results showed that both taxonomic hierarchical structure and trait variability affected aggregated trait affinities, although only to a small degree, and that aggregation method mattered in terms of the ranges of trait affinities produced over simulation replicates. The effect of taxonomic hierarchy differed across aggregation methods, but, as we expected, the produced range of trait affinities increased with increasing trait variability for all aggregation methods (Figure 3).

Taxonomic hierarchical structure appeared to influence the outcomes of the different aggregation methods. For the *sim_base* scenario (equal numbers of genera and species), the mean and weighted aggregation methods yielded similar ranges of aggregated trait affinities within each level of trait variability, and the median aggregation methods consistently produced greater ranges of aggregated trait affinities than the other methods. The largest ranges were produced by the *stepwise_agg $_{median}$* method. For the more complex taxonomic hierarchical structures, *sim_extreme* (one genus with a much larger number of species than the other four) and *sim_variation* (all genera with a different number of species), the *stepwise_agg $_{median}$* method still produced the largest ranges of trait affinities for most levels of trait variability (and *direct_agg$_{mean}$* produced the narrowest ranges for most levels of trait variability), but there was much less consistency in the ranges of trait affinities for all aggregation methods (i.e., the ranges were different among all aggregation methods).

Although trait aggregation methods were affected by taxonomic hierarchical structure and trait variability to some degree, in most simulated datasets, the different aggregation methods resulted in similar trait affinities. Only 1.42 %, or 213 out of 15.000 total comparisons (3 scenarios $*$ 5 levels of trait variability $*$ 10 unique comparisons of trait aggregation methods $*$ 100 replicates) showed a difference equal or greater than an absolute trait affinity of 0.1. Most (83.5 %) of these differences occurred in the *sim_extreme* scenario and were found between the aggregation methods *direct_agg$_{mean}$* and *stepwise_agg$_{median}$*, *direct_agg$_{median}$* and *stepwise_agg$_{median}$*, and *stepwise_agg$_{median}$* and *weighted_agg* (Figure 4).
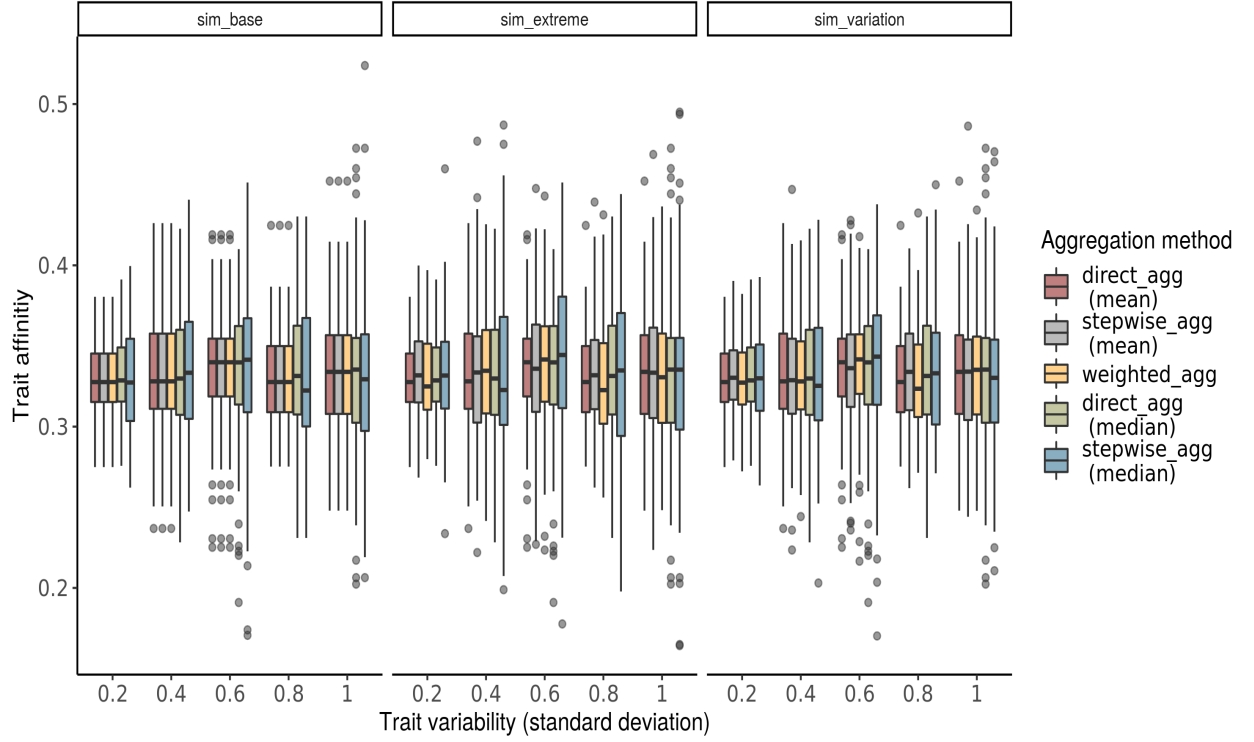
17

Figure 3: Ranges of aggregated trait affinities for the three examples of taxonomic hierarchies and simulated levels of trait variability. Shown are the results only for one simulated trait (T1). Similar results were obtained for the other simulated traits (Figure S2). Boxplots depict results for 100 replicated simulations of each trait aggregation method. The boxplot depicts the median and encompasses the 25th and 75th percentile. Horizontal black lines depict the median. Whiskers extend to the largest and smallest value respectively no further than $1.5 *$ the inter-quartile range. Outliers beyond the end of the whiskers are plotted as grey dots. Trait aggregation methods are in order of least to greatest produced ranges to improve visual inspection.
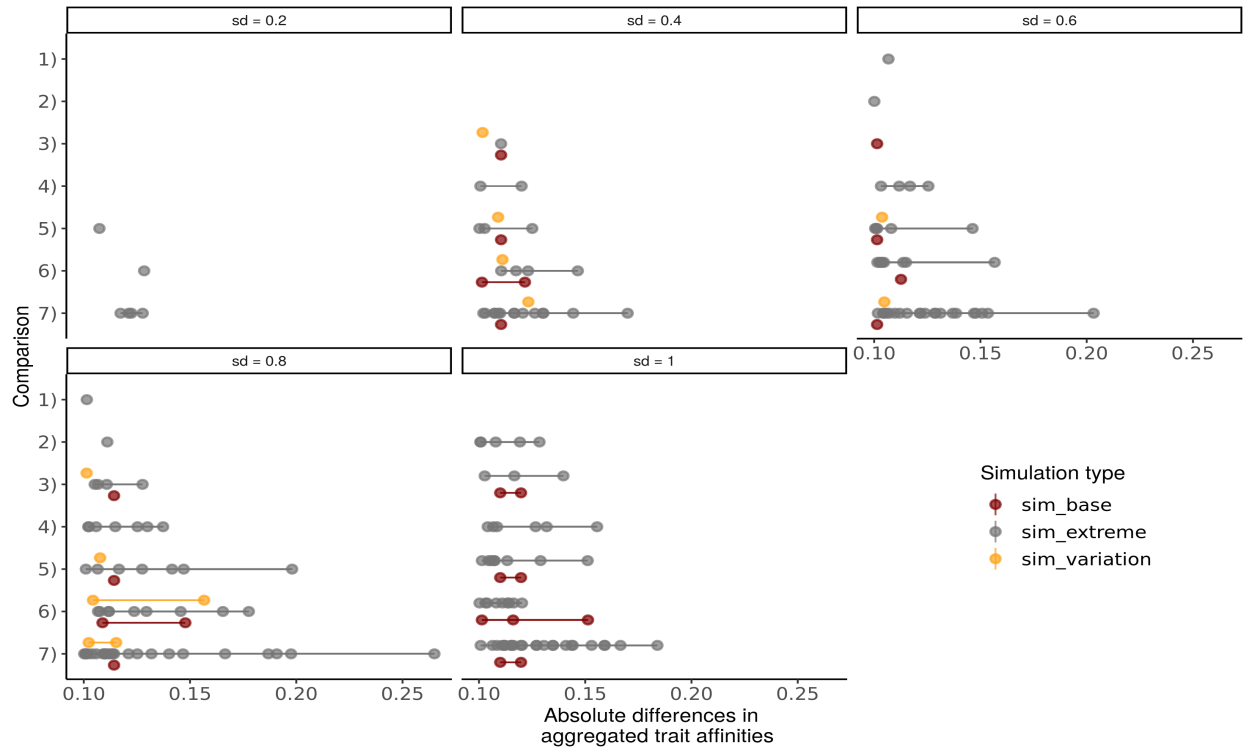
Figure 4: Comparison between the aggregated trait affinities produced by the different trait aggregation methods for every simulated dataset across all 3 simulated traits (5,000 comparisons per taxonomic hierarchy scenario). Results are shown for different levels of trait variability which is represented by the standard deviation (SD) of the simulated traits. Dots depict comparisons where absolute differences between aggregated trait affinities were greater than 0.1. Overall, there are 10 possible unique comparisons of which 7 produced absolute differences greater than 0.1.
Comparisons:
1) *direct_agg (median) - stepwise_agg (mean)*
2) *direct_agg (median) - weighted_agg*
3) *stepwise_agg (mean) - stepwise_agg (median)*
4) *stepwise_agg (mean) - weighted_agg*
5) *direct_agg (mean) - stepwise_agg (median)*
6) *direct_agg (median) - stepwise_agg (median)*
7) *stepwise_agg (median) - weighted_agg*

