

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

Stefan Kunz¹, Ben J. Kefford², Astrid Schmidt-Kloiber³, Christoph D. Matthaei⁴, Philippe Usseglio-Polatera⁵, Wolfram Graf³, N. LeRoy Poff⁶, Leon Metzeling⁷, Laura Twardochleb⁸, Charles P. Hawkins⁹,
and Ralf B. Schäfer¹

¹Institute for Environmental Sciences, University of Koblenz-Landau,
Landau, Germany

²Centre for Applied Water Science, Institute for Applied Ecology,
University of Canberra, Canberra, Australia

³Institute of Hydrobiology and Aquatic Ecosystem Management,
University of Natural Resources and Life Sciences Vienna (BOKU),
Vienna, Austria

⁴Department of Zoology, University of Otago, Dunedin, New Zealand

⁵University of Lorraine, CNRS, LIEC, Metz, France

⁶Department of Biology, Colorado State University, Fort Collins, USA

⁷Environment Protection Authority Victoria, Applied Sciences
Division, Macleod, Australia

⁸Department of Fisheries and Wildlife, Michigan State University,

East Lansing, USA

⁹Department of Watershed Sciences, National Aquatic Monitoring Center, and the Ecology Center, Utah State University, Logan, 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 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 taxonomic composition 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 phylogenetic structure and trait variability influence results from different aggregation methods.

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 when aggregation was based on the median, and that trait coding method, uneven phylogenetic structure, and high trait variability can influence aggregated trait affinities. We further showed that harmonised and aggregated data performed similarly to not aggregated and harmonised data from the case study in identifying trait–environment relationships.

4. By identifying 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.

Keywords— trait definitions, trait aggregation, data synthesis

Introduction

Explaining and predicting how communities are shaped by environmental factors is a primary goal of ecology. Species traits are measurable properties of an organism (**mcgill`rebuilding`2006**), and comparing communities based on the traits possessed by species may facilitate testing of a range of ecological hypotheses (**statzner`perspectives`200**). Traits are assumed to be adaptations (e.g., physiological, behavioural) of organisms to their environment, which represent either direct or indirect linkages between the biological response of an organism or a population and its environment (**southwood`habitat`1977**; **verberk`delivering`2013**). In addition to providing local-scale mechanistic-based expressions of species-environment relationships, trait-based approaches might be suitable for large-scale analyses of environmental risks to communities because the occurrence and distribution of traits are less constrained by biogeographic boundaries than that of taxa (**baird`toward`2011**; **bonada`taxonomic`2007**).

Invertebrate traits have been increasingly used in freshwater ecology, e.g. by relating macroinvertebrate trait composition to environmental factors such as climate change or salinisation (**bhowmik`large`2015**; **poff`developing`2010**; **szocs`effects`2014**). In the last decades, freshwater ecologists have compiled comprehensive invertebrate trait databases, which typically include data for a single region or continent (**kefford`integrated`2020**; **Philips`and`Smith`NZ`DB`2018**; **schmidt-kloiber`www.freshwaterecology.info`2015**; **tomanova`trophic`2006**; **ussegliopolatera`biological`2000**; **vieira`database`nodate**). 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`taxonomic`2007** compared trait composition for Mediterranean and temperate regions in Europe based on traits from **ussegliopolatera`biological`2000** (typically referred to as the Tachet database), **poff`developing`2010** characterised trait composition across sites in the Western USA based on traits from **poff`functional`2006**, and **botwe`effects`2018**

used trait definitions from **poiff'functional'2006** and **schafer'trait'2011** to test for effects of salinity on invertebrate traits across different sites in South Australia. Analyses that synthesise invertebrate trait information from more than two different continents are rare, but see **brown'functional'2018** and **statzner'reproductive'1997**.

The variability and inconsistencies of information in freshwater invertebrate trait databases, besides the diversity of taxa across regions, is likely a major reason for the lack of comparisons across continents. Several challenges currently prevent the use of invertebrate trait data from different established databases in ecological studies. One challenge is inconsistency in terminology. In this study, we follow the terminology proposed by **schmera'proposed'2015**, in which a grouping feature (e.g. feeding mode) refers to a set of related traits (e.g. predator, shredder, etc.) that together define an ecological property that is shared among individuals or species. Thus, to use consistent terminology in this study, 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 harmonise 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'towards'2019**). Ideally, the same traits and grouping features would be reported across databases or standardised terminology would exist to easily harmonise them; however, standardised terminology of trait definitions is lacking and poor metadata quality in many trait databases make harmonisation of grouping features difficult (**baird'toward'2011**; **kissling'towards'2018**). To our knowledge, only **brown'functional'2018** harmonised 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 the functional diversity of invertebrate assemblages.

A third challenge to comparing across invertebrate trait databases is inconsistent coding of trait data (**culp'incorporating'2011**). 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 what is being eaten (**moog'comprehensive'nodate**). Some traits can be described in categorical terms, 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. The values derived from the different coding approaches represent affinities (aka affinity scores) that indicate how strongly a taxon expresses a particular trait. The use of different types of coding for trait affinities in different databases limits comparisons. For example, in the study from **brown'functional'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 whereas the North American database used binary coding).

Finally, discrepancies in the taxonomic resolutions (e.g., species, genus, or family) used when comparing among trait databases, or when linking observed taxonomic data to trait databases, presents another challenge. When taxonomic composition or data in on trait database are at a more precise taxonomic level than other trait 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. **szocs'effects'2014**; **vos'taxonomic'2017**). If trait information is only available at more precise taxonomic levels than for taxonomic composition, traits are aggregated to a less precise taxonomic level (e.g. **aspin'extreme'2019**; **piliere'a.f.h.'importance'2016**; **poff'functional'2006**; **szocs'effects'2014**). To date, studies have used different methods of trait aggregation, e.g. the mean (**magliozzi'functional'2019**), median (**szocs'effects'2014**) or mode (**piliere'a.f.h.'importance'2016**), but studies on how and to what extent different trait aggregation methods influence trait-based analyses are missing.

These challenges motivated us to assess the characteristics and comparability of

different freshwater invertebrate trait databases with the goal of facilitating future syntheses. We aimed to identify 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 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 grouping features for comparison across geographic locations; and (3) by comparing the influence of different trait aggregation methods and grouping-feature harmonisation on the detection 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. Discrepancies in trait definitions served as a starting point to develop four novel invertebrate trait datasets by harmonising seven grouping features present in the existing trait databases. The harmonised trait datasets were used to evaluate the influence of different trait aggregation methods 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 phylogenetic structure 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 (**szocs·effects·2014**) and compared our results with those of the original study.

Selection of traits and harmonisation of trait databases

We assessed how six databases differed in how they assigned and described traits of freshwater invertebrates. These databases (from Europe, North America, Australia, and New Zealand) differed in how grouping features were defined and the taxonomic resolution used to assign traits. For Europe we obtained trait information from the *freshwaterecology.info* database (**schmidt-kloiber·www.freshwaterecology.info·2015**), which assigns traits at the species level, and the Tachet database (**ussegliopolatera·biological·2000**), which assigns traits at species, genus, and family levels. For North America we primarily obtained trait information from **twardochleb·freshwater·nodate** (hereafter referred to as the CONUS database) and where taxa were missing from that database, we obtained data from **vieira·database·nodate** (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 (**polatera'personal'information'2020**). For Australia and New Zealand, we used trait databases from **kefford'integrated'2020** (hereafter referred to as the Australian database) and **Philips'and'Smith'NZ'DB'2018** (hereafter referred to as the New Zealand database). These databases assigned traits at species, genus, family, and coarser 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 aspects 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 7 grouping features that collectively included 26 traits (Table ??). To do so we combined similar traits into one trait (e.g. crawlers and sprawlers into crawlers) by assigning for each taxon for the new combined trait the highest affinity score of those originally assigned. 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 used either fuzzy coding or, when unavailable, binary coding to create the final harmonised datasets. 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, therefore, had the same range as fuzzy coded traits. For example, in the *freshwaterecology.info* database, voltinism traits are associated with 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 lowest 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 (**polatera'personal'information'2020**).

