

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

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, is that 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'rebuilding'2006**), and comparing communities based on the traits possessed by species may facilitate testing of a range of ecological hypotheses (**heino'jani'macroecological'2013**). 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'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 because trait responses are less constrained by biogeographic boundaries and taxon distributional areas than taxonomic responses (**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 (**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** characterized trait composition across sites in the Western USA based on traits from **poff'functional'2006**, and **botwe'effects'2018** used trait definitions from **poff'functional'2006** and **schafer'trait'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'functional'2018** and **statzner'reproductive'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'proposed'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'towards'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'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 functional diversity and invertebrate assemblages.

A third challenge to comparing across invertebrate trait databases is inconsistent coding of trait data (**culp'incorporating'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 mouth-part morphology, feeding behaviour, and specific food sources (**moog'comprehensive'nodate**). 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'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, 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. **szocs'effects'2014**; **vos'taxonomic'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'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 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 trait 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 (**szocs’effects’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’www.freshwaterecology.info’2019**), which contains taxa at the species level, and the Tachet database (**ussegliopolatera’biological’2000**), which contains taxa 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’Smit’NZ’DB’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 ??). 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 (**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

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 ??).

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'2000**) 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.

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.

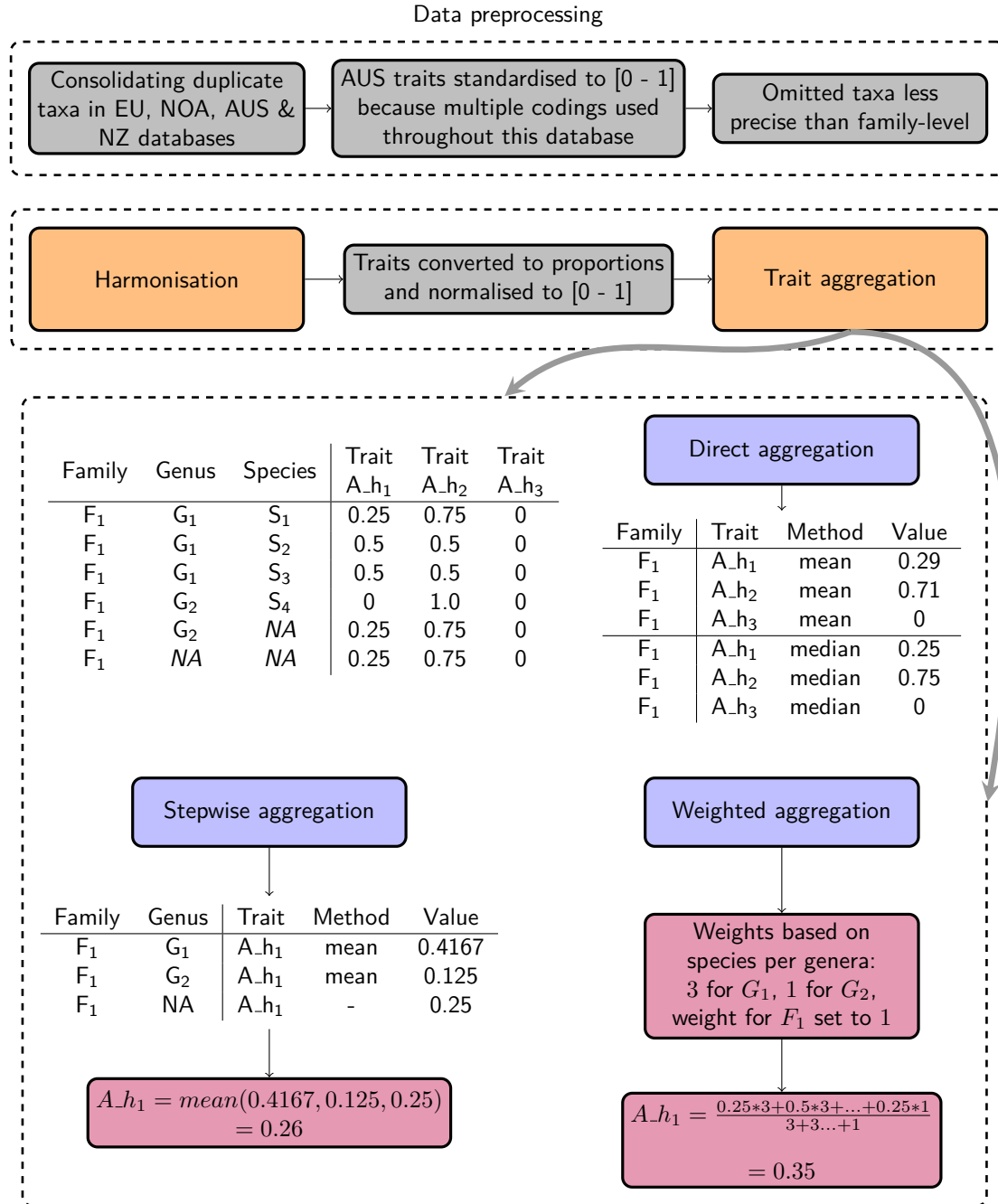


Figure 1: Data processing steps of the selected traits. Intermediate (gray) and main (orange) steps of data preparation are depicted. The dashed bottom box illustrates the different trait aggregation methods (hypothetical data in the upper left corner). The aggregation methods (purple) and intermediate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

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

Aggregated trait affinities of the five trait aggregation methods (*direct_agg_{median}*, *direct_agg_{mean}*, *stepwise_agg_{median}*, *stepwise_agg_{mean}*, and *weighted_agg*) were compared to trait affinities assigned at family-level by experts, which were available for Australia and North America for a subset of grouping features and taxa. For the Australian dataset, we compared aggregated trait affinities with assigned trait affinities resolved at family-level for the grouping features feeding mode and size using data from **chessman’dissolved-oxygen’2018**. In **chessman’dissolved-oxygen’2018** feeding mode is classified similarly as in the Australian dataset except that the trait parasite is missing. We conducted the comparison for the 220 families available in **chessman’dissolved-oxygen’2018**. Considering each factor combination of family and trait (in total 8) this amounted to 1760 cases.

For the North American dataset, we compared aggregated trait affinities with assigned trait affinities on family-level for the grouping features feeding mode, respiration, size, voltinism, and locomotion. The assigned trait affinities at family-level are part of the CONUS database and originate from expert knowledge. Trait information was available for 94 families of which all were present in the North American dataset (total number of cases 1598). The traits were on the categorical scale and were converted to binary traits prior to the comparison with aggregated trait affinities.