Effects of harmonisation and trait aggregation on trait-environment relationships

Our re-analysis of the Szöcs et al. (2014) study on the effects of salinisation on invertebrate traits showed that the use of harmonised and aggregated trait data may be useful for ecological studies on trait–environment relationships. Overall, harmonised trait data and all aggregation methods produced similar distributions of RDA species scores (5) compared with the original analysis, and they resulted in similar, but fewer, traits that distinguished upstream and downstream sites than the original analysis (Figures 5 and 6).

RDA using harmonised, but not aggregated, grouping features identified some, but not all, of the same traits that were identified by the original analysis (Figures 5 and 6). The original RDA in Szöcs et al. (2014) indicated that downstream sites (higher salinity) were characterised by the traits: ovoviviparity, multivoltinism, long life-cycle duration (> 1 year), shredder, and gill respiration, whereas RDA with the harmonised data identified the first three traits but did not include shredder or gill respiration traits. The original RDA indicated that upstream sites (lower salinity) were characterised by the traits: univoltinism, oviposition in clutches, and short life-cycle duration (< 1 year), and RDA with the harmonised data characterised upstream sites by univoltine taxa with short life-cycle durations that lay their eggs in an aquatic environment (aquatic eggs). This difference in oviposition trait is not substantial because the trait aquatic eggs of the harmonised grouping feature oviposition was derived by combining the trait oviposition in clutches and other related traits (Table 1).

RDA using harmonised traits that were aggregated to family level also produced results similar to the original analysis but with fewer traits distinguishing upstream and downstream sites (Figures 5 and 6). The $direct\_agg_{mean}$, $direct\_agg_{median}$, and $weighted\_agg$ characterised the downstream sites with the same traits as the original analysis except the trait shredder. As for the RDA with harmonised but not aggregated data, RDA with these three aggregation methods characterised upstream sites by the traits univoltinism, short life-cycle duration, and aquatic eggs. The $stepwise\_agg_{mean}$ and $stepwise\_agg_{median}$ methods produced the same results except that none of the life-cycle traits characterised upstream or downstream sites. Thus, compared with the original study's findings, aggregating by the direct and weighted methods yielded less change than the stepwise methods in interpretation of the RDA results.
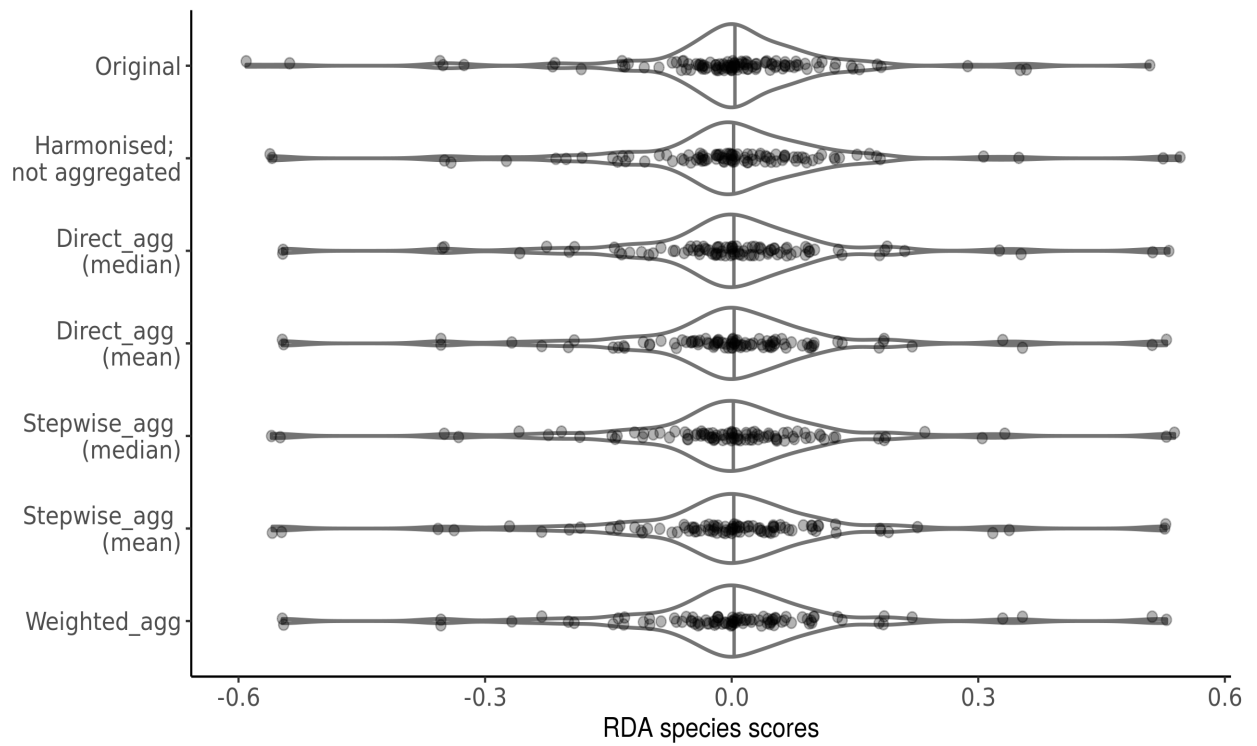
Figure 5: Species scores obtained by RDA from the original analysis (Szöcs et al., 2014), using harmonised grouping features, and using harmonised grouping features with traits aggregated to family level with five different aggregation methods. The dots represent the individual species scores for each analysed trait along the conductivity axis. The violin plot shows the density estimate of the species scores. Grey vertical lines indicate the median of the obtained species scores.
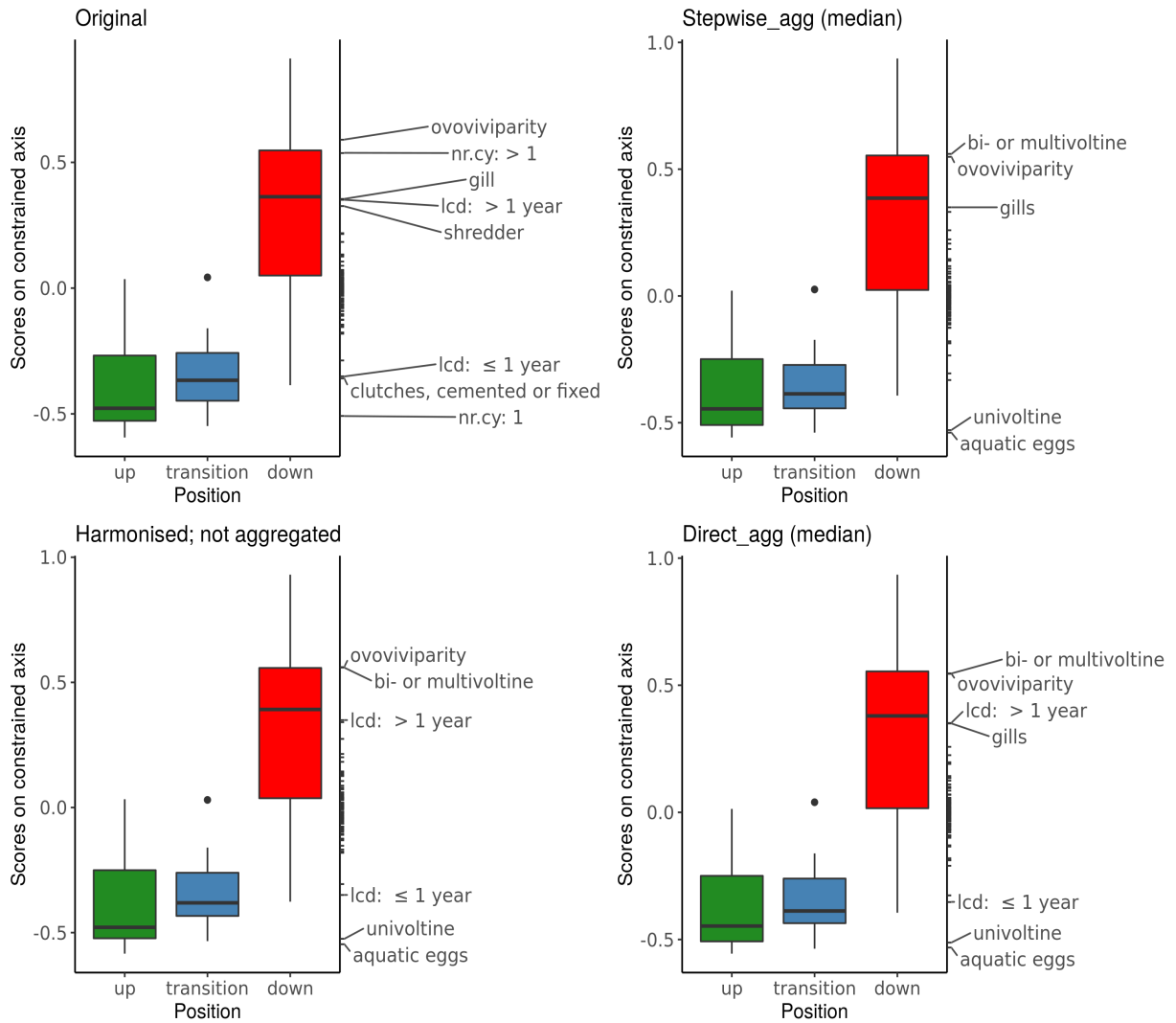
Figure 6: RDA of traits constrained by electric conductivity for harmonised but not aggregated data, data aggregated with *stepwise_agg* $_{median}$ (largest difference in terms of responsive traits to salinity to the original study), data aggregated with *direct_agg* $_{median}$ (smallest difference in terms of responsive traits to salinity to the original study), and the original study. Shown are boxplots of the site scores along the conductivity axis. The rug on the right side of each plot indicates species scores of the traits on the conductivity axis. Only traits with a Mahalanobis distance greater than the 97.5% quantile of the Chi-square distribution (5.02) were labelled. For better comparability, species scores of the original analysis were multiplied by -1. Results for the other aggregated methods are shown in the supporting information (Figure S5). Abbreviations: lcd, life cycle duration; nr.cy, potential number of cycles per year.

