DRAFT: Harmonisation and trait aggregation paper

Summary

1. Invertebrate traits are increasingly used in ecological studies because they provide a mechanistic understanding of species-environment relationships and are suitable for large scale analysis due to their consistent response to environmental gradients. The development of invertebrate trait databases for different continents in recent years should have created impetus for trait-based research, especially on large scales. However, a number of data-related problems hamper the work with trait-based approaches. For example, standardised trait definitions are lacking and coding of traits varies across trait databases impeding trait data synthesis. Another issue is different taxonomic resolution between observational taxonomic data and trait databases, leading to the need to aggregate traits.

2. In this study, we describe discrepancies in trait definitions of trait databases from Europe, North America, New Zealand and Australia for traits frequently used in ecological studies. Based on these databases we established 4 novel trait datasets by harmonising 7 commonly used grouping features. Two of these datasets were used to evaluate different trait aggregation methods by comparing aggregated traits with assigned traits by experts at family-level. Furthermore, we explored the effect of grouping feature harmonisation and trait aggregation by re-analysing a study on the effect of anthropogenic salinisation on biological traits using harmonised grouping features and aggregated traits.

3. We found that investigated trait databases are distinct with regard to how grouping features are differentiated into traits, the scope of the grouping feature definitions, how traits are defined, and the coding to describe traits. Comparison of aggregated and assigned traits showed that aggregated traits are generally in conclusion with expert assignments, especially approaches using the median. Grouping feature harmonisation and trait aggregation had a minor impact in the re-analysis, although averaging effects can occur due to harmonisation diminishing the signal of a trait.

4. By outlining discrepancies in trait definitions we hope to motivate the development of a standardized terminology for invertebrate traits. When trait aggregation is needed, our results provide guidance under which circumstances the choice of a trait aggregation method becomes important.

Introduction

Explaining and predicting how communities are shaped by environmental factors is one of the main goals of ecology. Species traits, defined as measurable properties of an organism (McGill et al. 2006), might be beneficial in achieving this goal (Heino et al. 2013). Traits evolve through adaptations (e.g., physiological, behavioural, etc.) of organisms to their environment and indicate direct or indirect linkages between the biological response of an organism or a population and its environment (Southwood 1977; Verberk et al. 2013). Besides providing a mechanistic explanation of species-environment relationships, trait-based approaches might be suitable for large scale analysis because the variability in trait responses is lower than in taxonomic responses (Bonada et al. 2007; Baird et al. 2011). Traits of freshwater invertebrate individuals are difficult to determine because - unlike plant traits - they often can not be measured directly. For example, to gain knowledge on feeding habits requires evaluating mouthpart morphology, consumed food, and the organisms function within its community (Moog and Hartmann 1995). Nevertheless, invertebrate traits have been increasingly used in freshwater ecology, e.g. by relating macroinvertebrate trait composition to environmental factors and as trait metrics in biomonitoring (N. Poff et al. 2010; Szöcs et al. 2014; Bhowmik and R. B. Schäfer 2015; Menezes et al. 2010).

In the last decades, freshwater ecologists have compiled comprehensive invertebrate trait databases for various continents (Usseglio-Polatera et al. 2000; Schmidt-Kloiber and Hering 2015; Vieira et al. 2006; Philips and Smith 2018; Kefford et al. 2020; Tomanova et al. 2006). The availability of invertebrate trait data from different continents enables comparisons of trait variation and analyses of their relation to environmental factors across large scales. However, such analyses have been carried out mostly within continents, using information from one or two trait databases. For example, Bonada et al. (Bonada et al. 2007) compared trait composition for Mediterranean and temperate regions in Europe using traits from Usseglio-Polatera et al. (Usseglio-Polatera et al. 2000, typically referred to as Tachet database); Poff et al. N. Poff et al. 2010 characterized trait composition across sites in the Western US using traits from Poff et al. (N. L. Poff et al. 2006), and Botwe et al. Botwe et al. 2018 investigated the effect of salinity on invertebrate traits in different sites in South Australia using trait data from Poff et al. (N. L. Poff et al. 2006) and Schäfer et al. (R. B. Schäfer et al. 2011). Analyses of invertebrate traits that synthesize information on invertebrate grouping features from more than two different continents are rare.

In this study, we follow the terminology proposed by Schmera et al. (Schmera et al. 2015), where a grouping feature is defined as a general property (e.g. feeding mode) that comprises a "group of related traits (e.g., predator, shredder, etc.) that vary among species or among individuals within a species". Thus, we use the term grouping feature in place of which many studies use the term "trait", and the term trait instead of "trait state", "modality" or "trait category". Traits can be described using different codings (binary, fuzzy) that represent the prevalence of a characteristic in an organism.

To our knowledge, only Brown et al. (Brown et al. 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 community assembly of invertebrates. We suspect that 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. To harmonize grouping features from different regions, first commonly accepted and unambiguous trait definitions are required (Schneider et al. 2019). In the best case, grouping features would be classified into the same traits across

databases or they could easily be harmonised using standardised terminology. However, a lack of standardised terminology of trait definitions and poor metadata quality in many trait databases are common issues throughout the field of trait-based ecology (Baird et al. 2011; Kissling et al. 2018, October). Secondly, consistent coding of traits facilitates the compatibility of trait data from different databases.

Traits can be described in a binary fashion, or with multiple categories, ignoring uncertainty how the trait is expressed in any particular organism (e.g. adult terrestrial stage, presence of gills) or continuous (e.g. tolerance of pollution, body size). One approach for dealing with uncertainty is the use of fuzzy coding, where the traits are assigned probabilistic values. Fuzzy codes are used to account for plasticity in traits, variability in traits within taxonomic groups above species, and incomplete knowledge and are usually converted to proportions. In this study we denote the probabilistic values (0-1) of traits as affinities or affinity scores.