As mentioned above, trait affinities ranged from 0 to 1. Hence, the maximum difference possible in trait affinities is 1 or -1 (corresponds to 100 %). For convenience and to improve interpretation, we report absolute trait differences.

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 **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, upstream, and close to salt discharge (transition). Further details about the original study can be found in **szocs’effects’2014**.

We re-analysed the **szocs’effects’2014** study by performing a RDA with harmonised grouping features from the dataset for Europe and additionally 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 **szocs’effects’2014**.

Data analysis

All data processing and RDA analyses were carried out using R (**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, 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 ??). 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 ??). 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'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. We provide a summary of discrepancies in trait definitions in the appendix (Table ??). 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 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 ??).

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 assignment system	7 traits; fuzzy [0 - 3]	6 traits; binary	8 traits; binary	16 traits [†] ; binary, proportional 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 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 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*, *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 overall approach (direct, stepwise, or weighted) 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 and weighted 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 ?? and 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.

Database	Comparison to traits at family level	Differing cases [%]	Min. differences	Max. differences	Mean abs. differences	SD abs. differences
Australia	<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
North America	<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

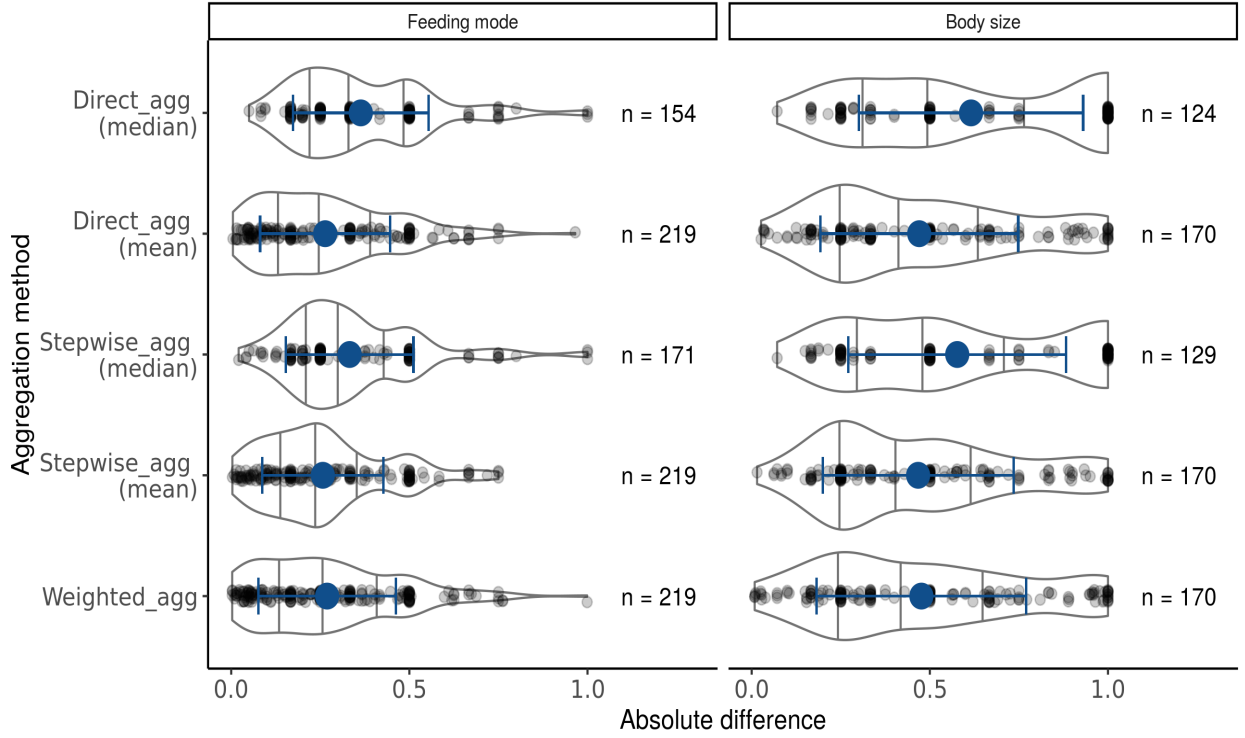


Figure 2: Cases (factor combination of investigated families and traits) where differences occurred between aggregated traits and expert assigned traits at family level for the Australian dataset for the grouping features feeding mode and body size. Violin plots - mirrored density plots - show the density of the absolute trait affinity differences. 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.

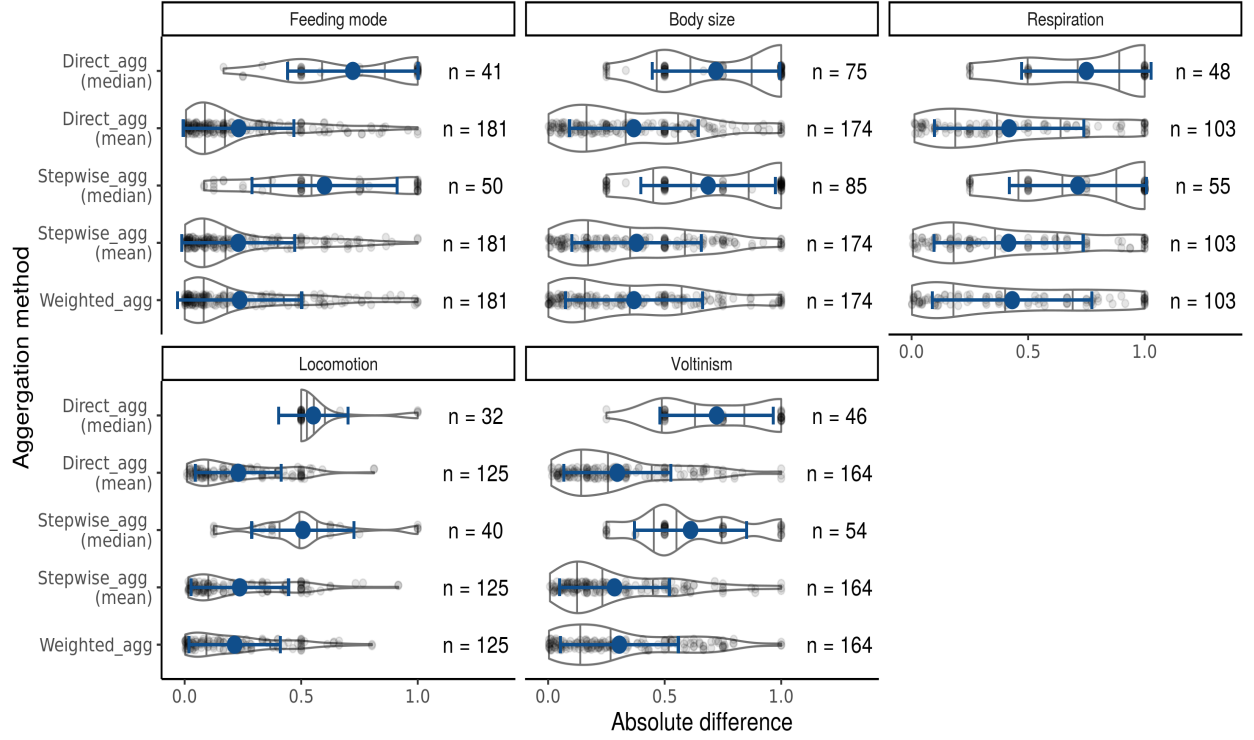


Figure 3: Cases (factor combination investigated families and traits) where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the North America dataset for the grouping features feeding mode, locomotion, respiration, body size and voltinism. For more details see Figure ??.