Discussion

Freshwater invertebrate trait databases are a valuable source of data that are freely available to researchers and that can offer opportunities for large-scale, cross-regional ecological study without the need for time-consuming and expensive field work. However, a lack of consistency among regional and continent-wide databases in trait definitions, trait affinity coding, and taxonomic resolution presents challenges to the use of multiple databases for a single study. We explored the effects of harmonising grouping features and aggregating trait affinities at different taxonomic levels from databases from four different continents on the outcomes of trait affinities and on analysis of trait–environment relationships. We found that despite discrepancies in trait definitions and taxonomic resolution across databases, harmonised and aggregated data were useful in identifying trait–environmental relationships, as evidenced by similar results from RDA as were found using observational data. Our results have implications for future multi-regional ecological studies that may benefit from the use of existing databases and provide insight into methods for harmonising traits with differing trait definitions and selecting aggregation methods that minimize either cases or ranges of differences in trait affinity scores when aggregating from higher to lower taxonomic precision.

Harmonising trait databases

Synthesizing trait information from multiple trait databases is crucial for developing the full potential of trait-based approaches, such as for studying community trait responses to environmental gradients across geographic regions. Others have noted the lack of standardised trait terminology in freshwater ecology (Baird et al., 2011; Brink et al., 2011), but we are the first to describe invertebrate trait definition discrepancies for some commonly used grouping features. Our experience can offer some insight into the need for, and process of, harmonising invertebrate trait information from trait databases of different regions. Some of the challenges we found were that grouping features were resolved differently into traits in different databases and that various codings (fuzzy, categorical, binary) were used to describe traits. In addition, the same traits were sometimes defined differently, requiring expert knowledge to reclassify these traits (e.g. the trait piercer). Missing information was another challenge. We aimed to use most of the available invertebrate trait information for different regions to establish harmonised grouping feature datasets, but the availability of trait information varied strongly across grouping features and databases. For instance, there was extensive information available for grouping features that are often used in trait–environment studies, such as feeding mode and respiration, but there was surprisingly little information on body form, a trait that is relatively easy to determine.

To resolve definition discrepancies and facilitate data synthesis in the future, terminological standards are needed. Harmonised definitions and concepts of traits have been developed in the past for other organism groups, such as for plants with the *Thesaurus of Plant Characteristics* (TOP) initiative (Garnier et al., 2017). The core of this harmonisation initiative is to provide standardised trait definitions and to draw connections to synonyms, related terms, and surrounding concepts by linking to other controlled vocabularies or ontologies in the field. Such initiatives could be a role model for freshwater ecologists to establish unambiguous terminologies for invertebrate traits. The existing freshwater invertebrate trait databases could be linked through standardised terminologies or ontologies, as suggested by Baird et al. (2011). By following the recently proposed *Ecological Trait-data Standard Vocabulary* (ETS), providers of invertebrate trait data could connect their traits

to such ontologies (Schneider et al., 2019). Once standardised trait definitions are established, these will improve trait data sharing, trait data processing when working with multiple trait databases, and, ultimately, interpretation of the derived results.

Trait aggregation method choice

Observational data for freshwater invertebrates are often resolved to a variety of taxonomic levels because species-level identification can be complicated, time-consuming, and expensive (Marshall et al., 2006; Resh, 2008). Consequently, trait information at family level has been widely used by freshwater ecologists, for example in bioassessment (Beketov et al., 2009). However, assigned traits at family level may not reflect the real trait diversity within, for example, a taxon in the large, ecologically diverse family Chironomidae (Serra et al., 2016). Additionally, various trait aggregation methods have been used by researchers to allow comparison among traits at different taxonomic levels. Evaluation of the differences between aggregated and assigned traits is difficult because it remains unclear what the true value of a particular trait is for a particular family. Aggregation of trait information at species or genus level to a point estimate at family level suggests a precision that is not necessarily present, especially not for traits with high variability or if trait information at species level is missing. Indeed, some traits can vary strongly at more precise taxonomic levels than the family level. For example, Monaghan and Soares (2013) found a high intra-family trait dispersion in the Tachet database for traits of the grouping features body size, flow preference, and reproduction. Further, studies focusing on the lability of traits, (i.e. how much traits are constrained by phylogeny), found that traits of ecological preferences (e.g. thermal preference), body size, resistance forms, and, to a lesser extent, feeding mode are labile, and thus possibly highly variable (Poff et al., 2006; Wilkes et al., 2020).

If trait aggregation is necessary, however, our study can offer insight about the circumstances under which the choice of aggregation method may be important. The 5 aggregation methods we tested used either the mean or the median and different weightings to assign family-level affinities. Our results indicated that for both the Australian and North American datasets (1) the median aggregation methods provided affinities that were often closer to the assigned traits than were those provided by the mean aggregation methods and (2) the different weighting approaches exerted only a minor influence on the aggregated trait affinities. However, when differences occurred between median aggregation methods and assigned traits, these differences were greater in absolute value compared with the mean aggregation methods, particularly for the North American dataset where mean absolute differences for median aggregations were twice as high as for mean aggregations. This pattern of greater differences for median than mean aggregations can, in large part, be explained by the binary coding used in the North American trait dataset and by the assigned traits. Binary coded traits have affinities of either 1 or 0; therefore, traits of a particular grouping feature showed, in most cases, a higher difference in trait affinities to each other than if they were fuzzy coded. Median-aggregated binary-coded taxa often resulted in a value of 0 or 1, whereas mean aggregation often yielded values between 0 and 1. Thus, using the median led for the North American dataset either to agreement with the assigned traits or to higher differences compared to the mean aggregation methods. This pattern inherent in using median-aggregation methods to aggregate binary-coded data should be considered when researchers choose methods to resolve affinities at different taxonomic levels.