However, invertebrate trait databases are heterogeneous regarding the coding they use for their traits (Culp et al. 2011). For example, in the study from Brown et al. (Brown et al. 2018) the authors needed to reclassify traits using expert knowledge given that the individual databases employed different coding approaches (i.e. European and New Zealand databases fuzzy coding, North American database binary coding). Databases with standardised and unambiguous traits and a consistent coding of traits would minimize the data processing effort. Discrepancies in the taxonomic resolutions (e.g. species, genus or family) when linking observational taxonomic data to trait database represents another challenge. When observations are on a more precise taxonomic level than data available in the trait databases (e.g. observations on species-level, trait data on genus-level) trait information of the less precise taxonomic level is often assigned (e.g. Szöcs et al. 2014; Voß and R. Schäfer 2017). Conversely, if trait information is only available on more precise taxonomic levels than the observed taxa, traits are aggregated to a less precise taxonomic level (e.g. N. L. Poff et al. 2006; Szöcs et al. 2014; Pilière A. F. H. et al. 2016; Aspin et al. 2019). Studies have used different methods of trait aggregation, e.g. the mean (Magliozzi et al. 2019), median (Szöcs et al. 2014) or the mode (Pilière A. F. H. et al. 2016). Up to now, studies on how and to which extent different trait aggregation methods influence trait-based analysis are missing. However, related knowledge would inform future studies regarding the choice of the aggregation method.

The outlined difficulties when working with or synthesizing trait data motivated us to (1) investigate discrepancies in trait definition between trait databases from Europe, North America, Australia, and New Zealand. As a basis for further evaluating different trait aggregation methods and how harmonising definition discrepancies between grouping features can influence trait-environment relationships, we established 4 novel invertebrate trait datasets from these trait databases by harmonising 7 grouping features. We evaluate different trait aggregation methods by (2) comparing aggregated traits with assigned traits by experts at family-level for two of the established trait datasets. (3) we re-analysed data on the effect of anthropogenic salinisation on biological traits by Szöcs et al. (Szöcs et al. 2014) using trait data from the established harmonised European trait datasets and aggregated traits. By comparison with the original analysis, we investigated how harmonising and aggregating trait data can modify the outcome of trait-environment relationships.

Methods

Selection of traits and harmonisation of trait databases

We extracted information from 6 trait databases from Europe, North America, Australia, and New Zealand and harmonised 7 grouping features. Trait information for Europe was obtained from the freshwaterecology.info database Schmidt-Kloiber and Hering 2015 and the Tachet database Usseglio-Polatera et al. 2000. The freshwaterecology.info contains taxa on species-level, while taxa recorded in the Tachet database are on species, genus and family-level. Trait information for North America was obtained from Twardochleb et al. Twardochleb et al. 2020 and complemented by Vieira et al. Vieira et al. 2006. Data on body form for European and North American taxa was based on expert knowledge Usseglio-Polatera 2020b. For Australia and New Zealand, we used trait databases from Kefford et al. Kefford et al. 2020 and Philips and Smith respectively Philips and Smith 2018. To increase readability we refer to the databases as well as the datasets we extracted from them by the name of the continent they originate from, except for European databases, which we refer to by their commonly used names (freshwaterecology.info and Tachet database).

We selected traits of seven grouping features 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. The grouping features were differently classified across the databases, we therefore harmonised them into 26 traits (Table 1). Harmonisation was undertaken by amalgamating similar traits into one trait (e.g. crawlers and sprawlers into crawlers). Thereby, for a particular taxa the highest trait affinity score among the amalgamated traits was taken.

We used fuzzy coding for establishing our harmonised datasets unless data quality prohibited. In the latter case we used binary traits, i.e. categorical and continuous traits were converted into binary traits. For example, in the freshwaterecology.info database, the classification of the trait voltinism accounts for different faunistic regions. Hence, entries such as "arctic" or "boreal" in the voltinism traits were substituted with a value of 1. Implicitly, we assumed for binary traits that a value of 1 and 0 corresponds to the highest and no affinity for a particular trait. Fuzzy codes are reported with different ranges in the trait databases (e.g. freshwaterecology.info 0 to 10, Tachet database 0 to 3 or 0 to 5). We standardised these to a range between 0 and 1 and converted trait affinities to percentages. Thus, fuzzy coded and binary traits were in the same range.

Prior to harmonisation we consolidated duplicate taxa on species, genus or family-level if present within a dataset, by either applying the median for fuzzy coded traits, or the maximum for binary traits. We omitted taxa with a lower taxonomic precision than family-level.

Table 1: Traits of the harmonised grouping features. The last column indicates traits that were amalgamated for harmonisation (no amalgamation needed if empty).

Grouping feature	Trait	Amalgamated 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 [a]
	Spherical	Spherical, round (humped)
	Streamlined	Streamlined, fusiform
Size	Small	<9 mm, <10 mm [b]
	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)[c]
	Parasite	
	Predator	Piercer (animals)[c], 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

a The trait "bluff (blocky)" occurred in the Vieira et al. Vieira et al. 2006 database and was newly classified by expert knowledge into cylindrical and flattened Usseglio-Polatera 2020b.

b Reflects the different size classifications by the North American trait databases and the other trait databases.

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

Trait aggregation

Traits of the harmonised grouping feature datasets were aggregated to family-level using three approaches. I) Direct aggregation of taxa to family-level giving equal weight to every taxon using the mean or median, denoted $direct\_agg_{mean}$ and $direct\_agg_{median}$, respectively. II) Stepwise aggregation, i.e. first to the genus-level and subsequently to the family-level using the mean or median. This approach gives equal weights to each genus. Hereafter, we denote this aggregation type as $stepwise\_agg_{mean}$ or $stepwise\_agg_{median}$, respectively. III) Aggregation using a weighted mean approach, denoted as $weighted\_agg$. This method weights the genera according to the number of their species in the trait datasets regardless if for every used grouping feature information was available (Figure 1). When we refer to $direct\_agg_{mean}$, $stepwise\_agg_{mean}$, and $weighted\_agg$ together we denote these methods as mean aggregation methods and when we refer to $direct\_agg_{median}$ and $stepwise\_agg_{median}$ together, we denote these methods as median aggregation methods.