Simulation of varying taxonomic hierarchies

We evaluated the simulation results based on the produced range of aggregated trait affinities and by comparing results of every simulated dataset.

As expected the produced range of trait affinities increased for all aggregation methods with increasing trait variability (Figure ??).

For the *sim_base* scenario, the mean aggregation methods yielded similar ranges of aggregated trait affinities within each level of trait variability whereas the median aggregation methods produced mostly greater ranges of aggregated trait affinities. By contrast, the ranges of trait affinities slightly differed for all aggregation methods for the simulation scenarios *sim_extreme* and *sim_variation*. *Weighted_agg* and *stepwise_agg_mean* produced a wider range of trait affinities than the *direct_agg_mean* in the *sim_extreme* scenario. For most levels of trait variability the *stepwise_agg_median* resulted in the largest range of trait affinities in the *sim_extreme* and *sim_variation* scenarios.

Each trait aggregation method yielded one aggregated affinity for every simulated dataset, in total 15000 comparisons (100 replicates * 5 levels of trait variability * 10 unique comparisons of trait aggregation methods * 3 scenarios). In most simulated datasets the different aggregation methods often resulted in similar trait affinities. Only 1.42 % (213 out of 15.000) of all comparisons showed a difference equal or greater than an absolute trait affinity of 0.1. Most of these differences occurred for the *sim_extreme* scenario (83.5 %). The majority of the differences equal or above 0.1 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 ??).

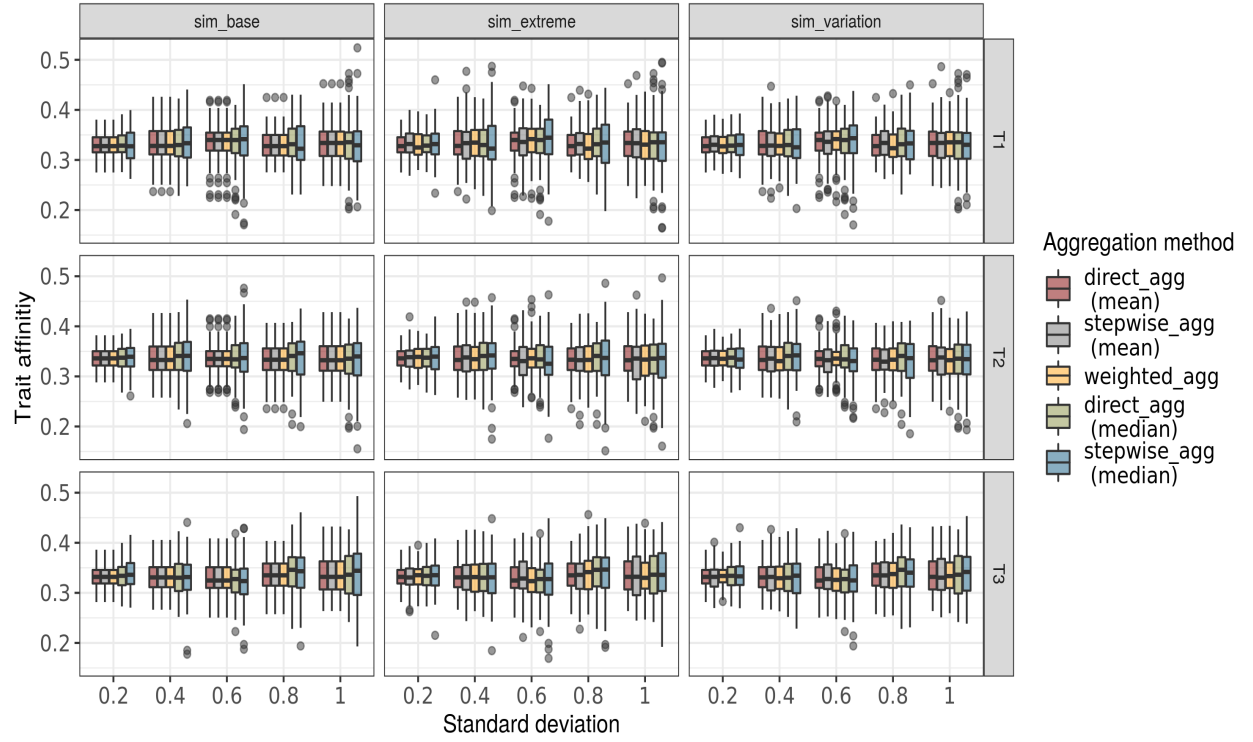


Figure 4: Ranges of aggregated trait affinities for the three examples of taxonomic hierarchies and simulated levels of trait variability. Boxplots depict results for each trait aggregation method of 100 simulations. T1, T2, and T3 are the simulated traits.