We expected that, in addition to averaging measure, different weighting approaches to aggregation (i.e., equal weight to each taxon, equal weight to each genus, or weighted by number of

species) would affect affinities for families with varying taxonomic hierarchical structure. However, we found through simulations that although taxonomic structural unevenness appeared to produce variation in affinities among weighting approaches, especially for a family where one genus had a much larger number of species than its other genera, trait variability had a greater effect on the range of affinities (i.e., greater ranges of affinities at higher levels of variance) and on the differences in affinities between weighting methods. Also, the number of differing cases and the mean absolute differences in trait affinities, as well as the distributions of absolute trait affinity differences to assigned traits, were similar across the mean aggregation and across the median aggregation methods, which suggests a small influence of the weighting approach on the aggregation outcomes. The minor impact of the weighting approaches on trait aggregation may be explained by the fact that a considerable portion of taxa had low numbers of genera or species. Of the taxa that were compared from the North American trait dataset, 14 % were identified at family and 62 % at genus level, 52 % comprised five or fewer genera, and 13 % contained just one genus (Figure S3). In the Australian dataset, 21 % of the compared taxa were identified at family, 40 % at genus, and 39 % at species level, 68 % of the taxa contained five or fewer genera, and 40 % just one genus (Figure S4). Hence, these results could change when more species-level trait information becomes available. Our findings show that the choice of aggregation method may matter less when taxonomic hierarchical structure is fairly even and traits are less variable, but the $stepwise\_agg_{median}$ method tended to produce the widest range in affinities, especially with high taxonomic unevenness and high trait variability.

Using harmonised and aggregated datasets for ecological research

By reanalysing trait–environment relationships identified in Szöcs et al. (2014) with our harmonised and family-level aggregated datasets, we have illustrated the potential for the use of harmonised and aggregated data in ecological research. As far as we are aware, ours is the first study to compare trait–environment relationships using harmonised and non-harmonised data. Our re-analysis of the data on salinisation effects on biological traits yielded only slightly different results compared with the original study, which agrees with previous findings that family-level traits can be sensitive enough to detect environmental impacts (Beketov et al., 2009). However, some of the traits that had responded in the original analysis did not respond in the re-analysis: feeding mode shredder, respiration gills, and life-cycle duration traits. There are several possible reasons for these discrepancies. The non-responsive traits were those closest to the threshold for association with either higher or lower salinity, and so the precision lost through trait harmonisation may have decreased our ability to detect a relatively weak association. Feeding mode shredder was the only trait that was non-responsive when using harmonised (but not aggregated) data as well as when using aggregated data. This result was likely because of the harmonisation procedure, when this trait was combined based on three traits (miner, xylophagus, and shredder). Consequently, the trait affinities in the original data had a higher mean and standard deviation than in the harmonised and aggregated data (Table S2), suggesting that the signal in the original data was weakened by harmonisation. Following this result, we suggest that if harmonisation is necessary, harmonised and non-harmonised data, if available, should be compared, and possible averaging effects should be considered in further analyses. However, our results illustrate that where site-level data are not available, harmonised and aggregated datasets may be useful in identifying at least the strongest associations between traits and environmental conditions.

Outlook

Although comprehensive freshwater invertebrate trait databases have been developed for several regions in the past, data synthesis presents challenges because of discrepancies in trait definitions, coding of trait affinities, and taxonomic resolution. By providing an overview of definition discrepancies, we have set a starting point for the development of standardised trait terminology through which invertebrate trait databases can be linked. Agreement on standard terminology and the subsequent development of ontologies are the next steps to facilitate trait-based analysis at large geographical scales. As our analysis showed, some grouping features might need to be re-classified to fit into such a standardised terminology. Although we showed that trait affinities from fuzzy and binary coding can be used together, we also recommend that a uniform coding of traits should be addressed during trait standardisation.

As ecological knowledge increases, and with the aid of new methods such as computer vision for species and trait identification (Høye et al., 2020), more observational and species-level trait data may become available in the future, reducing the need for trait aggregation. However, we showed that trait aggregation is often in agreement with expert assignments, especially through aggregation approaches that use the median. No aggregation method is objectively the best, and except for cases where traits are highly variable and taxonomic hierarchical structure is highly uneven, choice of aggregation method does not appear to be influential on the outcome. Weighting trait information based on the taxa present in the database used does not seem to have much influence except when the taxonomic hierarchical structure is highly uneven. However, we recommend that the type of coding used to describe traits (binary, fuzzy) should be considered, because median and mean aggregations can yield very different results for binary-coded traits. We also suggest that a measure of trait dispersion should be reported to indicate the uncertainty of the aggregated estimate. Through the use of carefully synthesised data from freshwater invertebrate trait databases, while acknowledging the limitations inherent in harmonised and aggregated data, the ability to investigate a wide range of large-scale ecological questions is enhanced.

Data availability statement

Harmonised invertebrate trait datasets for Australia, Europe, North America and New Zealand are deposited *Zenondo/Github?*

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

References

Aspin, T. W. H., Khamis, K., Matthews, T. J., Milner, A. M., O'Callaghan, M. J., Trimmer, M., ... Ledger, M. E. (2019). Extreme drought pushes stream invertebrate communities over functional thresholds. *Global Change Biology*, *25*(1), 230–244. doi:10.1111/gcb.14495

Baird, D. J., Baker, C. J. O., Brua, R. B., Hajibabaei, M., McNicol, K., Pascoe, T. J., & Zwart, D. d. (2011). Toward a knowledge infrastructure for traits-based ecological risk assessment. *Integrated Environmental Assessment and Management*, *7*(2), 209–215. _eprint: https://setac.onlinelibrary.wiley.com/do doi:10.1002/ieam.129

Bäthe, J. & Coring, E. (2011). Biological effects of anthropogenic salt-load on the aquatic fauna: A synthesis of 17 years of biological survey on the rivers werra and weser. *Limnologica*. Salinisation of running waters, *41*(2), 125–133. doi:10.1016/j.limno.2010.07.005

Beketov, M. A., Foit, K., Schäfer, R. B., Schriever, C. A., Sacchi, A., Capri, E., ... Liess, M. (2009). SPEAR indicates pesticide effects in streams – comparative use of species- and family-level biomonitoring data. *Environmental Pollution*, *157*(6), 1841–1848. doi:10.1016/j.envpol.2009.01.021

Bhowmik, A. K. & Schäfer, R. B. (2015). Large scale relationship between aquatic insect traits and climate. *PLoS ONE*, *10*(6). doi:10.1371/journal.pone.0130025

Bonada, N., Dolédec, S., & Statzner, B. (2007). Taxonomic and biological trait differences of stream macroinvertebrate communities between mediterranean and temperate regions: Implications for future climatic scenarios. *Global Change Biology*, *13*(8), 1658–1671. doi:10.1111/j.1365-2486.2007.01375.x

Botwe, P. K., Carver, S., Magierowski, R., McEvoy, P., Goonan, P., Madden, C., & Barmuta, L. A. (2018). Effects of salinity and flow interactions on macroinvertebrate traits in temporary streams. *Ecological Indicators*, *89*, 74–83. doi:10.1016/j.ecolind.2018.01.036

Brink, P. J. V. d., Alexander, A. C., Desrosiers, M., Goedkoop, W., Goethals, P. L., Liess, M., & Dyer, S. D. (2011). Traits-based approaches in bioassessment and ecological risk assessment: Strengths, weaknesses, opportunities and threats. *Integrated Environmental Assessment and Management*, *7*(2), 198–208. _eprint: https://setac.onlinelibrary.wiley.com/doi/pdf/10.1002/ieam.109. doi:https://doi.org/10.1002/ieam.109

Brown, L. E., Khamis, K., Wilkes, M., Blaen, P., Brittain, J. E., Carrivick, J. L., ... Milner, A. M. (2018). Functional diversity and community assembly of river invertebrates show globally consistent responses to decreasing glacier cover. *Nature Ecology & Evolution*, *2*(2), 325–333. doi:10.1038/s41559-017-0426-x

Chessman, B. C. (2018). Dissolved-oxygen, current and temperature preferences of stream invertebrates estimated from field distributions: Application to assemblage responses to drought. *Hydrobiologia*, *809*(1), 141–153. doi:10.1007/s10750-017-3455-1