‡ 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'biomonitoring'2000**). Taxa exhibiting this trait have been assigned to predators or herbivores based on expert knowledge (**polatera'personal'information'piercer'2020**).

Comparing trait aggregation methods

To compare the effects of different aggregation methods on trait-environment relationship inferences, 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, in which we aggregated either trait affinities by the median or mean for species and genera within a family, giving equal weight to every taxon. We denote these aggregation methods as *direct_agg_median* and *direct_agg_mean*, respectively; and (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 ??). 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’dissolved-oxygen’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 \times 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 \times 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 is a continuous variable in the Australian database, so, we converted size into three traits (small, medium, and large) each with a binary coding. Because trait affinities ranged from 0 to 1, the maximum difference possible in trait affinity was 1 to -1 , both of which corresponds to 100 %. For convenience and to improve interpretation, we calculated absolute values of the trait differences, e.g. a difference of -1 was converted to 1.

Effect of phylogeny and trait variability on aggregation outcomes

To examine the influence of phylogeny and trait variability on the outcomes of the different trait aggregation methods, we created three hypothetical scenarios. We simulated three different families, each containing 25 total species but with different phylogenetic 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 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 \times 5 levels of trait variability \times 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 phylogenetic scenario

as well as the differences in trait affinities obtained by each aggregation method.

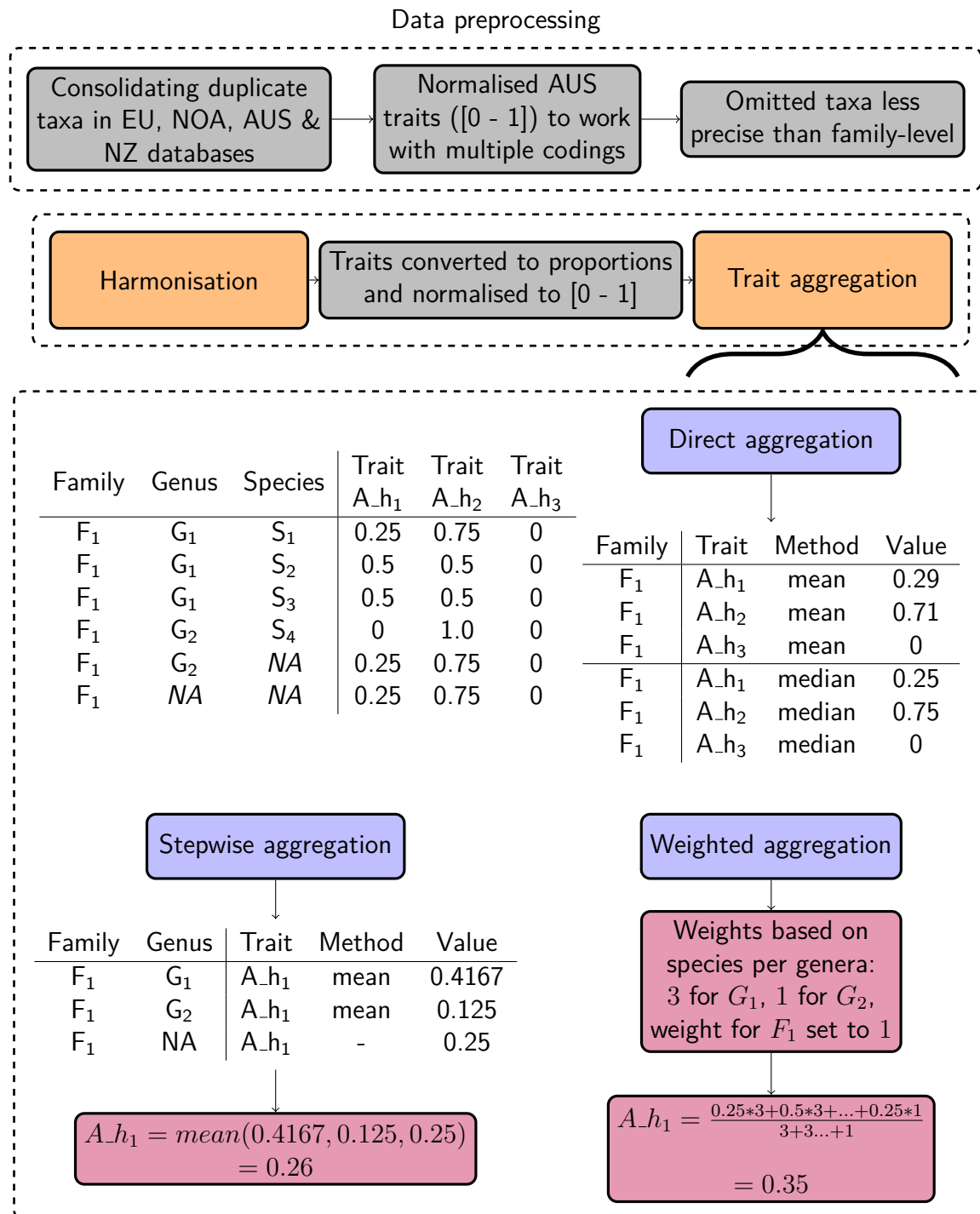


Figure 1: Data processing steps of the selected traits. Intermediate (grey) and main (orange) steps of data preparation are depicted. The dashed bottom box illustrates the different trait aggregation methods (hypothetical data in the upper left corner). The aggregation methods (purple) and intermediate 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 inferences regarding trait-environment relationships

To further evaluate the effects of grouping-feature harmonisation and trait aggregation on characterisation of trait–environment relationships, we repeated the analysis of **szocs’effects’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 (**bathe’biological’2011**). **szocs’effects’2014** used redundancy analysis (RDA) to compare trait compositions among sites downstream and upstream to the salt discharge. Details about the original study can be found in **szocs’effects’2014**.

We re-analysed the **szocs’effects’2014** data 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 the established 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, trait composition, expressed as community weighted mean (CWM) traits, was ordinated along an electrical 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, to test for the effect of different aggregation methods on inferences about CWM trait–salinisation relationships, we repeated the RDA five times, each with family-level aggregated trait values assigned to corresponding taxa from **szocs’effects’2014**.

Software used in data analyses

All data processing and RDA analyses were carried out using R version 3.6.1 (**cite’R**). RDA was computed using the **vegan** package (**cite’vegan**). 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 ??) 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, North American, and New Zealand datasets contained the 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
EUR	4601	3942 (86)	3739 (81)	704 (15)	158 (3)
NA	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: EUR, Europe; NOA, North America; AUS, Australia; 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 ??). By contrast, the New Zealand dataset contained complete trait information for most of the investigated grouping features (94–100%).