To examine the influence of the taxonomic hierarchy and the trait variability on the outcome of the different trait aggregation methods we created three hypothetical examples of different taxonomic hierarchies (scenarios). 1) A family with an equal number of genera and species (5 genera each with 5 species respectively), denoted as *sim_base*. 2) A family where one genus has a much larger number of species than the other 4 genera (1 genus with 13 species, 4 genera with 3 species respectively), denoted as *sim_extreme*. 3) A family where all genera have a different number of species (8, 2, 7, 3, 5), denoted as *sim_variation*. In total every family consisted of 25 species. To each hypothetical taxonomic hierarchy a hypothetical grouping feature with 3 traits was assigned. The 25 affinities for each trait were simulated by sampling from a truncated normal distribution with a mean value of 0.5 and 5 levels of standard deviations (0.2, 0.4, 0.6, 0.8, and 1) respectively to simulate different levels of trait variability. The truncated normal distribution was bound to 0 and 1. Simulated trait affinities were converted to percentages, similar to the data processing of the trait databases. The sampling was repeated 100 times for each standard deviation. Hence, in total 100 datasets for each level of trait variability were simulated, resulting in 12500 simulated trait affinities for each simulated trait (25 species per family $*5$ levels of trait variability $*100$ replicates). These 12500 simulated affinities per trait were assigned to each scenario. The above mentioned trait aggregation methods were applied to each simulated dataset. The results were compared based on the produced range of aggregated trait affinities between levels of trait variability and scenario. We also compared for each simulated dataset the differences in trait affinities obtained by each aggregation method.

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

Data preprocessing

Consolidating duplicate taxa in EU, NOA, AUS & NZ DBs → AUS traits standardized to [0-1] because multiple codings used throughout this database → Taxa more precise than family-level

Harmonisation → Traits converted to percentages and normalized to [0-1] → Trait aggregation

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

Direct aggregation

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

Stepwise aggregation

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

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

Weighted aggregation

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

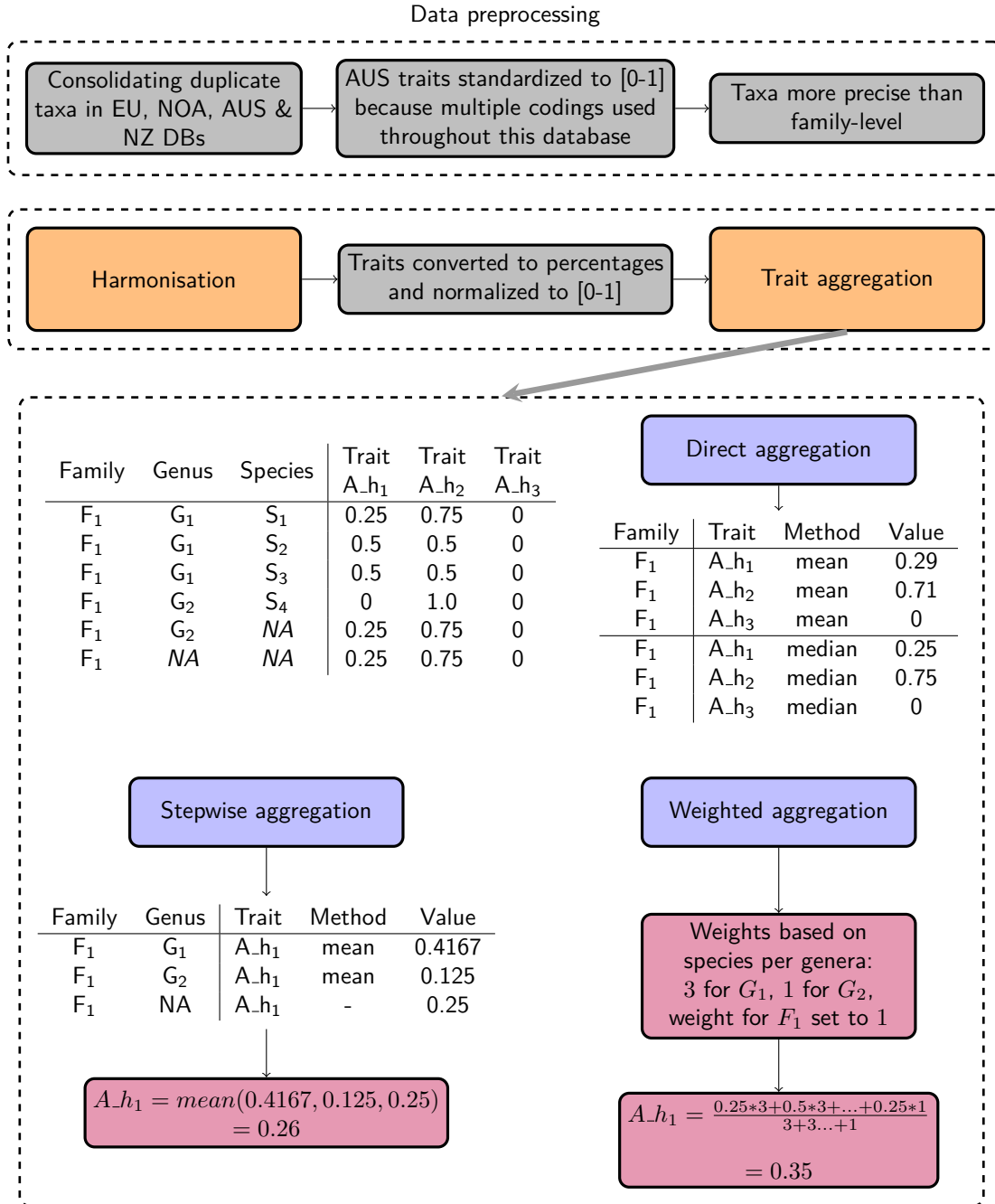$$A\_h_1 = \frac{0.25*3 + 0.5*3 + ... + 0.25*1}{3 + 3 ... + 1} = 0.35$$