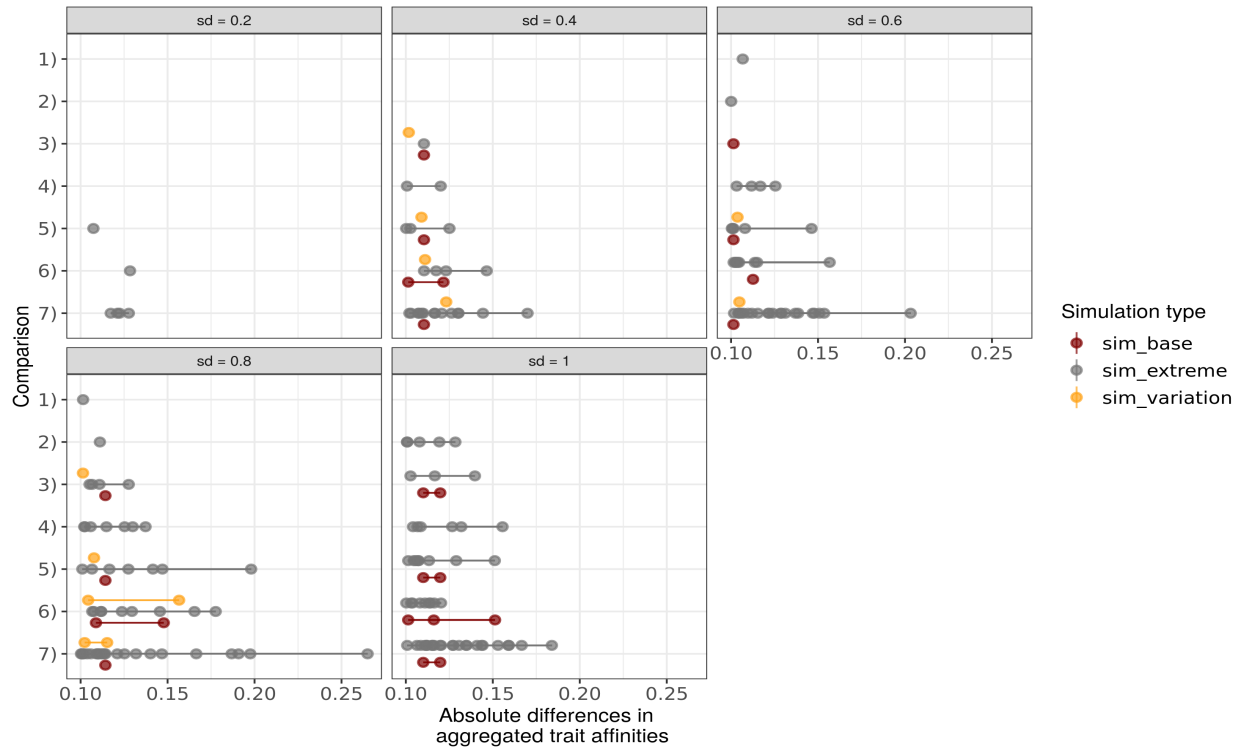


Figure 5: Comparison of the aggregated trait affinities produced by the different trait aggregation methods for every simulated dataset across all 3 simulated traits. Dots depict comparisons where absolute differences between aggregated trait affinities were greater than 0.1.

Comparison:

- 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

Table 6: Frequency of differing cases, minimum, maximum, means, and standard deviations of absolute differences between trait affinities assigned at family-level and aggregated trait affinities.

Database	Comparison to traits at fam.-lvl.	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

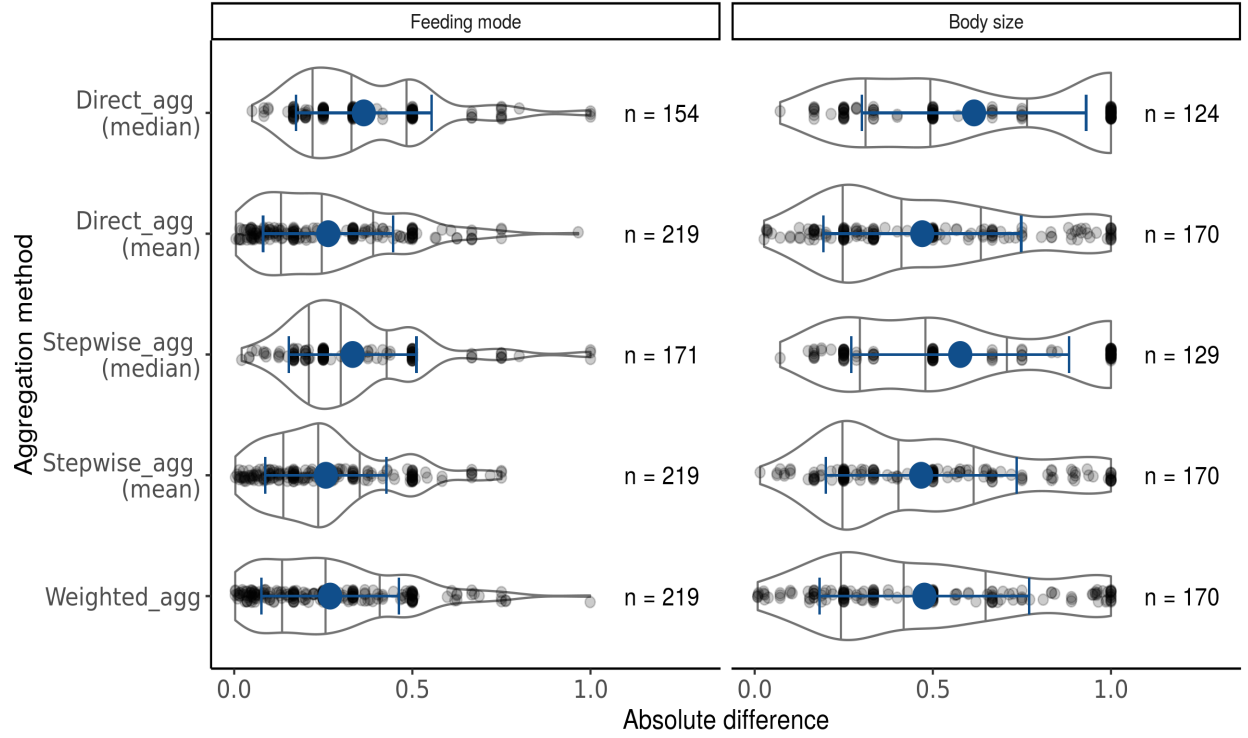


Figure 6: Cases (factor combination investigated families and traits) where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the Australian dataset for the grouping features feeding mode and body size. Absolute differences in trait affinities are depicted as gray 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 the standard deviation. The gray vertical lines show the 25th, 50th and 75th quantile of the density estimate.