The number of entries with available information for the selected grouping features varied strongly across the established European, North American, and Australian datasets (Table ??). In contrast, the New Zealand dataset contained com-

plete 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
EUR	8	15	23	36	11	57	76
NA	28	13	47	52	73	44	63
AUS	4	46	49	39	75	68	99
NZ	100	94	100	99	100	100	99

Abbreviations: EUR, Europe; NA, North America; AUS, Australia; NZ, New Zealand.

Discrepancies in invertebrate trait definitions across databases

Definitions of grouping features and traits varied in their level of detail in the six 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'integrated'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 (see Table ?? for a summary of discrepancies in trait definitions). Varying levels of differentiation among the trait databases were present in all investigated grouping features (Table ?? and Table ??).

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 feeding mode in the *freshwaterecology.info* database was primarily described in terms of type of food (except for filterers), whereas the other databases described feeding mode in terms of 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 the basis for their definitions. The Tachet database captured the food source in a separate 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, the Tachet database as substrate relation and locomotion, the Vieira database as how organisms deal with flow, the Australia database as attachment, and the CONUS database as not only the method of movement but also the location of movement. Similarly, databases differed in the way they characterise 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 ??).

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;	7 traits; fuzzy [0 - 3]	6 traits; binary	8 traits; binary	16 traits [†] ; binary, proportional fuzzy [0 - 3]	6 traits; fuzzy [0 - 3]
	10 point assignment system					
Voltinism	6 traits;	3 traits; fuzzy [0 - 3]	3 traits; binary	3 traits; binary	7 traits; binary, proportional fuzzy [0 - 3]	3 traits; fuzzy [0 - 3]
	single category assignment system					
Locomotion	6 traits;	8 traits; fuzzy [0 - 5]	10 traits; binary	9 traits; binary	9 traits; binary, fuzzy [0 - 3]	4 traits; fuzzy [0 - 3]
	10 point assignment system					
Respiration	7 traits;	5 traits; fuzzy [0 - 3]	3 traits; binary	8 traits; binary	10 traits; binary, proportional fuzzy [0 - 3]	4 traits; fuzzy [0 - 3]
	binary					
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*, *Detritivore*, 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 ??). 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 ??).

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.

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

Abbreviations: Min., Minimum; Max., Maximum; abs., absolute; SD, Standard deviation; AUS, Australia; NA, North America.

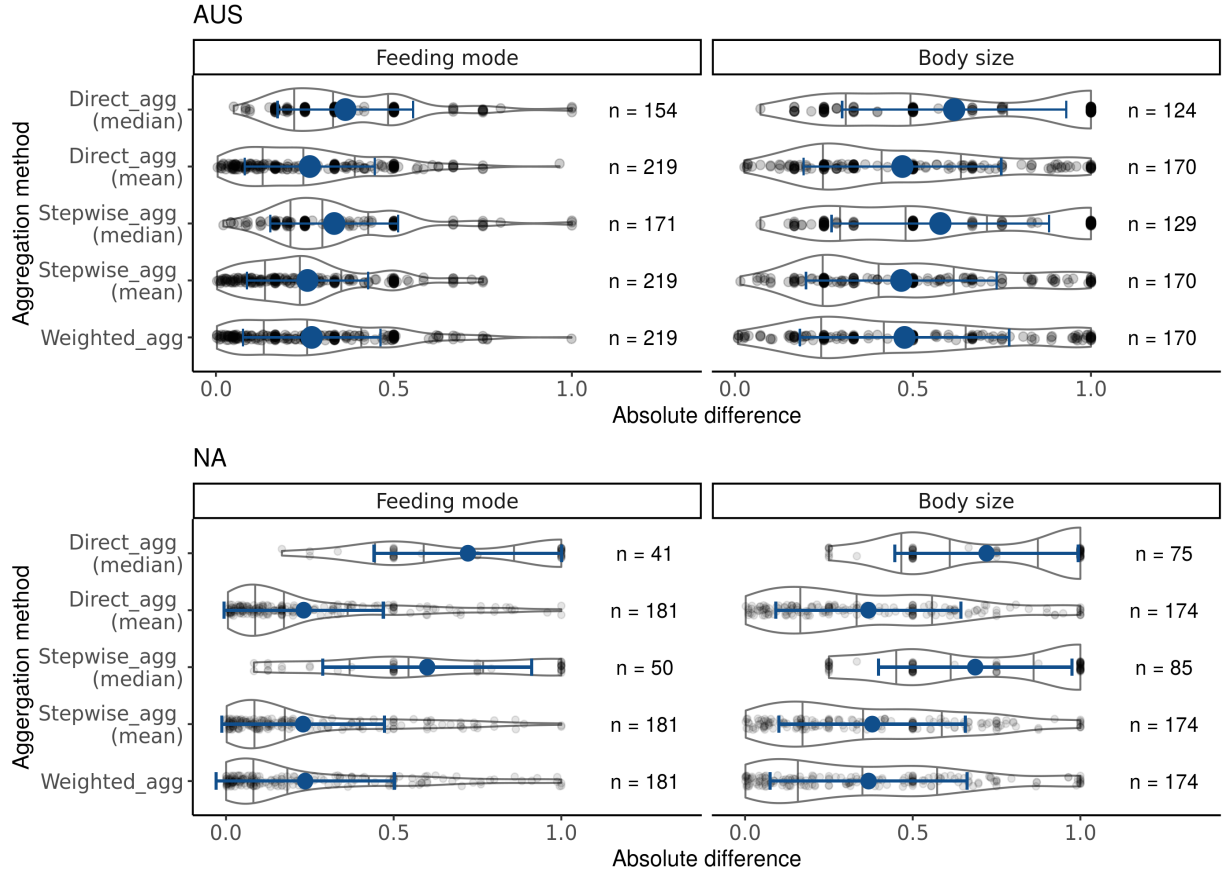


Figure 2: Cases (factor combination of investigated families and traits) where differences occurred between aggregated traits and expert assigned traits at family level for the Australian and North American dataset. Violin plots - mirrored density plots - show the density of the absolute trait affinity 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 ??). 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; NA, North America.

Comparison of aggregation methods with varying phylogenies and trait variability

The simulations showed that both phylogenetic 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 phylogenetic structure differed across aggregation methods, but, as we expected, the range of trait affinities increased with increasing trait variability for all aggregation methods (Figure ??).

Phylogenetic 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 phylogenetic 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 phylogenetic structure and trait variability to some degree, in most simulated datasets, the different aggregation methods resulted in similar trait affinities. Only 1.4 %, or 213 out of 15,000 total comparisons (3 scenarios \times 5 levels of trait variability \times 10 unique comparisons of trait aggregation methods \times 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 ??).