Figure 1: Data processing steps of the selected traits. Intermediate (gray) and main (orange) steps of data preparation are depicted. The dashed bottom box illustrates the different trait aggregation methods using a small made-up example (data in the upper left corner). The aggregation methods (blue) and intermediate steps of the aggregation methods (purple) are displayed. Abbreviations: EU: Europe, NOA: North America, AUS: Australia, NZ: New Zealand, DB: Database.

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 et al. Chessman 2018. In Chessman et al. Chessman 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 et al. Chessman 2018. Considering each factor combination of family and trait (in total 8) this amounts 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 North American database (Twardochleb et al.) Twardochleb et al. 2020 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.

Analysis of the effect of harmonisation and trait aggregation on trait-environment relationships

We repeated the analysis of Szöcs et al. Szöcs et al. 2014 who studied the effect of anthropogenic salinization on invertebrates in the River Werra in Germany. For the re-analysis we used the established harmonised grouping feature dataset for Europe and additionally aggregated traits using the aforementioned aggregation methods.

The river Werra has been subject to effluents from the potash industry since the mid of the 20th century and allows to study responses of invertebrates and their trait compositions to salinization Bäthe and Coring 2011, May 1. Sites downstream, upstream, and close to the salt discharge (transition) were compared regarding their trait composition. Further details can be found in Szöcs et al. Szöcs et al. 2014.

We substituted 6 of the grouping features from the original data with harmonised grouping features from the European trait dataset. We compared the results of the redundancy analysis (RDA) from the original study to the case when using 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 identified traits associated with high or low salinity based on their distance to the ordination axis median using the mahalanobis distance. CWM traits with a mahalanobis distance greater than the 97.5% -quantile of the Chi-square distribution (5.02) were regarded as traits responding to either low or high salinity. For our analysis, we used the same 21 grouping features that Szöcs et al. Szöcs et al. 2014 used. The 6 harmonised grouping features used were *Size, Feeding mode, Locomotion, Oviposition, Respiration*, and *Voltinism*. Additionally, for testing the effect of aggregated traits we assigned to each taxon in Szöcs et al. Szöcs et al. 2014 the aggregated trait value for its corresponding family and repeated the RDA.

Data analysis

The data processing and aforementioned analyses were carried out using R (Version 3.6.1). Raw data and the R code for data processing and grouping feature harmonisation is located in the Github repository: https://github.com/KunzstLD/Invertebrate_traits. Scripts and data to reproduce the trait aggregation and analysis with aggregated traits are located in the Github repository https://github.com/KunzstLD/Trait-aggregation.

Results

Taxonomic coverage of the harmonised trait datasets

Regarding the taxonomic coverage, the New Zealand dataset has the smallest taxon pool (478 taxa, Table 2). By contrast, the European trait dataset has the largest taxon pool with 4601 taxa followed by the North American trait dataset that contained trait information on 3753 taxa. The Australian dataset contains 1402 taxa. The European, New Zealand, and North American datasets have most taxa on the species-level whereas the Australian dataset has a similar number of taxa on species and genus-level.

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

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

Completeness of trait information

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

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

Database	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

Discrepancies of invertebrate trait definitions

Definitions of grouping features and traits varied in their level of detail in the investigated trait databases. The freshwaterecology.info database, the Tachet database and the North American database (Twardochleb et al.) provided more detailed descriptions of their trait information compared to the North American (Vieira et al.) and New Zealand database. An exception is the Australian trait database which is a collection of seven specific trait datasets Kefford et al. 2020. Thus, grouping features occur multiple times with varying differentiation into traits. Depending on the dataset trait information is described with more or less detail.

The definition of grouping features varied across databases mainly concerning their differentiation into traits but also in their scope. We provide a summary of discrepancies in trait definitions in the appendix (Table S1). Both, differences in differentiation and scope can lead to discrepancies in trait definitions. For example, for the grouping feature feeding mode discrepancies arise because traits are assigned in different ways. The Tachet database defines predators as carvers, engulfers and swallowers. By contrast, in the North American (Twardochleb et al.) database predators are defined as engulfers and carnivorous piercers. In turn, in the Tachet database, piercers are defined as a separate trait encompassing herbivorous and carnivorous piercers. Furthermore, the scope in the freshwaterecology.info database for feeding mode is primarily on the food source of a species (except for filterers), while the other databases focus on the strategies of food acquisition. Therefore, the freshwaterecology.info database defines e.g. predator as "eating from prey", while the other databases use the mouthpart morphology as basis of their definition. The Tachet database captures the food source in an additional grouping feature. Varying levels of differentiation are also present in all other investigated grouping features between the trait databases (Table 4 and Table S1). Locomotion definitions differ also in scope between databases. Freshwaterecology.info and New Zealand databases describe locomotion as the way of movement of an organism, Tachet as substrate relation and locomotion, the North American (Vieira et al.) as how organisms deal with flow, Australia as attachment, and the North American (Twardochleb et al.) database includes not only the way of movement, but also the location of movement. Similarly, regarding the reproduction traits, databases differ in their scope. Reproduction is captured in one grouping feature and defined as location of oviposit clutches and mode of reproduction in the freshwaterecology.info and Tachet databases. North America (Vieira et al.) provides information on the oviposition location but not on reproductive behavior. The Australian database report traits for reproductive behavior but also on oviposition site. The New Zealand database distinguishes three grouping features related to reproduction: reproductive technique, oviposition site (e.g. water surface, terrestrial), and egg/egg mass (e.g. free, cemented).