Culp, J. M., Armanini, D. G., Dunbar, M. J., Orlofske, J. M., Poff, N. L., Pollard, A. I., ... Hose, G. C. (2011). Incorporating traits in aquatic biomonitoring to enhance causal diagnosis and prediction. *Integrated Environmental Assessment and Management*, *7*(2), 187–197. _eprint: https://setac.onlinelibrary.wiley.com/doi/pdf/10.1002/ieam.128. doi:10.1002/ieam.128

Garnier, E., Stahl, U., Laporte, M.-A., Kattge, J., Mougenot, I., Kühn, I., ... Klotz, S. (2017). Towards a thesaurus of plant characteristics: An ecological contribution. *Journal of Ecology*, *105*(2), 298–309. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2745.12698. doi:https://doi.org/10.1111/1365-2745.12698

Høye, T. T., Ärje, J., Bjerge, K., Hansen, O. L. P., Iosifidis, A., Leese, F., . . . Raitoharju, J. (2020). Deep learning and computer vision will transform entomology. *bioRxiv*, 2020.07.03.187252. Publisher: Cold Spring Harbor Laboratory Section: New Results. doi:10.1101/2020.07.03.187252

Kefford, B. J., Botwe, P. K., Brooks, A. J., Kunz, S., Marchant, R., Maxwell, S., . . . Thompson, R. M. (2020). An integrated database of stream macroinvertebrate traits for australia: Concept and application. *Ecological Indicators*, *114*, 106280. doi:10.1016/j.ecolind.2020.106280

Kissling, W. D., Walls, R., Bowser, A., Jones, M. O., Kattge, J., Agosti, D., . . . Guralnick, R. P. (2018). Towards global data products of essential biodiversity variables on species traits. *Nature Ecology & Evolution*, *2*(10), 1531–1540. Number: 10 Publisher: Nature Publishing Group. doi:10.1038/s41559-018-0667-3

Magliozzi, C., Usseglio-Polatera, P., Meyer, A., & Grabowski, R. C. (2019). Functional traits of hyporheic and benthic invertebrates reveal importance of wood-driven geomorphological processes in rivers. *Functional Ecology*, *33*(9), 1758–1770. doi:10.1111/1365-2435.13381

Marshall, J., Steward, A., & Harch, B. (2006). Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an australian dryland river: The benefits and costs of using species abundance data. *Hydrobiologia*, *572*, 171–194. doi:10.1007/s10750-005-9007-0

McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, *21*(4), 178–185. doi:10.1016/j.tree.2006.02.002

Monaghan, K. A. & Soares, A. M. (2013). Improving the accuracy of aquatic macroinvertebrate trait assignment — the importance of taxon-weighting. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(6), 885–894. doi:10.1139/cjfas-2012-0402

Moog, O. & Hartmann, A. (1995). A comprehensive species inventory of austrian aquatic organisms with ecological notes, 86.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2019). *Vegan: Community ecology package*. R package version 2.5-6. Retrieved from https://CRAN.R-project.org/package=vegan

Philips, N. & Smith, B. (2018). New zealand freshwater macroinvertebrate trait database. *NIWA*.

Pilière, A. F. H., Verberk, W. C. E. P., Gräwe, M., Breure, A. M., Dyer, S. D., Posthuma, L., . . . Schipper, A. M. (2016). On the importance of trait interrelationships for understanding environmental responses of stream macroinvertebrates. *Freshwater Biology*, *61*(2), 181–194. doi:10.1111/fwb.12690

Poff, N. L., Olden, J. D., Vieira, N. K. M., Finn, D. S., Simmons, M. P., & Kondratieff, B. C. (2006). Functional trait niches of north american lotic insects: Traits-based ecological applications in light of phylogenetic relationships. *Journal of the North American Benthological Society*, *25*(4), 730–755. doi:10.1899/0887-3593(2006)025[0730:FTNONA]2.0.CO;2

Poff, N., Pyne, M., Bledsoe, B., Cuhaciyan, C., & Carlisle, D. (2010). Developing linkages between species traits and multiscaled environmental variation to explore vulnerability of stream benthic communities to climate change. *Journal of the North American Benthological Society*, *29*(4), 1441–1458. doi:10.1899/10-030.1

Resh, V. H. (2008). Which group is best? attributes of different biological assemblages used in freshwater biomonitoring programs. *Environmental Monitoring and Assessment*, *138*(1), 131–138. doi:10.1007/s10661-007-9749-4

Schäfer, R. B., Kefford, B. J., Metzeling, L., Liess, M., Burgert, S., Marchant, R., . . . Nugegoda, D. (2011). A trait database of stream invertebrates for the ecological risk assessment of single

and combined effects of salinity and pesticides in south-east australia. *Science of The Total Environment*, *409*(11), 2055–2063. doi:10.1016/j.scitotenv.2011.01.053

Schmera, D., Podani, J., Heino, J., Erős, T., & Poff, N. L. (2015). A proposed unified terminology of species traits in stream ecology. *Freshwater Science*, *34*(3), 823–830. doi:10.1086/681623

Schmidt-Kloiber, A. & Hering, D. (2015). Www.freshwaterecology.info – an online tool that unifies, standardises and codifies more than 20,000 european freshwater organisms and their ecological preferences. *Ecological Indicators*, *53*, 271–282. doi:10.1016/j.ecolind.2015.02.007

Schneider, F. D., Fichtmueller, D., Gossner, M. M., Güntsch, A., Jochum, M., König-Ries, B., . . . Simons, N. K. (2019). Towards an ecological trait-data standard. *Methods in Ecology and Evolution*, 2041–210X.13288. doi:10.1111/2041-210X.13288

Serra, S. R. Q., Cobo, F., Graça, M. A. S., Dolédec, S., & Feio, M. J. (2016). Synthesising the trait information of european chironomidae (insecta: Diptera): Towards a new database. *Ecological Indicators*, *61*, 282–292. doi:10.1016/j.ecolind.2015.09.028

Southwood, T. R. E. (1977). Habitat, the templet for ecological strategies? *Journal of Animal Ecology*, *46*(2), 337–365. doi:10.2307/3817

Statzner, B., Bis, B., Dolédec, S., & Usseglio-Polatera, P. (2001). Perspectives for biomonitoring at large spatial scales: A unified measure for the functional composition of invertebrate communities in european running waters. *Basic and Applied Ecology - BASIC APPL ECOL*, *2*. doi:10.1078/1439-1791-00039

Statzner, B., Hoppenhaus, K., Arens, M.-F., & Richoux, P. (1997). Reproductive traits, habitat use and templet theory: A synthesis of world-wide data on aquatic insects. *Freshwater Biology*, *38*(1), 109–135. doi:10.1046/j.1365-2427.1997.00195.x

Szöcs, E., Coring, E., Bäthe, J., & Schäfer, R. B. (2014). Effects of anthropogenic salinization on biological traits and community composition of stream macroinvertebrates. *Science of The Total Environment*, *468-469*, 943–949. doi:10.1016/j.scitotenv.2013.08.058

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

Twardochleb, L., Hiltner, E., Pyne, M., & Zarnetske, P. (2021). Freshwater insects CONUS: A database of freshwater insect occurrences and traits for the contiguous united states. *Global Ecology and Biogeography*, *n/a*. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/geb.13257. doi:https://doi.org/10.1111/geb.13257

Usseglio-Polatera, P., Bournaud, M., Richoux, P., & Tachet, H. (2000). Biomonitoring through biological traits of benthic macroinvertebrates: How to use species trait databases? *Hydrobiologia*, *422*(0), 153–162. doi:10.1023/A:1017042921298

Usseglio-Polatera, P. (2020a). Body form traits for european and north american taxa. Personal information.

Usseglio-Polatera, P. (2020b). Classification of piercers as herbivores or predators. Personal information.

Usseglio-Polatera, P., Bournaud, M., Richoux, P., & Tachet, H. (2000). Biological and ecological traits of benthic freshwater macroinvertebrates: Relationships and definition of groups with similar traits. *Freshwater Biology*, *43*(2), 175–205. doi:10.1046/j.1365-2427.2000.00535.x

Verberk, W. C. E. P., van Noordwijk, C. G. E., & Hildrew, A. G. (2013). Delivering on a promise: Integrating species traits to transform descriptive community ecology into a predictive science. *Freshwater Science*, *32*(2), 531–547. doi:10.1899/12-092.1

version 3.6.1; R Project for Statistical Computing, Vienna, Austria. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Vieira, N. K. M., Poff, N. L., Carlisle, D. M., II, S. R. M., Koski, M. L., & Kondratieff, B. C. (2006). A database of lotic invertebrate traits for north america, 19.