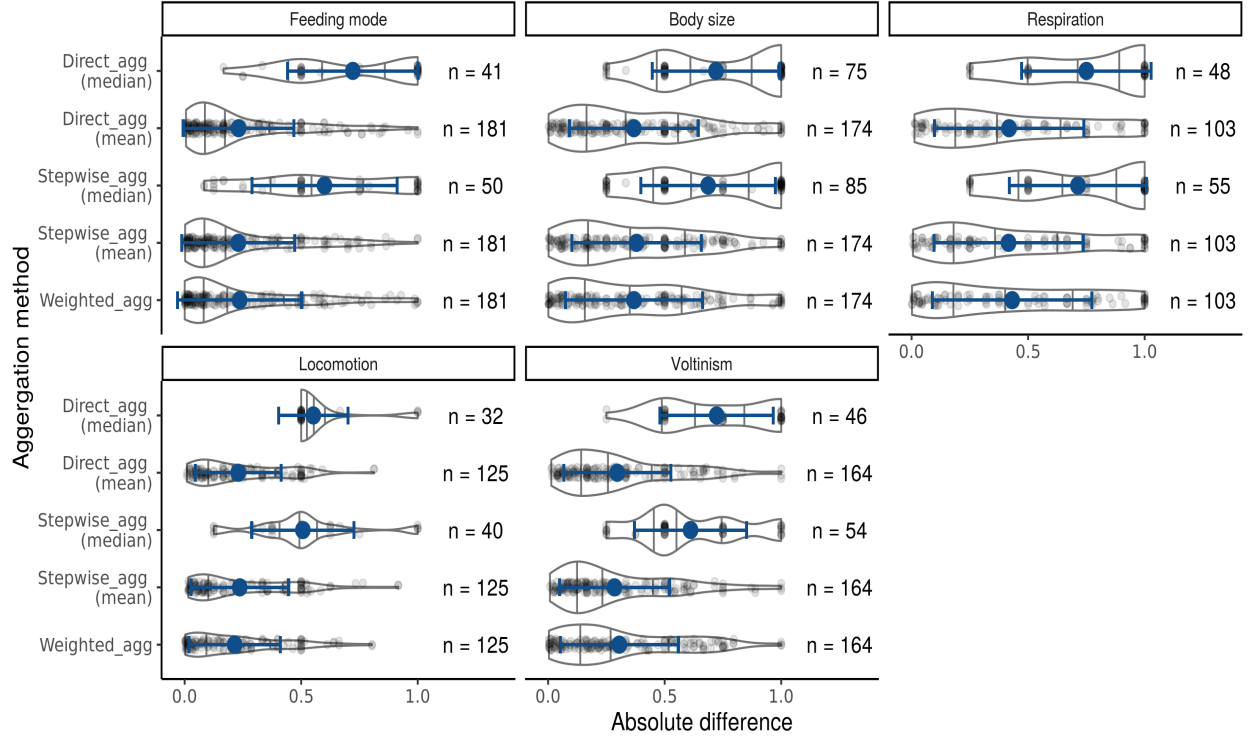


Figure 7: Cases (factor combination investigated families and traits) where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the North America dataset for the grouping features feeding mode, locomotion, respiration, body size and voltinism. For more details see Figure ??.

===== »»»j dcfb038be0251b833001b350d3e336fd4f22a9fc

Effects of harmonisation and trait aggregation on trait-environment relationships

The original RDA in **szocs'effects'2014** indicated that downstream sites (higher salinity) were characterised by the traits shredder, ovoviviparity, multivoltinism, long life cycle duration (> 1 year), and gill respiration, whereas upstream sites (lower salinity) were characterized by univoltinism, oviposition in clutches and short life cycle duration (< 1 year).

Overall, harmonised trait data and all aggregation methods produced similar distributions of RDA species scores (Figures ??) compared to the original analysis in **szocs'effects'2014**. Using harmonised grouping features resulted in less traits that distinguish upstream and downstream sites in comparison to the original analysis (Figures ?? and ??). According to the RDA of the trait composition, downstream sites were characterised by taxa with the traits multivoltinism, ovoviviparity and long life cycle duration. Upstream sites were characterised by univoltine taxa with a short life cycle duration that lay their eggs in an aquatic environment (aquatic eggs). The trait aquatic eggs of the harmonised grouping feature oviposition was derived by amalgamating the trait oviposition in clutches and other related traits (Table ??). The traits shredder and gill respiration did not characterise sites with higher salinisation.

Using traits aggregated at family-level from the harmonised dataset also showed results similar to the original analysis, but with less 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 that downstream sites were not characterised by the trait shredder. Upstream sites were characterised by the traits univoltinism, aquatic eggs and short life cycle duration. The same results were obtained when re-analysing the data with traits aggregated by the *stepwise_agg_mean* and *stepwise_agg_median*, with the exception that none of the life cycle traits characterised upstream or downstream sites. Thus, in our comparison aggregating by *direct_agg_mean*, *direct_agg_median*, and *weighted_agg* yielded the least change in interpretation of the RDA results compared to the original findings.