Various codings to describe traits are used throughout the databases (e.g. binary, fuzzy, continuous). The freshwaterecology.info and Australian use different codings in their databases. Tachet and the New Zealand database exclusively use fuzzy coding. Both North American trait databases contain categorical grouping features that can be converted into traits using a binary coding (Table 4).

Table 4: Number of traits per grouping feature and type of coding of the traits for the respective grouping feature per database.

Database	Feeding Mode	Voltinism	Locomotion	Respiration	Reproduction	Size	Body Form
Freshwater-ecology.info	10; 10 point assignment system	6; single category assignment system	6; 10 point assignment system	7; binary	9; binary	-	-
Tachet	7; fuzzy [0 − 3]	3; fuzzy [0 − 3]	8; fuzzy [0 − 5]	5	8; fuzzy [0 − 3]	7	-
North America (Twardochleb)	6	3	10	3; binary	10	3	-
North America (Vieira)	8	3	9	8; binary	10	3	5
Australia	16 [a]; binary; proportional [0 − 1]; fuzzy [0 − 3]	7; binary; proportional [0 − 1]; fuzzy [0 − 3]	9; binary; fuzzy [0 − 3]	10; binary; proportional scale [0 − 1]; fuzzy [0 − 3]	13 [b]; categorical	9; binary; continuous; fuzzy [0 − 3]	4; fuzzy codes [0 − 3]
New Zealand	6; fuzzy [0 − 3]	3	4	4	4	5	4

a Some of the traits were similar (e.g. trait *Shredder*, *Shredder*, *Detrivore*, and *Collector*, *Shredder*).
b Many traits were rather comments than traits in the original database and were not considered.