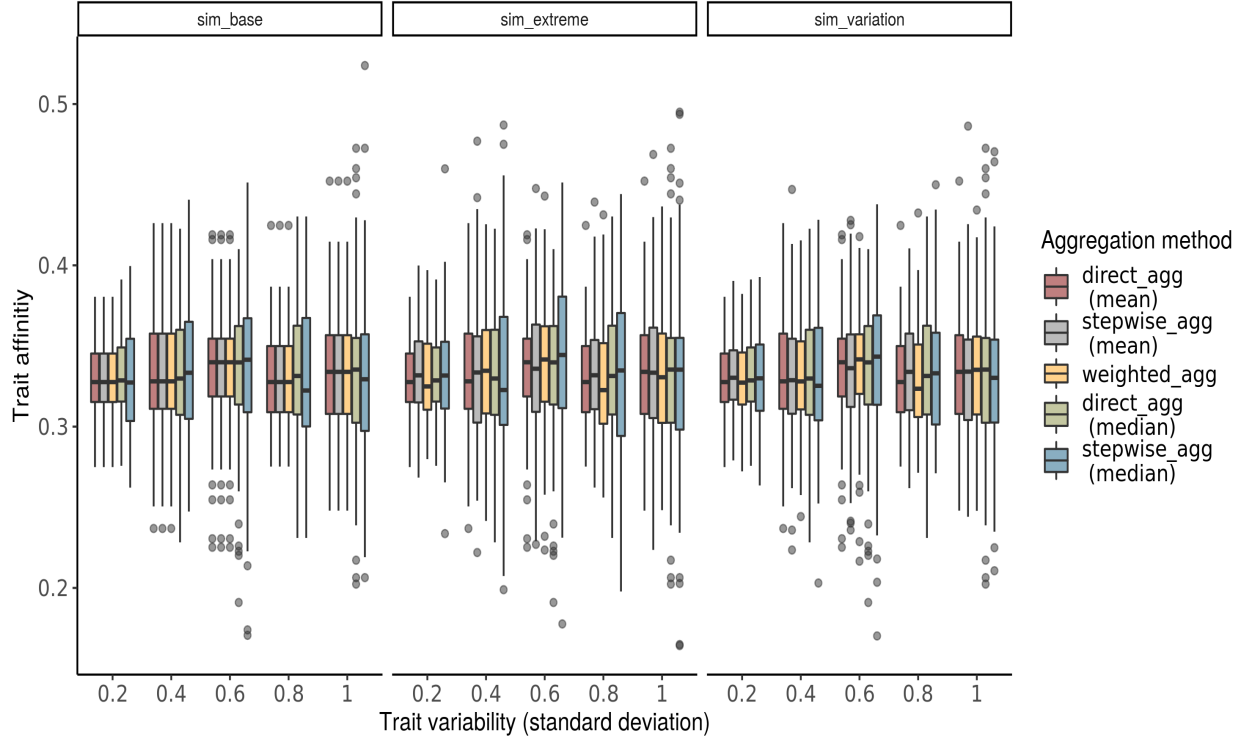


Figure 3: Ranges of aggregated trait affinities for the three examples of taxonomic hierarchies and simulated levels of trait variability. Shown are the results only for one simulated trait (T1). Similar results were obtained for the other simulated traits (Figure ??). Boxplots depict results for 100 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 $1.5 \times$ 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.

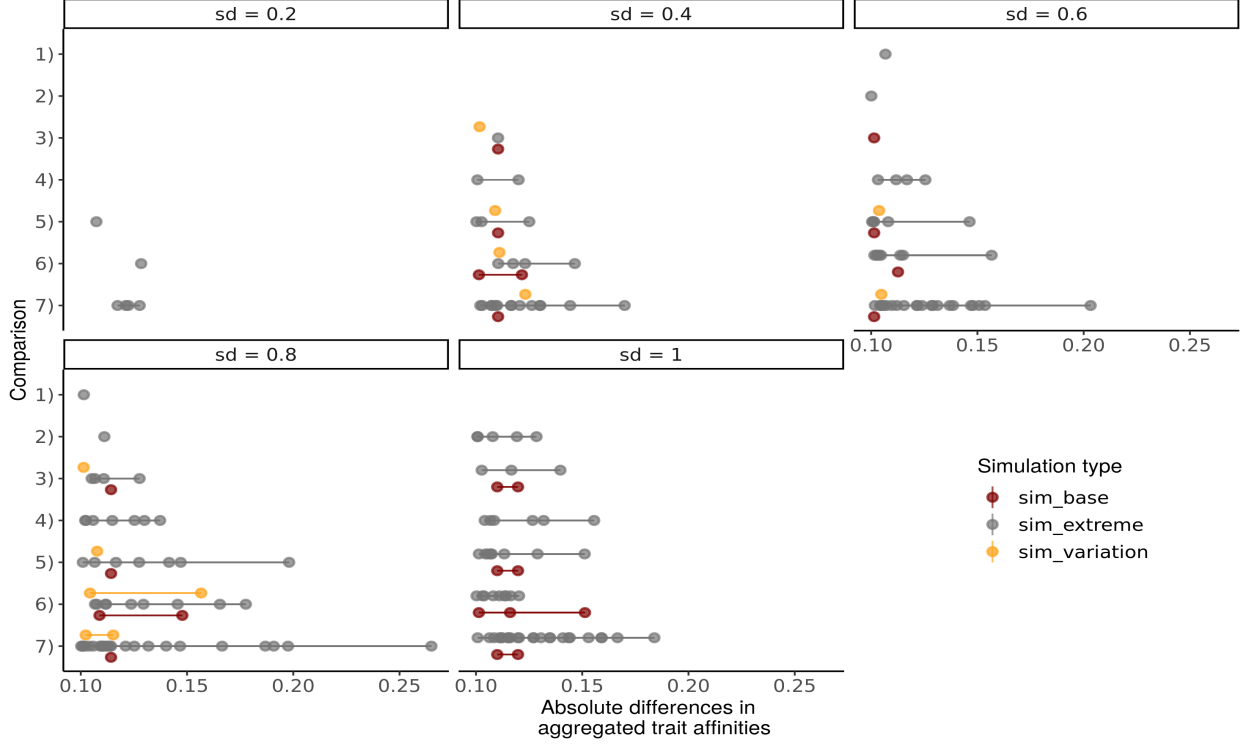


Figure 4: Comparison between the aggregated trait affinities produced by the different trait aggregation methods for every simulated dataset across all 3 simulated traits (5,000 comparisons per scenario). Results are shown for different levels of trait variability which is represented by the standard 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 inferences regarding trait-environment relationships

Our re-analysis of the **szocs'effects'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 (??) compared with the original analysis, and they resulted in similar, but fewer, traits that distinguished upstream and downstream sites than the original analysis (Figures ?? and ??).

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 ?? and ??). The original RDA in **szocs'effects'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 the 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 ??).

RDA based on 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 ?? and ??). 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.