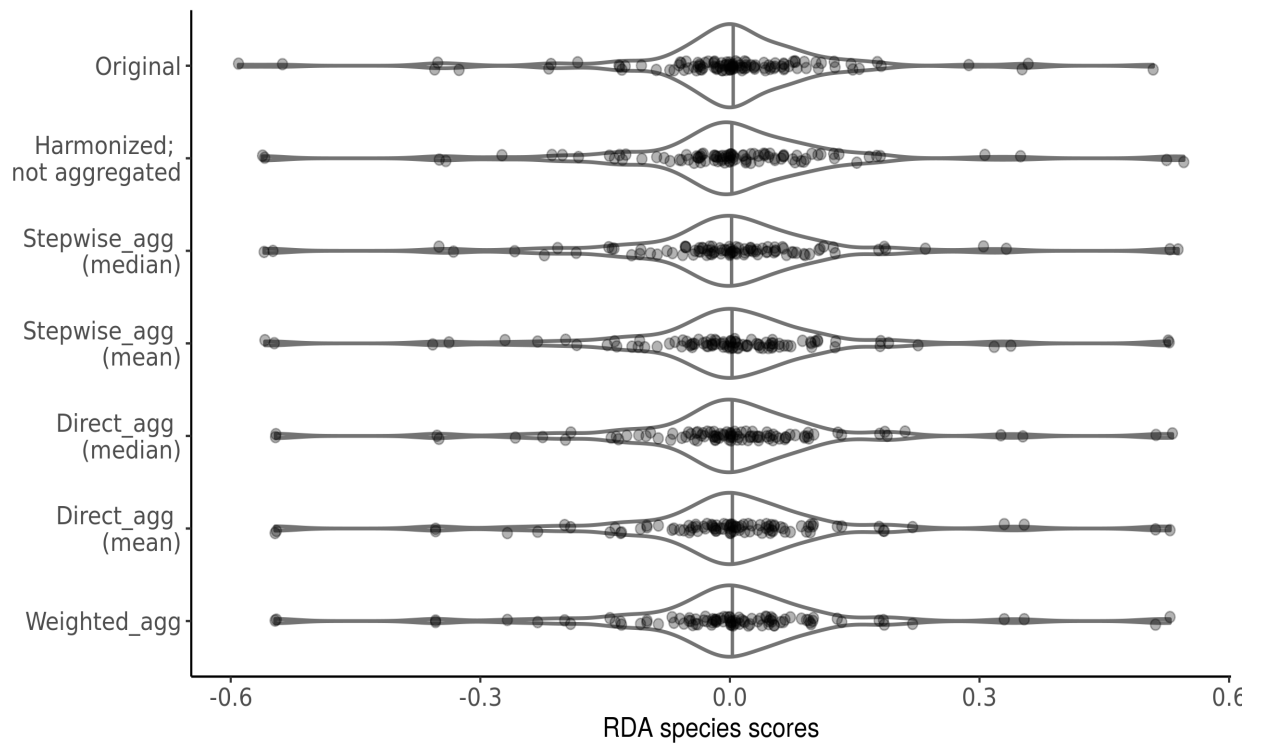


Figure 8: Species scores obtained by RDA from the original analysis (*szocs'effects'2014*), using harmonised grouping features, and using harmonised grouping features with traits aggregated to family-level. The dots represent the individual species scores for each analysed trait along the conductivity axis. The violin plot shows the density estimate of the species scores. Gray vertical lines indicate the median of the obtained species scores.

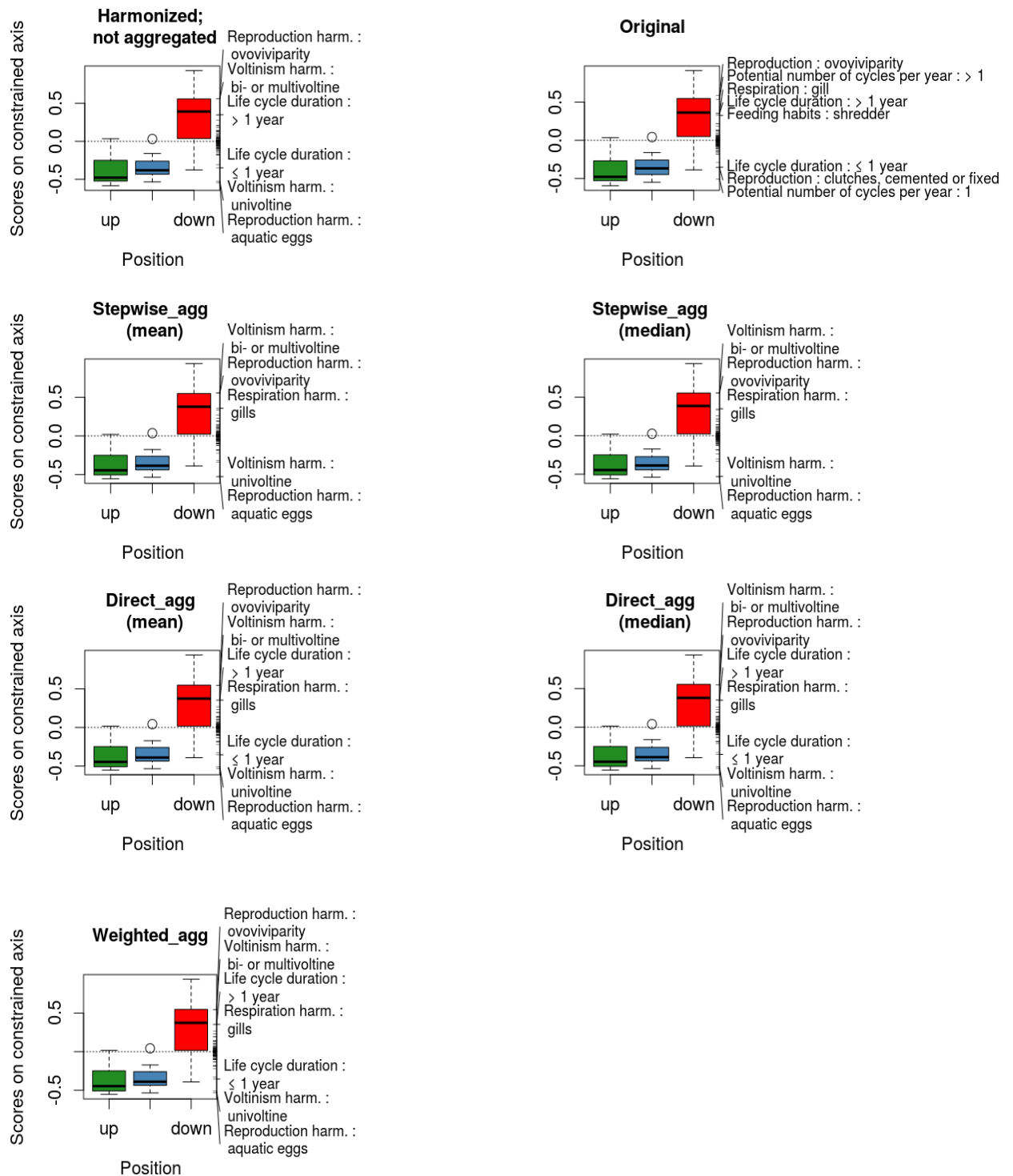


Figure 9: RDA of traits constrained by electric conductivity for the tested methods and the original study. Shown are boxplots of the site scores along the conductivity axis. The rug on the right side of each plot indicates species scores of the traits on the conductivity axis. Only traits with a Mahalanobis 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.

Discussion

Trait definition discrepancies and taxonomic resolution

Synthesizing trait information from multiple trait databases is crucial for developing the full potential of trait-based approaches, e.g. for studying community trait responses to environmental gradients worldwide. Our attempt at harmonising invertebrate trait information from trait databases of different regions showed that harmonisation was labour-intensive because grouping features are resolved differently into traits and various codings are used to describe traits in the various databases. In addition, the same traits are sometimes defined differently, requiring expert knowledge to reclassify these traits (e.g. the trait piercer in the Tachet database). Although, others have noted the lack of standardised trait terminology in freshwater ecology (**baird'toward'2011**; **brink'traits-based'2011**), we are the first to describe invertebrate trait definition discrepancies for some commonly used grouping features. 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, e.g. 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 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'toward'2011** a decade ago. 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, these will improve trait data sharing, trait data processing when working with multiple trait databases, and ultimately interpretation of the derived results.

Our analysis aimed to use most of the available invertebrate trait information for different regions to establish harmonised grouping feature datasets. Although many taxa are covered in the trait databases from which we extracted our datasets, the availability of trait information varies strongly across grouping features and taxonomic resolution differs between databases. Much information is available for grouping features that are often used in trait-environment studies, such as feeding mode and respiration. Surprisingly, there is little information on body form, a trait that is relatively easy to determine.