Simulation of varying taxonomic hierarchies

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

As expected with increasing trait variability the produced range of trait affinities increased for all aggregation methods (Figure 2). The mean aggregation methods yielded similar ranges of aggregated trait affinities within each level of trait variability for the *sim_base* scenario. The median aggregation methods yielded mostly greater ranges of aggregated trait affinities for the *sim_base* scenario. 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 $*$3 scenarios). Overall, differences between the results of the aggregation methods for each simulated dataset increased with increasing trait variability. However, in most simulated datasets the different aggregation methods often resulted in similar trait affinities. Only 1.42 % (213 out of 15.000) of all comparisons showed a difference equal or greater than an absolute trait affinity of 0.1. Most of these differences occurred for the *sim_extreme* scenario (83.5 %). The majority of the differences equal or above 0.1 were found between the aggregation methods *direct_agg$_{mean}$* and *stepwise_agg$_{median}$*, *direct_agg$_{median}$* and *stepwise_agg$_{median}$*, and *stepwise_agg$_{median}$* and *weighted_agg* (Figure 3).
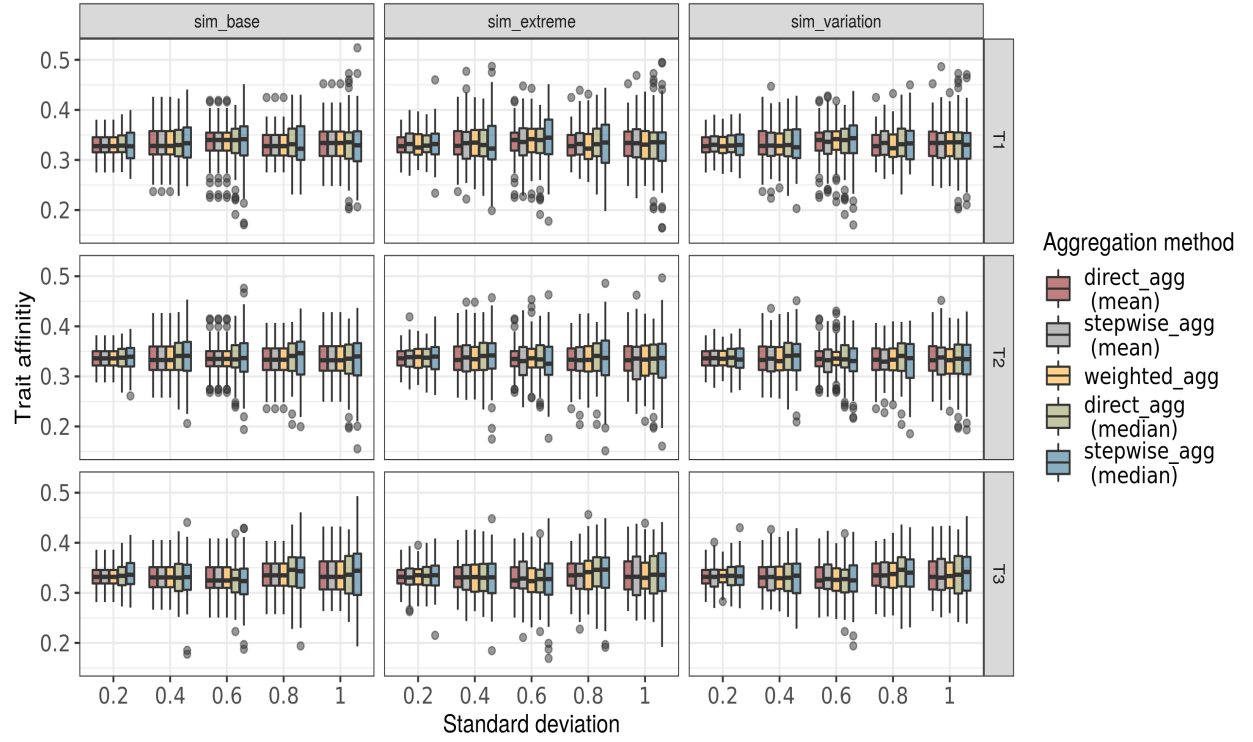
Figure 2: Ranges of aggregated trait affinities for the three examples of taxonomic hierarchies and simulated levels of trait variability. Boxplots depict results for each trait aggregation method of 100 simulations. T1, T2, and T3 are the simulated traits.
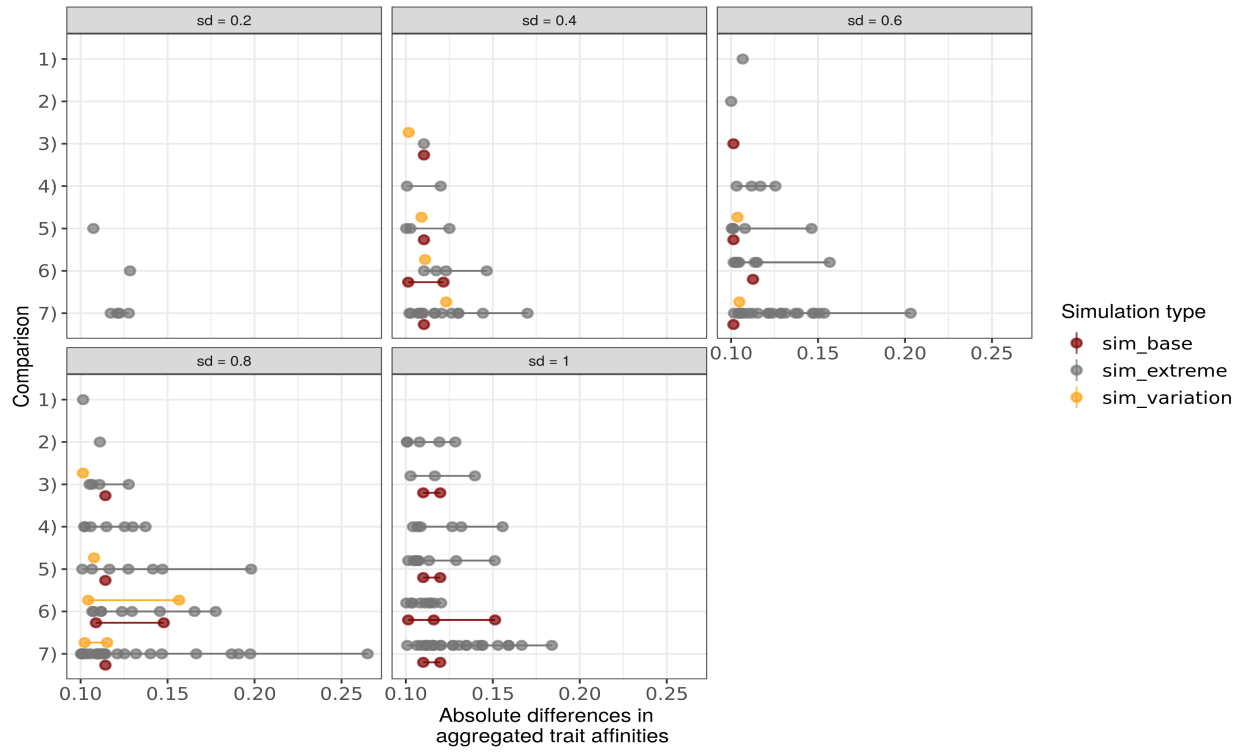
Figure 3: Comparison of the aggregated trait affinities produced by the different trait aggregation methods for every simulated dataset across all 3 simulated traits. Dots depict comparisons where absolute differences between aggregated trait affinities were greater than 0.1.

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

14

Differences in trait affinities obtained by trait aggregation methods compared to traits assigned at family-level

The percentage of differing cases of trait affinities obtained by the trait aggregation methods compared to trait affinities originally assigned at family-level varied between 16.2 % and 22.9 % for the Australian dataset. For the North American dataset, comparison of the trait aggregation methods to trait affinities assigned at family-level yielded between 15.3 % and 47 % differing cases (Table 5). In general, median aggregation methods yielded fewer cases with differences compared to mean aggregation methods. However, median aggregation methods produced greater differences for both datasets. Standard deviations of absolute differences were similar for all tested aggregation methods. For both datasets maximum differences of 1 occurred for all investigated grouping features (Figure 4 and Figure 5).