Voß, K. & Schäfer, R. (2017). Taxonomic and functional diversity of stream invertebrates along an environmental stress gradient. *Ecological Indicators*, *81*, 235–242. doi:10.1016/j.ecolind.2017.05.072

Wilkes, M. A., Edwards, F., Jones, J. I., Murphy, J. F., England, J., Friberg, N., . . . Brown, L. E. (2020). Trait-based ecology at large scales: Assessing functional trait correlations, phylogenetic constraints and spatial variability using open data. *Global Change Biology*, gcb.15344. doi:10.1111/gcb.15344

Supporting Information

Harmonised trait datasets

Discrepancies in trait definitions

Table S1: Comparison of trait definitions between invertebrate trait databases. Only traits that are differently described across databases are listed. The definition is quoted if it enables differences to be identified, otherwise the differences are described. The hyphen indicates a missing trait. Reproduction was captured in multiple grouping features per database. Hence, differences for reproduction have been described in the paper. Body form traits are not different between databases, except that the Vieira database contains the trait Bluff (blocky) which does not appear in the other databases.

Trait	Freshwater-ecology.info	Tachet	CONUS	Vieira	Australia	New Zealand
Feeding shredder	"Feed from fallen leaves, plant tissues, CPOM"	"Eat coarse detritus, plants or *animal material*"	• "Shred decomposing vascular plant tissue" • Trait herbivore includes among others insect that shred *living aquatic plants*	Shredder	• Detrivore† • Trait herbivore includes among others the trait shredder	Shredders
Feeding predator	"Eating from prey"	• Carvers, engulfers & swallowers • Piercers (plants & animals) are an additional trait	Engulfers ("ingest prey whole or in parts") & piercers ("prey tissues and suck fluids")	Predator	Piercer & engulfer	Predator

	Distinguishes between active and passive	No distinction between active and passive	No distinction between active and passive	No distinction between active and passive	No distinction between active and passive	No distinction between active and passive
Feeding filter-feeder						
Semivoltine	"One generation in two years"	"Life cycle lasts *at least* two years"	"< 1 generation per year"	"< 1 generation per year"	"< 1 generation per year"	"< 1 reproductive cycle per year"
Multivoltine	"*Three* or more generations per year"‡	"Able to complete *at least* two successive generations per year"	"> 1 generations per year"	"> 1 generations per year"	• 1-2 generations per year • bi/multivoltine • up to 5 generations per year • up to 10 generations per year	"> 1 reproductive cycles per year"
Locomotion swimming	• Passive movement like floating or drifting (trait swimming/scating) • Active movement (trait swimming/diving)	• Surface swimmers (over and under the water surface) • Full water swimmers (e.g. Baetidae).	"Adapted for 'fish-like' swimming"	Swimmer	Distinguishes swimmer and skater	Swimmers (water column)

Locomotion burrowing	"Burrowing in *soft* substrates or boring in *hard* substrates"	• Burrowing "within the first centimeters of the benthic fine sediment" • Differentiates also the trait interstitial (endobenthic)	"Inhabiting *fine* sediment of streams and lakes"	Burrower	"Moving deep into the substrate and thus avoiding flow"	Burrowers (infauna)
Locomotion sprawling & walking	"Sprawling or walking actively with legs, pseudopods or on a mucus"	-	Sprawling: "inhabiting the surface of floating leaves of vascular hydrophytes or fine sediments"	Sprawler	-	-
Locomotion crawling	-	"Crawling over the bottom substrate"	Defined as crawling on the surface of floating leaves or fine sediments on the bottom	-	Database contains traits crawler, climber sprawler, and clinger.	Crawlers (epibenthic)
Locomotion sessil	Does not distinguish temporarily and permanently attached	Distinguishes temporarily and permanently attached	Does not distinguish temporarily and permanently attached	Does not distinguish temporarily and permanently attached	Distinguishes temporarily and permanently attached	Does not distinguish temporarily and permanently attached

Respiration plastron & spiracle	Plastron and spiracle (aerial) are two separate traits	Definition includes respiration using air stores of aquatic plants	Plastron and spiracle combined into one trait	Distinguishes spiracular gills, plastron, atmospheric breathers and plant breathers	Plastron and spiracle (termed aerial) occur as separate and combined traits. Contains also traits: air (plants), atmospheric, and functional spiracles	Distinguishes plastron and spiracle (termed aerial)
Body size small	-	Multiple size classifications¶	< 9 mm	< 9 mm	< 9 mm †§	Multiple size classifications⋆
Body size medium	-		9 - 16 mm	9 - 16 mm	9 - 16 mm	
Body size large	-		> 16 mm	> 16 mm	> 16 mm	

† Traits from Botwe et al.
‡ Contains also bivoltine (two generations per year), trivoltine (three generations per year) and flexible.
§ Contains a size trait with numeric size values. Contains also traits classifying size like Tachet and like the North American trait databases.
¶ Size classifications: $<= 0.25$ *cm*, $> 0.25 - 0.5$ *cm*, $0.5 - 1$ *cm*, $1 - 2$ *cm*, $2 - 4$ *cm*, $4 - 8$ *cm*, > 8 *cm*. No distinction into small, medium and large.
⋆ Size classifications: $> 0.25 - 0.5$ *cm*, $0.5 - 1$ *cm*, $1 - 2$ *cm*, $2 - 4$ *cm*, $4 - 8$ *cm*. No distinction into small, medium and large.

Comparing aggregation methods

Comparison of family-level aggregated traits with family-level assigned traits
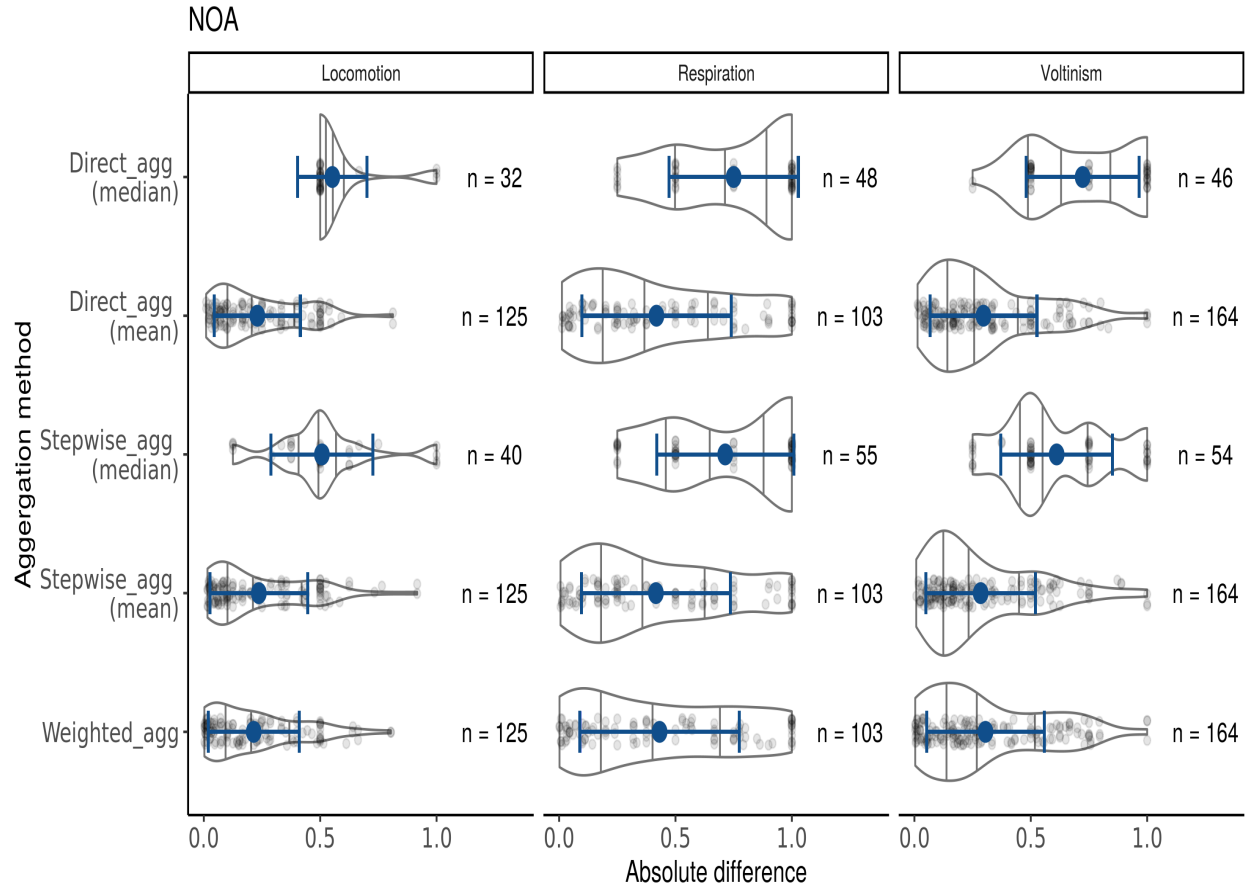

Figure S1: Cases (factor combination of investigated families and traits) where differences occurred between aggregated traits and expert assigned traits at family level for the North American dataset. Violin plots - mirrored density plots - show the density of the absolute trait affinity differences for the grouping features locomotion, respiration, and body size. For more details see Figure 2.