While trait information from Europe and New Zealand was mainly at species-level, considerable portions of trait data from North America and Australia were at genus or family-level. Assigned traits at family-level may not reflect the real trait diversity within a taxon, for example for the large, ecologically diverse family Chironomidae (**serra'synthesising'2016**). However, species-level identification can be complicated, time-consuming and expensive (**marshall'taxonomic'2006**; **resh'which'2008**). Consequently, trait information at family-level has been widely used by freshwater ecologists, for example in bioassessment (**beketov'spear'2009**).

Trait aggregation

When trait values are only available at species-level and observational data are at less precise taxonomic levels, various trait aggregation methods have been used by researchers. We aggregated traits from an Australian and a North American trait dataset to family-level using 5 different ag-

gregation methods and compared the results to traits originally assigned at family-level for these regions. Evaluation of the differences between aggregated and assigned traits is difficult because it remains unclear what the true value of a particular trait for a particular family is. 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 a high variability or if trait information at species-level is missing. Some traits can vary strongly at more precise taxonomic levels than the family-level. For example, **monaghan'improving'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'functional'2006**; **wilkes'traitbased'2020**). If trait aggregation is necessary, however, our study can give guidance, under which circumstances the choice of aggregation method may be important. The 5 aggregation methods we tested use either the mean or the median and different weightings. Our results indicated that 1) the median aggregation methods compared to mean aggregation methods were often closer to the assigned traits at family-level and 2) the different weightings tested exerted a minor impact on the aggregated trait affinities. We obtained for both datasets more strongly differing cases for mean aggregation methods, irrespective of the weighting. In the Australian dataset, the amount of differing cases was slightly higher for the mean aggregation methods (23 %) than for median aggregation methods (16 to 18 % differing cases). In the North American dataset, mean aggregation methods yielded 47 % cases being different to the assigned traits, a value much higher in comparison to the median aggregation methods (between 15 and 18 %). When differences occurred between median aggregation methods and assigned traits, these differences were greater compared to the mean aggregation methods, particularly for the North American dataset. There, the mean absolute differences for the median aggregations were twice as high (0.63 - 0.7) compared to the mean aggregations (0.3 - 0.31). These differences can be explained by the binary coding used in the North American trait dataset and by the assigned traits. As a result, 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. Using the median to aggregate binary coded taxa often resulted in a value of 0 or 1, while using the mean 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.

The weightings explored in this analysis should lead to different aggregated trait affinities in cases where genera within a family have different numbers of species. In fact, the simulation results indicated that for a particular family where one genus has a much larger number of species compared to the other genera, weighting the number of taxa can lead to different aggregated trait affinities, especially for traits with higher levels of variance. However, in the comparison of the trait aggregation methods to assigned traits the number of differing cases and the mean absolute differences in trait affinities were similar across the mean aggregation and across the median aggregation methods, which suggests a small influence of the weightings on the aggregation methods. Also, the distributions of absolute trait affinity differences to assigned traits were relatively similar for the mean aggregation methods and the median aggregation methods, respectively. The minor impact on trait aggregation of the weightings 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 5

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 5 or fewer genera, and 40 % just one genus (Figure ??). Hence, these results could change when more trait information at species-level becomes available.

Grouping feature harmonisation

We explored how using harmonised grouping features and aggregated traits might change the results of trait-environment relationships. We are not aware of other studies that compared trait-environment relationships using harmonised and non harmonised data. The re-analysis of the data on salinisation effects on biological traits from **szocs'effects'2014** yielded only slightly different results compared to the original study, i.e. some of the traits that had responded in the original analysis did not respond in the re-analysis. These non-responsive traits were those closest to the criterion that defined when a trait was associated with either higher or lower salinity, namely feeding mode shredder, respiration gills, and life cycle duration traits. Feeding mode shredder was the only trait that was non-responsive when using harmonised (but not aggregated) as well as when using aggregated data. This result was likely due to the harmonisation procedure, when this trait was amalgamated 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 aggregated and harmonised data (Table ??), suggesting that the signal in the original data was weakened by the harmonisation. These findings show 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. The fact that the family-level aggregated traits used in the re-analysis showed similar results compared to the original analysis agrees with previous findings that family-level traits can be sensitive enough to detect environmental impacts (**beketov'spear'2009**).

Outlook

Although comprehensive freshwater invertebrate trait databases have been developed for several regions in the past, data synthesis is difficult because of discrepancies in trait definitions. By providing an overview of these definition discrepancies we have set a starting point for the development of standardised trait terminology through which invertebrate trait databases can be linked. A consent 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. We could show that trait affinities resulting from fuzzy coding and binary coding can be used together, but a uniform coding of traits is another problem that should be addressed during trait standardisation. As ecological knowledge increases in the future, and with the aid of new methods such as computer vision for species and trait identification (**hoye'deep'2020**), more observational and trait data at species-level may become available in the future, reducing the need for trait aggregation. However, if traits need to be aggregated we showed that trait aggregation is often in agreement with expert assignments, especially aggregation approaches using the median. Ultimately, the question which aggregation method is the objectively best measure cannot be answered. Nevertheless, we make the following recommendations for situations when trait aggregation is necessary. 1) The aggregation method will tend to have little effect except for highly variable traits and taxa where a few genera have a high number of species. 2) The coding used to describe traits (binary, fuzzy) should be

considered, because median and mean aggregations can yield very different results for binary coded traits. 3) Weighting trait information based on the taxa present in the database used does not seem to have much influence except for families with few genera that have a high number of species. 4) A measure of trait dispersion should be reported to indicate the uncertainty of the aggregated estimate.

Supporting Information

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 herbi-vore includes among others insect that shred <i>living aquatic plants</i> 	Shredder	<ul style="list-style-type: none"> Detrivore ^a Trait herbi-vore includes among others the trait shredder 	Shredders
Feeding predator	"Eating from prey"	<ul style="list-style-type: none"> 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

Feeding filter-feeder	Distinguishes be- tween active and passive	No distinction be- tween active and passive	No distinction be- tween active and passive	No distinction be- tween active and passive	No distinction be- tween active and passive	No distinction be- tween 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"
Multivoltine					<ul style="list-style-type: none"> • 1-2 generations per year • bi/multivoltine • up to 5 generations per year • up to 10 generations per year 	"> 1 reproductive cycles per year"
		"Able to complete <i>at least</i> two successive generations per year"	"> 1 generations per year"	"> 1 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 "fish-like" 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 (in-fauna)
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 traits: sprawler, crawler, climber and clinger.	Crawlers (epibenthic)
Locomotion sessile	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 plastron, atmospheric 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 ^d	< 9 mm	< 9 mm	< 9 mm ^{a;c}	Multiple size classifications ^e
Body size medium	-		9 - 16 mm	9 - 16 mm	9 - 16 mm	
Body size large	-		> 16 mm	> 16 mm	> 16 mm	

^a Traits from Botwe et al.