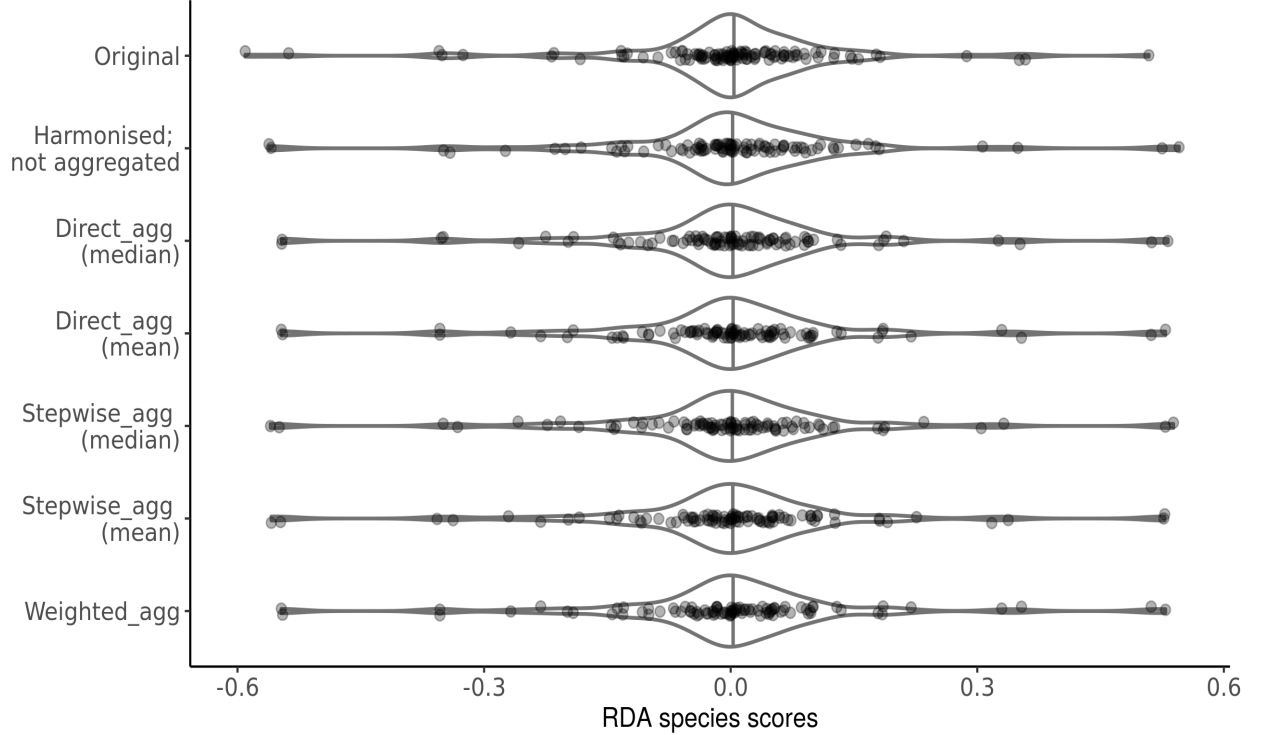


Figure 5: Species scores obtained by RDA from the original analysis (**szocs'effects'2014**), using harmonised grouping features, and using harmonised grouping features with traits aggregated to family level with five different aggregation methods. The dots represent the individual species scores for 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.

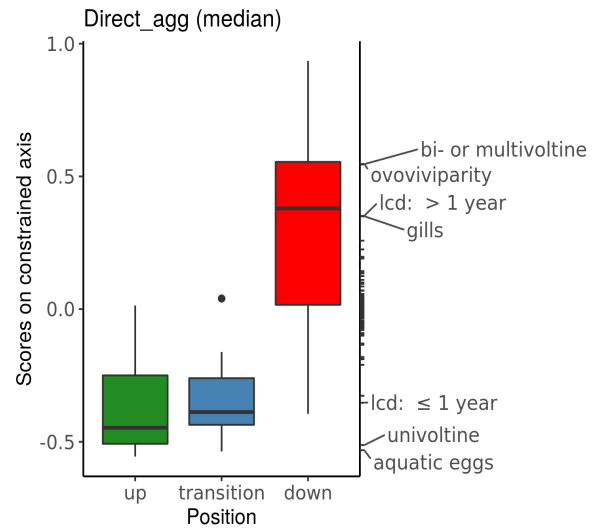
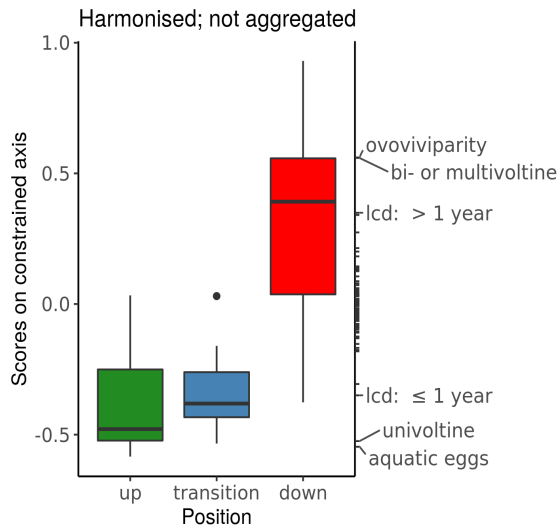
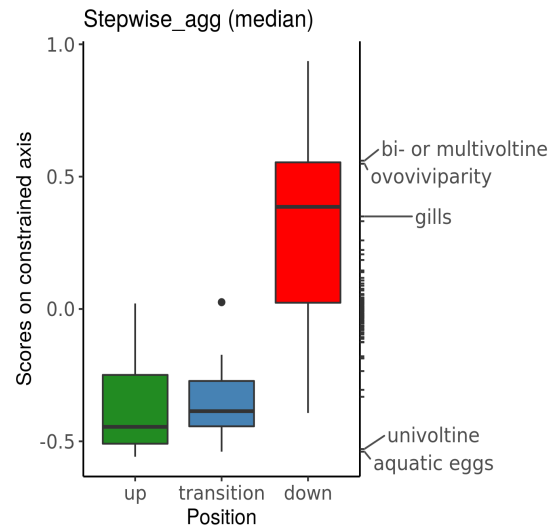
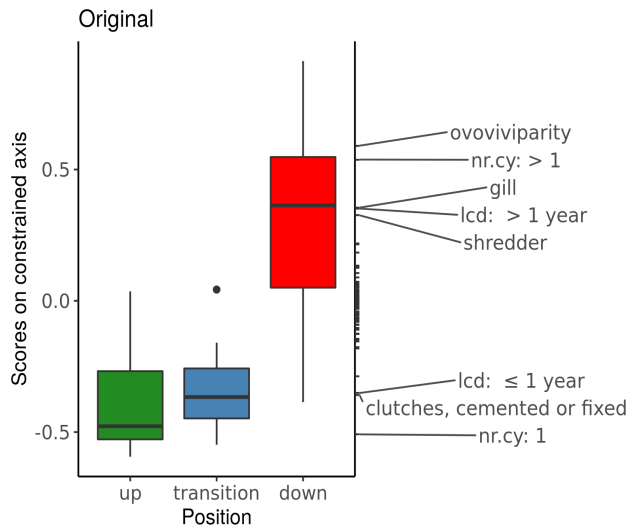


Figure 6: RDA of traits constrained by electric conductivity for harmonised but not aggregated data, data aggregated with *stepwise_agg* *median* (largest difference in terms of responsive traits to salinity to the original study), data aggregated with *direct_agg* *median* (smallest difference in terms 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 ??). 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 can greatly limit the use of combined databases. 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 generally effective 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

Synthesising 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'toward'2011**; **brink'traits-based'2011**), but in this study we show for the first time the discrepancies that exist across some commonly used grouping features. Furthermore, we offer 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, extensive information was 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 characterise.

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'towards'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 should encourage 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'toward'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'towards'2019**). Once standardised trait definitions are established, they 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

The taxonomic resolution used to describe the occurrence and abundance of freshwater invertebrates can be highly variable because species-level identifications require specialists and are also often time-consuming and expensive (**marshall'taxonomic'2006; resh'which'2008**). Consequently, trait information at the family level has been

widely used by freshwater ecologists, for example in bioassessment (**beketov'spear'2009**). However, traits assigned at the family level may not reflect the full trait diversity that exists within, for example, large, ecologically diverse families such as Chironomidae (**serra'synthesising'2016**). Additionally, various trait aggregation methods have been used by researchers to compare trait profiles based on 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 the species or genus level to a single estimate at the family level implies 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 resolved taxonomic levels than the family level. For example, **monaghan'improving'2013** found high intra-family trait dispersion in the Tachet database for traits within the grouping features body size, flow preference, and reproduction. Furthermore, 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'functional'2006**; **wilkes'traitbased'2020**).