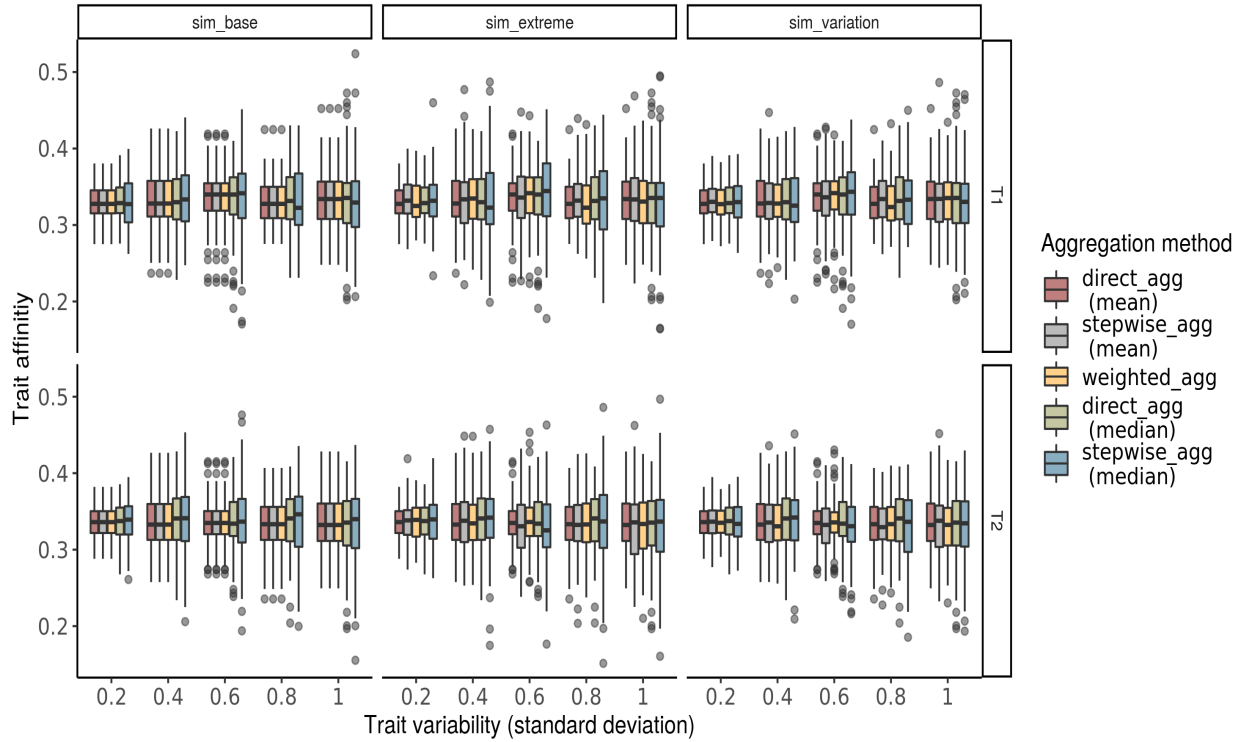**Comparison of aggregation methods with varying taxonomic hierarchies and trait variability**

Figure S2: Ranges of aggregated trait affinities for the three examples of taxonomic hierarchies and simulated levels of trait variability. Shown are the results for the simulated traits T2 and T3. Boxplots depict results for 100 replicated simulations of each trait aggregation method. Trait aggregation methods are in order of least to greatest produced ranges to improve visual inspection. For more details see Figure 3.

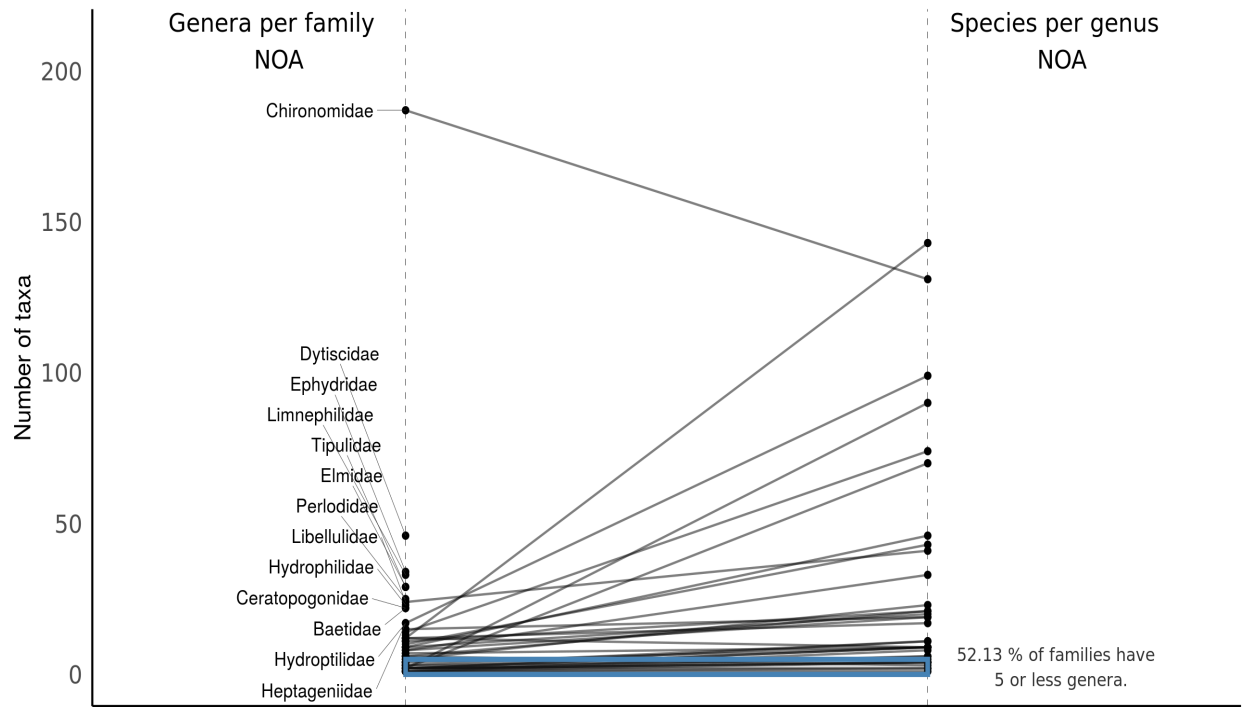**Taxonomic hierarchy in the trait datasets used for comparisons with assigned traits at family level**