Table 5: Amount 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 (Chessman)	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
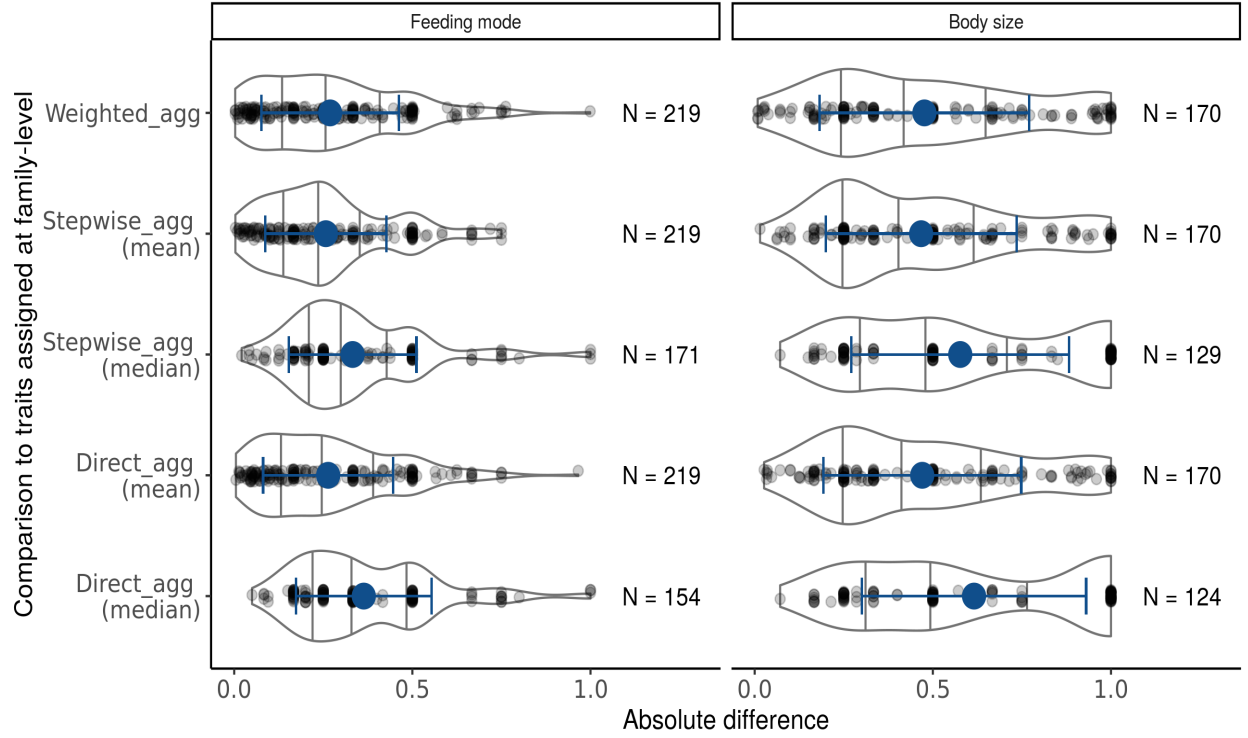
Figure 4: Display of the cases (factor combination investigated families and traits) where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the Australian dataset 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.
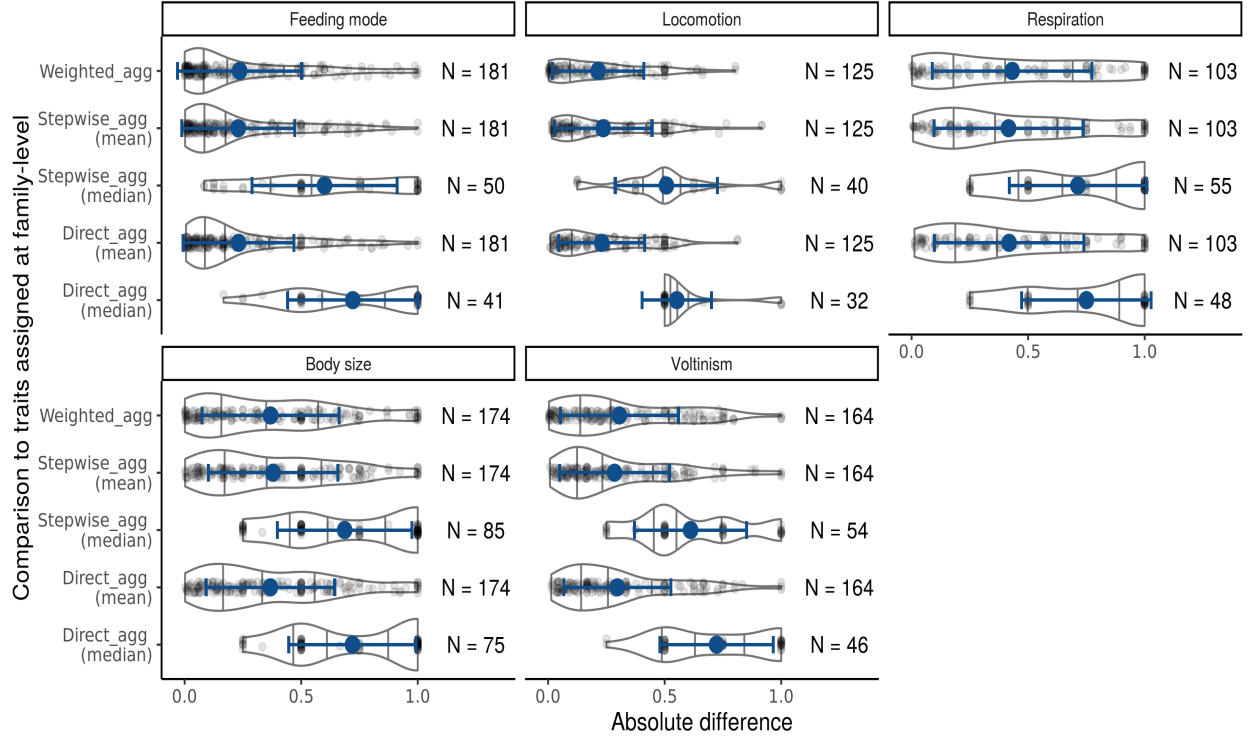
Figure 5: Display of the cases (factor combination investigated families and traits) where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the North America dataset for the grouping features feeding mode, locomotion, respiration, body size and voltinism. 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.

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

The original RDA in Szöcs et al. Szöcs et al. 2014 indicated that downstream sites (high salinity) were characterised by the traits shredder, ovoviviparity, multivoltinism, long life cycle duration (> 1 year), and gill respiration and upstream sites (low salinity) were characterized by univoltinism, oviposition in clutches and short life cycle duration (< 1 year).

Using harmonised grouping features resulted in less traits that distinguish upstream and downstream sites in comparison to the original analysis (Figure 6 and Figure 7). 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 has been derived by amalgamating the trait oviposition in clutches and other related traits (Table 1). The traits shredder and gill respiration did not characterise sites with high salinisation.