If trait aggregation is necessary, however, our study offers insights 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 is probably caused by the use of binary

coding 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 resulted in the North American dataset either to agreement with the assigned traits or to more substantial differences compared to the use of 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 measures, 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 phylogenetic structure. However, we found through simulations that 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 approaches. In the simulation, taxonomic structural unevenness appeared to produce some variation in affinities among weighting approaches, especially for a family where one genus had a much larger number of species than its other genera. The comparison between aggregated and assigned traits showed, that 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 might 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 ??). 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 ??). 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 phylogenetic 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 **szocs’effects’2014** with our harmonised and family-level aggregated datasets, we 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 inferences regarding trait–environment relationships based on 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’spear’2009**). However, some of the traits that had responded in the original analysis did not respond in the re-analysis, specifically 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. The shredder feeding mode 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, in which this trait was combined based on three traits (miner, xylophagus, and shredder) used in the original databases. Consequently, the trait affinities in the original data had a higher mean and standard deviation than in the harmonised and aggregated data (Table ??), suggesting that the signal in the original data was weakened by harmonisation. Given 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 also 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 help of new methods such as computer vision for species and trait identification (**hoye'deep'2020**), more observed taxa will be identified to the species level and trait data will be available at the species level 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. Based on this analysis, we cannot claim that a specific aggregation method is objectively the best, and except for cases where traits are highly variable and phylogenetic structure is highly uneven, the choice of aggregation method does not appear to be influential to the outcome. Weighting trait information based on the taxa present in the database used does not seem to have much influence except when the phylogenetic 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. Use of carefully harmonised and integrated trait data as described here should enhance our ability to study a wide range of large-scale ecological questions with greater rigor than currently exists.

Acknowledgments

We thank Brooke Cassell for technical editing of the manuscript. The project, in which this study was conducted, is funded by the German research society (DFG, project number 338785727).

Data availability statement

Harmonised invertebrate trait datasets for Australia, Europe, North America and New Zealand are deposited in the Github repository: <https://github.com/KunzstLD/Trait-aggregation>.

Conflict of interest

The authors declare no conflict of interest.

Author contribution

R.B.S. and S.K. designed the study; B.J.K. and L.M. provided trait data for Australia; P.U-P. provided body form trait data for European and North American invertebrates; C.P.H., N.L.P., and L.T. provided trait data for North America; C.D.M. provided trait data for New Zealand; B.J.K., A.S-K., C.D.M, W.G., N.L.P., L.M., and C.P.H. provided in depth information on trait data and discussed the procedure of analysis; S.K. conducted data processing and the statistical analyses.; S.K. wrote

the manuscript; B.J.K, A-S.K., C.D.M, P.U-P., C.P.H., and R.B.S. discussed the results and provided suggestions on the manuscript.

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 <i>animal material</i> "	<ul style="list-style-type: none"> • "Shred decomposing vascular plant tissue" • Trait herbivore includes among others insect that shred <i>living aquatic plants</i> 	Shredder	<ul style="list-style-type: none"> • Detrivore[†] • Trait herbivore includes among others the trait shredder 	Shredders
------------------	--	--	---	----------	---	-----------

Feeding predator	"Eating prey"	<ul style="list-style-type: none"> • Carvers, engulfers & swallowers • Piercers (plants & animals) are an additional trait 	Engulfers ("ingest whole or in parts") & piercers ("prey tissues and suck fluids")	Predator	Piercer & engulfer	Predator
Feeding filter-feeder	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
Semivoltine	"One generation in two years"	"Life cycle lasts <i>at least</i> two years"	"< 1 generation per year"	"< 1 generation per year"	"< 1 generation per year"	"< 1 reproductive cycle per year"

Multi-voltine	"Three or more generations per year" [†]	"Able to complete two successive generations per year"	"> 1 generations per year"	"> 1 generations per year"	<ul style="list-style-type: none"> • 1-2 generations per year • bi/multivoltine > 1 reproductive cycles per year" • up to 5 generations per year • up to 10 generations per year
---------------	---	--	----------------------------	----------------------------	---

Locomotion swimming	<ul style="list-style-type: none"> • Passive movement like floating or drifting (trait swimming/scating) • Active movement (trait swimming/diving) 	<ul style="list-style-type: none"> • Surface swimmers (over and under the water surface) • Full water swimmers (e.g. Baetidae). 	”Adapted for ”fishlike” swimming”	Swimmer	Distinguishes swimmer and skater	Swimmers (water column)
---------------------	--	---	-----------------------------------	---------	----------------------------------	-------------------------

Locomotion burrowing	"Burrowing in <i>soft</i> substrates or boring in <i>hard</i> substrates"	<ul style="list-style-type: none"> Burrowing "within the first centimeters of the benthic fine sediment" Differentiates also the trait interstitial (endobenthic) 	"Inhabiting <i>fine</i> 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 crawler, sprawler, climber 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	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

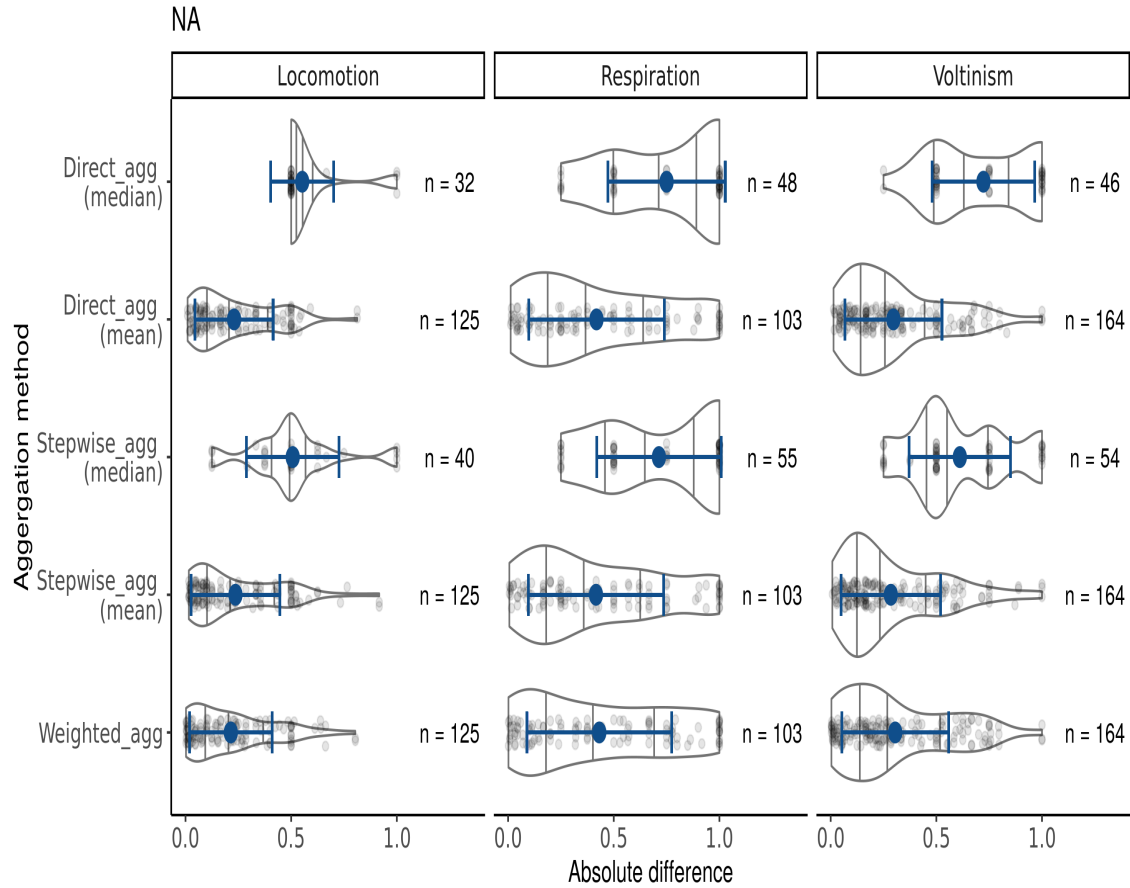


Figure S1: Cases (factor combination of investigated families and traits) where differences occurred between aggregated traits and expert assigned traits at family level for the North American dataset. Violin plots - mirrored density plots - show the density of the absolute trait affinity differences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