^b Contains also bivoltine (two generations per year), trivoltine (three generations per year) and flexible.

^c Contains a size trait with numeric size values. Contains also traits classifying size like Tachet and like the North American trait databases.

^d 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.

^e 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.

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

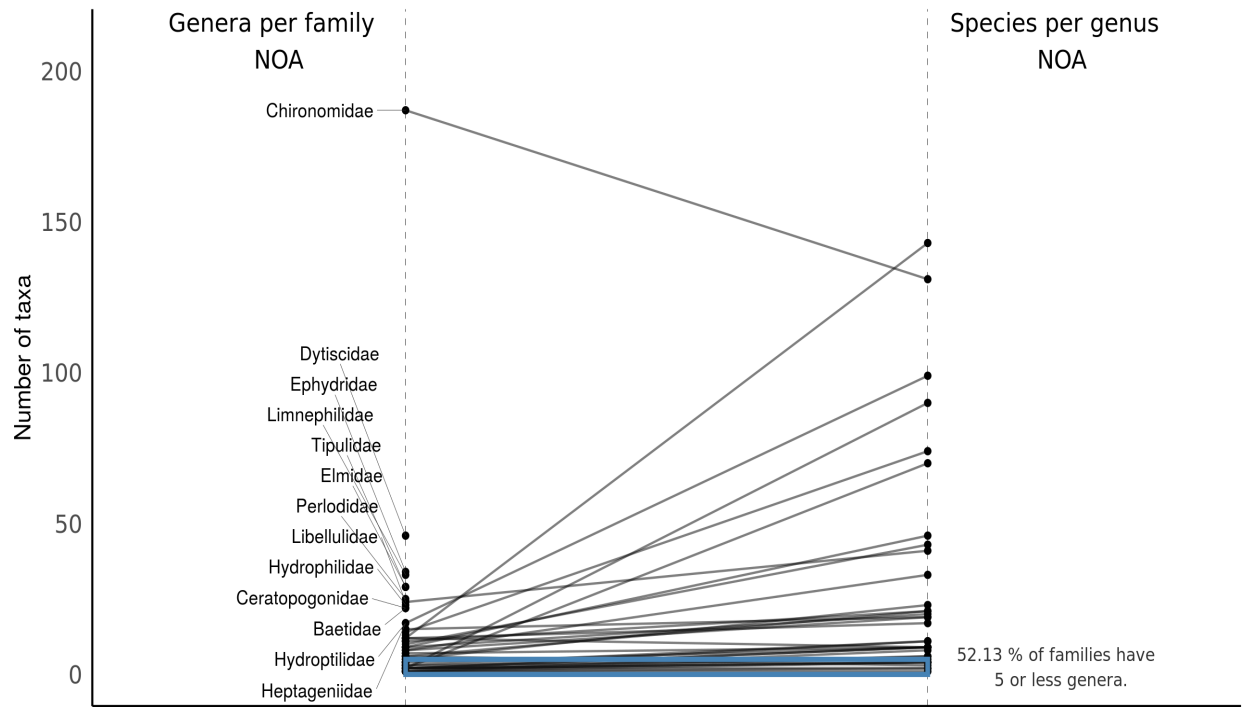


Figure S1: Number of genera per family and species per genus for those families of the North American trait dataset that have been compared to assigned traits at family level. For better visual display only families with more than 15 genera are displayed.

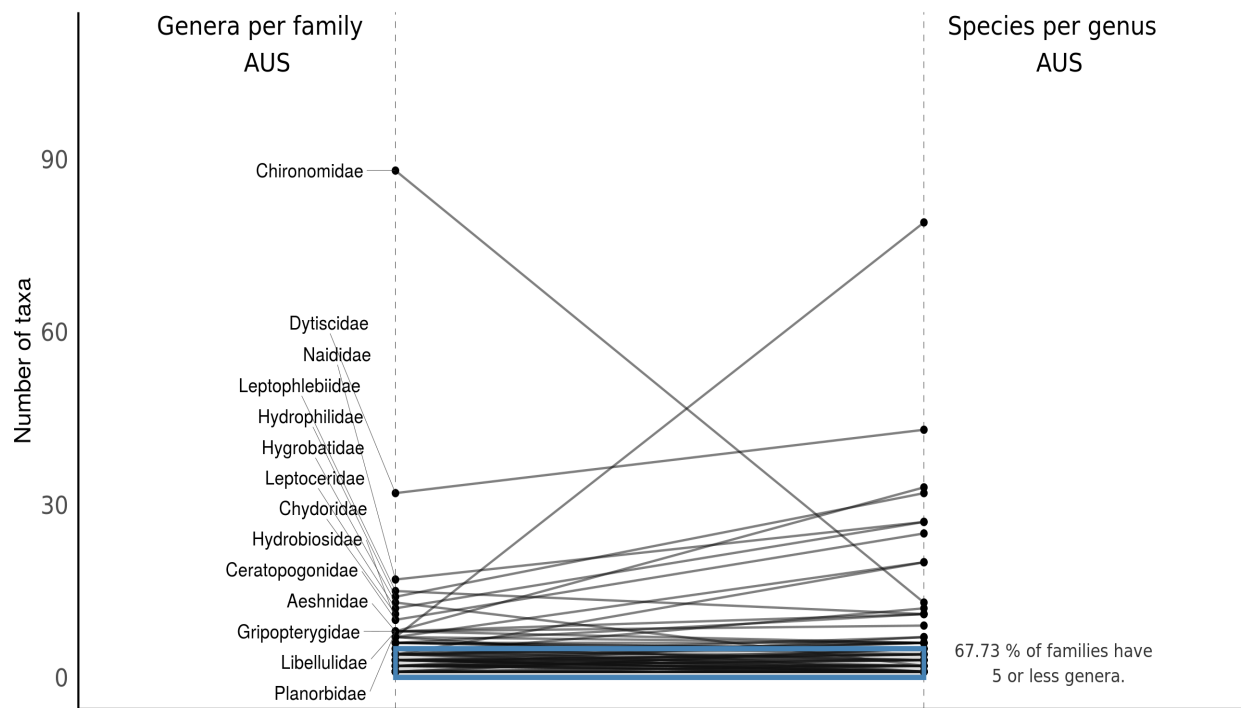


Figure S2: Number of genera per family and species per genus for the Australian trait dataset. For better visual display only families with more than 7 genera are displayed.

Re-analysis of Szöcs et al. using harmonised and aggregated grouping features

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