Figure S3: Number of genera per family and species per genus for those families of the North American trait dataset that have been compared with assigned traits at family level. For better visual display only families with more than 15 genera are displayed.
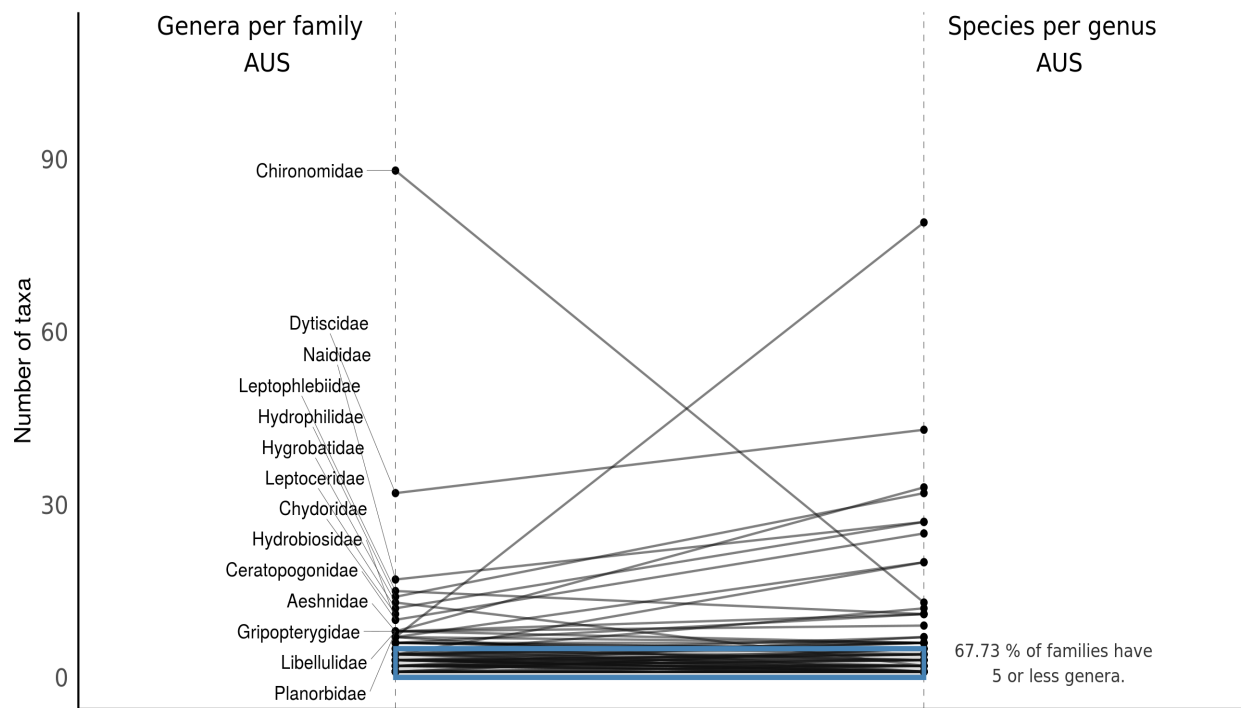
Figure S4: Number of genera per family and species per genus for the Australian trait dataset. For better visual display only families with more than 7 genera are displayed.

Effects of harmonisation and trait aggregation on trait-environment relationships

Table S2: Mean, median and standard deviation of the affinities of traits that were responsive to the salinity gradient in the original study but not in the re-analysis using the harmonised European trait dataset.

| Type	Trait	Mean	Median	SD	Responsive?
Stepw_median	Shredder	0.20	0.14	0.25	No
Stepw_mean	Shredder	0.18	0.12	0.22	No
Direct_median	Shredder	0.21	0.14	0.25	No
Direct_mean	Shredder	0.19	0.14	0.22	No
Weighted	Shredder	0.19	0.14	0.22	No
Harmonised; not_aggregated	Shredder	0.18	0.12	0.24	No
Original	Shredder	0.25	0.14	0.32	Yes
Stepw_median	Gills	0.30	0.27	0.32	Yes
Stepw_mean	Gills	0.29	0.22	0.32	Yes
Direct_median	Gills	0.30	0.30	0.32	Yes
Direct_mean	Gills	0.30	0.30	0.32	Yes
Weighted	Gills	0.30	0.30	0.32	Yes
Harmonised; not_aggregated	Gills	0.30	0.25	0.32	No
Original	Gills	0.28	0.00	0.33	Yes
Stepw_median	Short life cycle	0.64	0.75	0.39	No
Stepw_mean	Short life cycle	0.64	0.79	0.39	No
Direct_median	Short life cycle	0.67	0.75	0.37	Yes
Direct_mean	Short life cycle	0.67	0.79	0.38	Yes
Weighted	Short life cycle	0.67	0.79	0.38	Yes
Harmonised; not_aggregated	Short life cycle	0.64	0.75	0.40	Yes
Original	Short life cycle	0.64	0.75	0.40	Yes
Stepw_median	Long life cylce	0.36	0.25	0.39	No
Stepw_mean	Long life cylce	0.36	0.21	0.39	No
Direct_median	Long life cylce	0.33	0.25	0.37	Yes
Direct_mean	Long life cylce	0.33	0.21	0.38	Yes
Weighted	Long life cylce	0.33	0.21	0.38	Yes
Harmonised; not_aggregated	Long life cylce	0.36	0.25	0.40	Yes
Original	Long life cylce	0.36	0.25	0.40	Yes
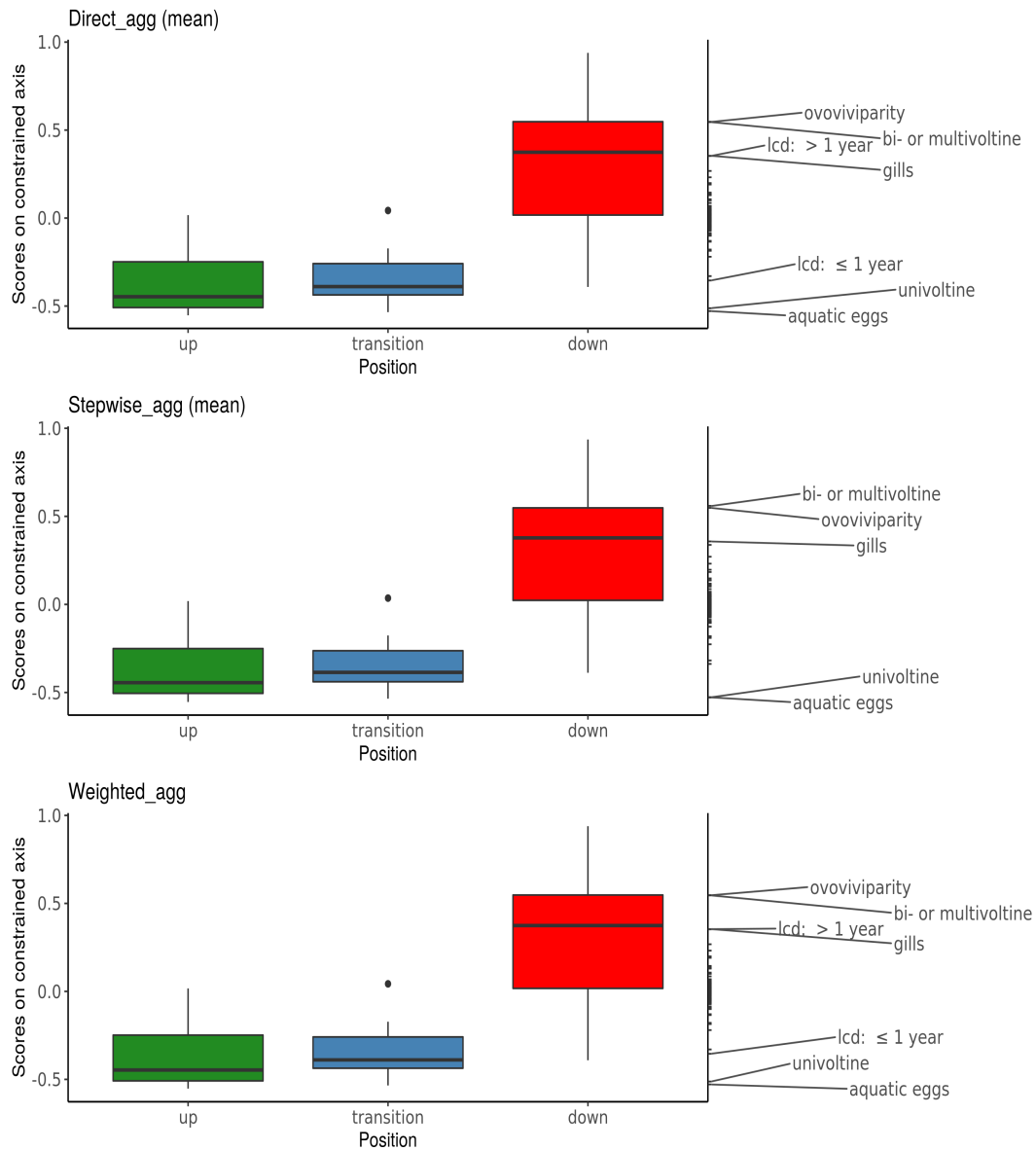
Figure S5: RDA of traits constrained by electric conductivity for the data aggregated with *direct_agg $_{mean}$*, *stepwise_agg $_{mean}$*, and *weighted_agg*. Shown are boxplots of the site scores along the conductivity axis. The rug on the right side of each plot indicates species scores of the traits on the conductivity axis. For more details see Figure 6. Abbreviations: lcd, life cycle duration; nr.cy, potential number of cycles per year.