Comparison of aggregation methods with varying taxonomic hierarchies and trait variability

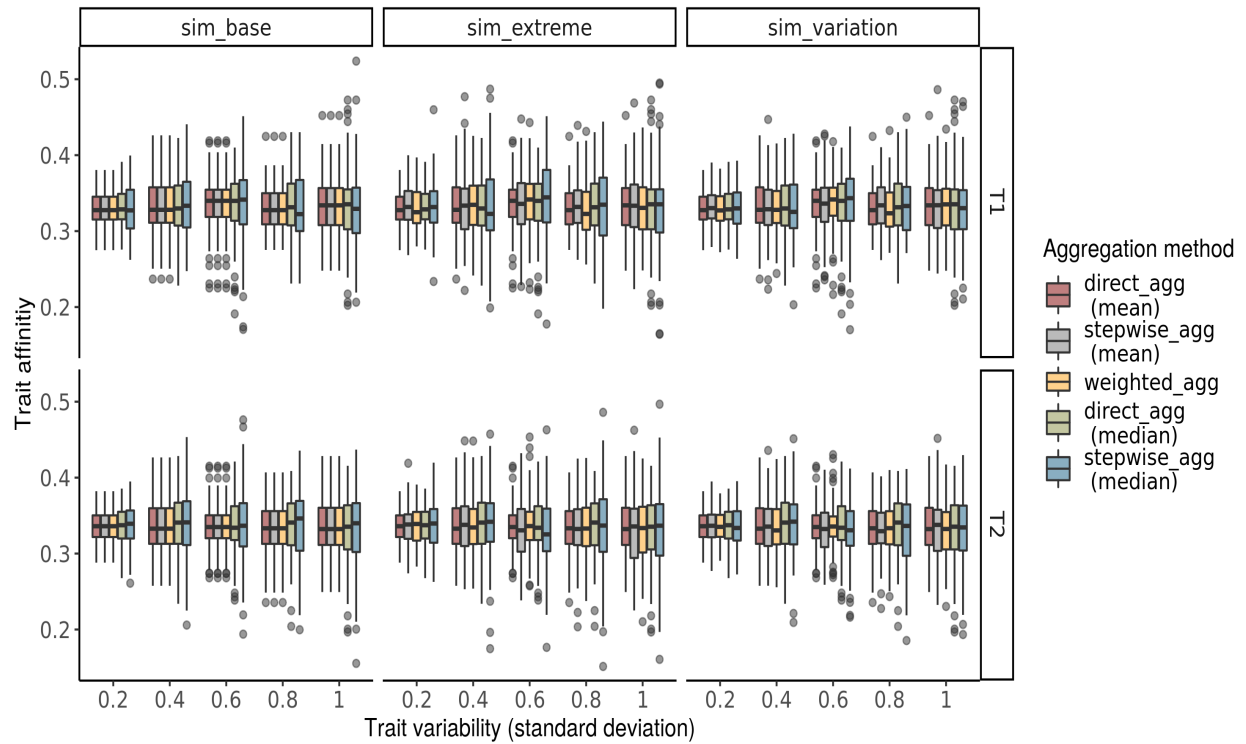


Figure S2: Ranges of aggregated trait affinities for the three examples of taxonomic hierarchies and simulated levels of trait variability. Shown are the results for the simulated traits T2 and T3. Boxplots depict results for 100 replicated simulations of each trait aggregation method. Trait aggregation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

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

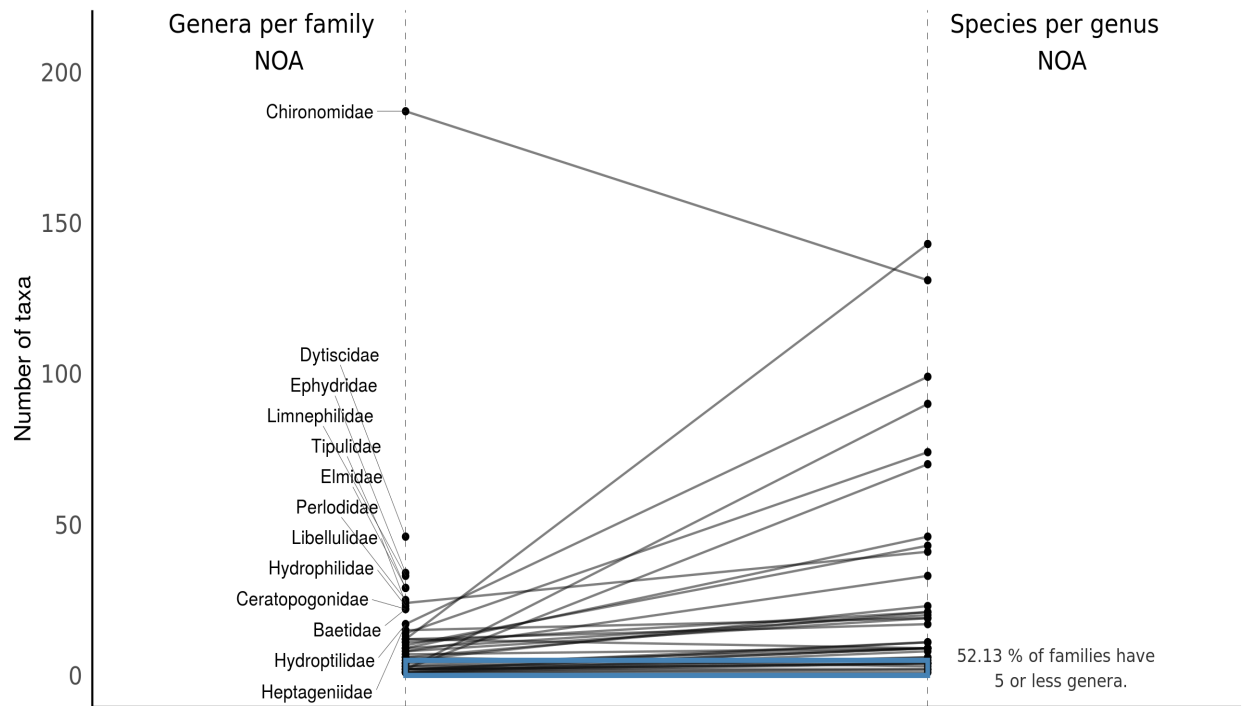


Figure S3: Number of genera per family and species per genus for those families of the North American trait dataset that have been compared with assigned traits at family level. For better visual display only families with more than 15 genera are displayed.

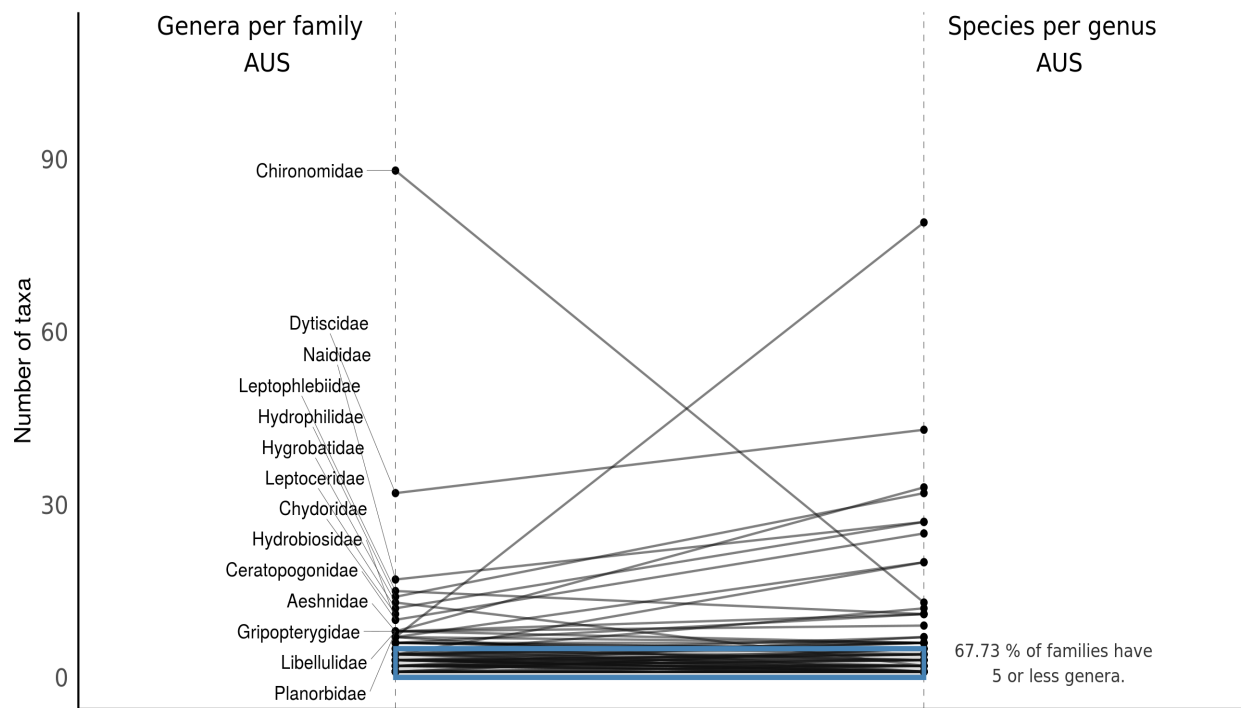


Figure S4: Number of genera per family and species per genus for the Australian trait dataset. For better visual display only families with more than 7 genera are displayed.

Effects of harmonisation and trait aggregation on inferences regarding 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

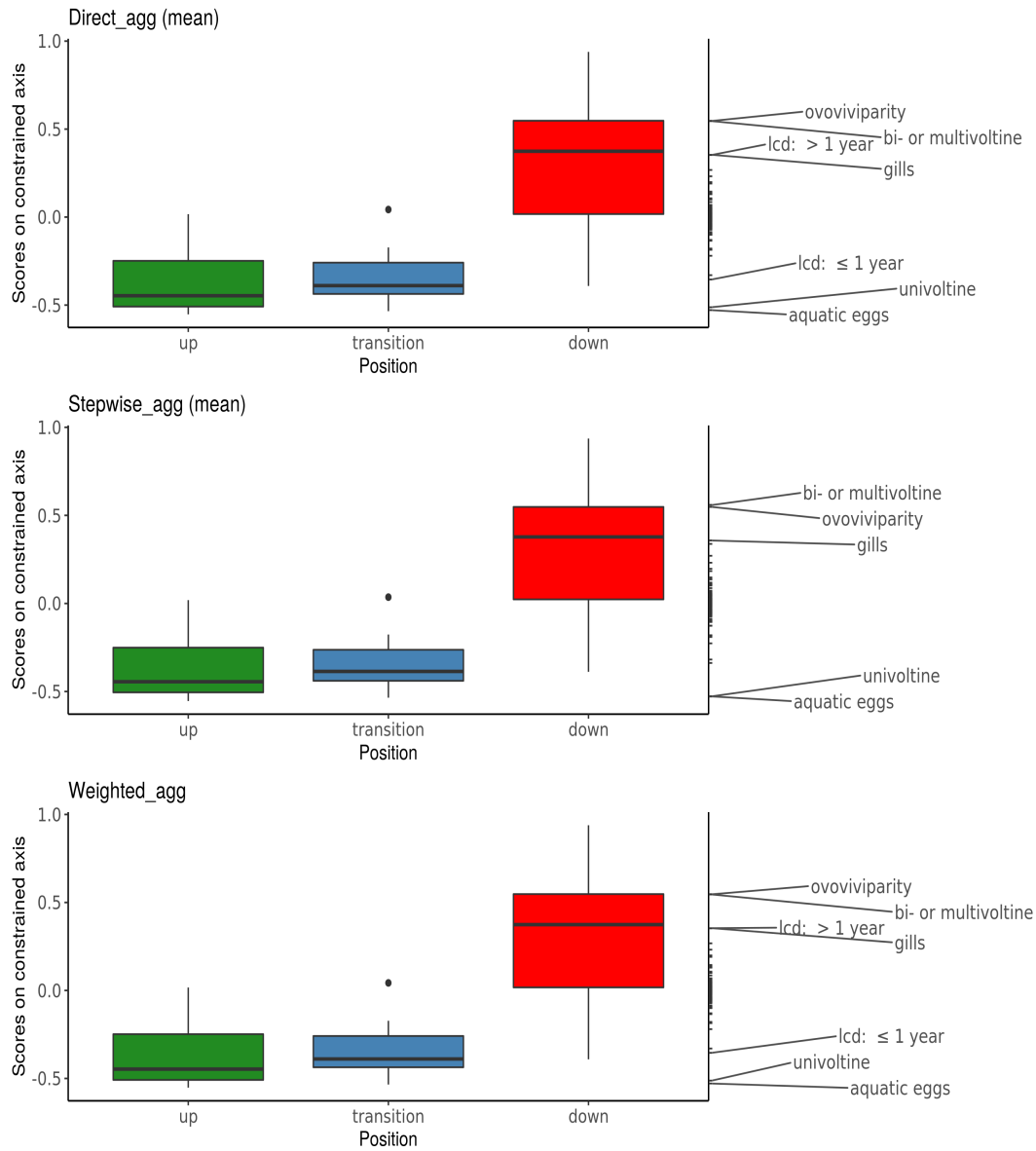


Figure S5: RDA of traits constrained by electric conductivity for the data aggregated with *direct_agg* *mean*, *stepwise_agg* *mean*, and *weighted_agg*. Shown are boxplots of the site scores along the conductivity axis. The rug on the right side of each plot indicates species scores of the traits on the conductivity axis. For more details see Figure ??.

Abbreviations: lcd, life cycle duration; nr.cy, potential number of cycles per year.