Using at family-level aggregated traits from the harmonised dataset showed also results similar to the original analysis but with fewer traits distinguishing upstream and downstream sites (Figure 6 and Figure 7). 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 to the least change in interpretation of the RDA results compared to the original results.
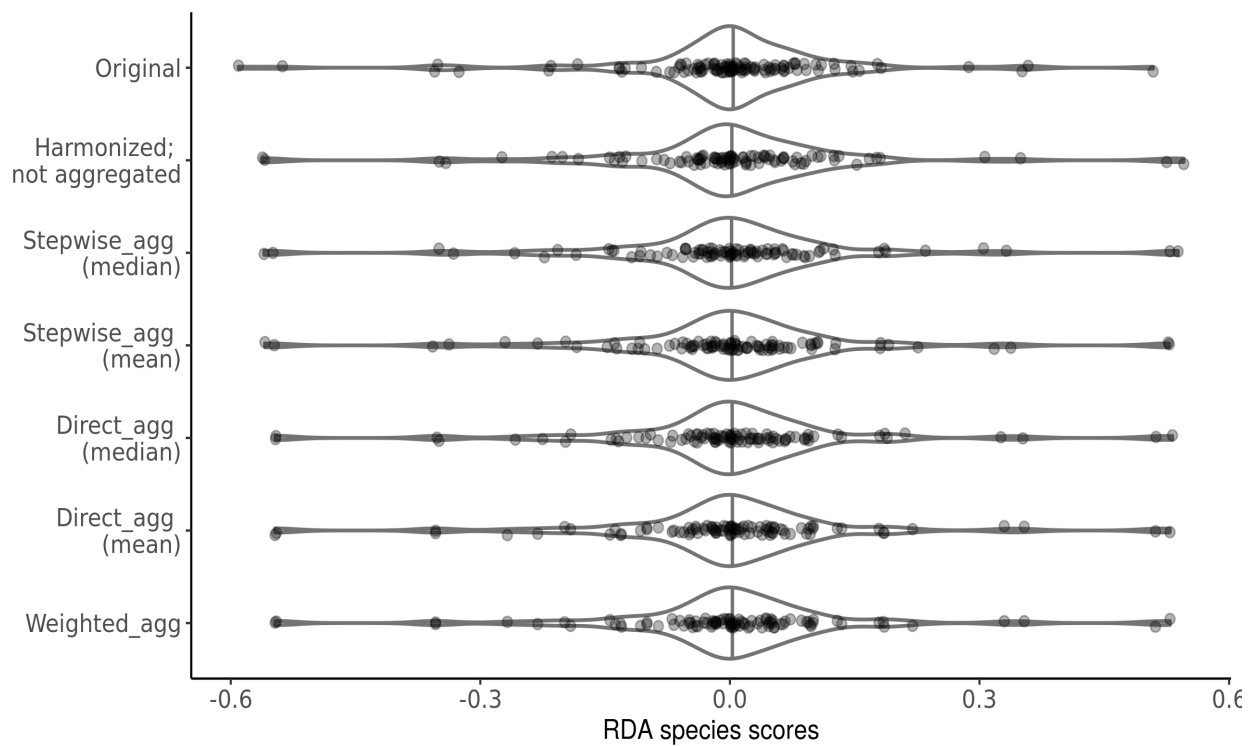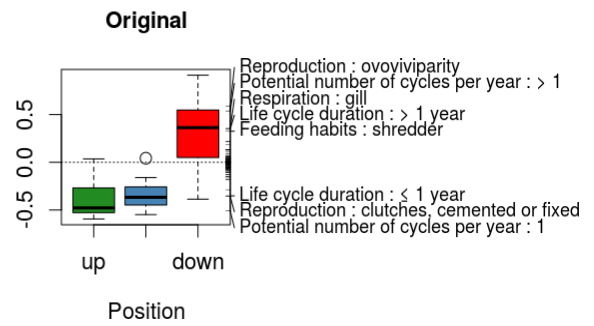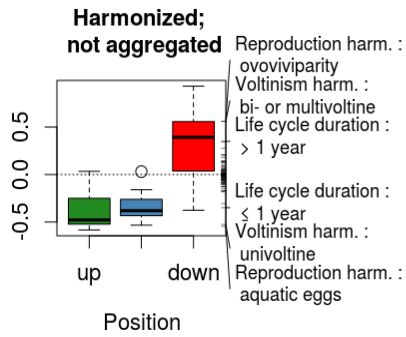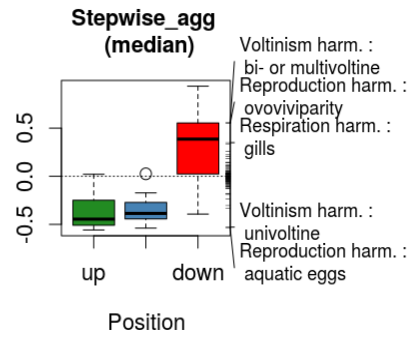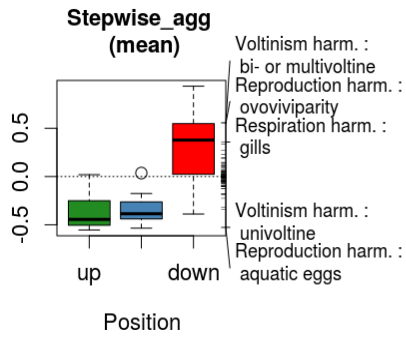
Figure 6: Species scores obtained by RDA from the original analysis Szöcs et al. 2014, using harmonised grouping features, and using harmonised grouping features with traits aggregated to family-level. The dots represent the individual species scores for each analysed trait along the conductivity axis. The violin plot shows the density estimate of the species scores. Gray vertical lines indicate the median of the obtained species scores.
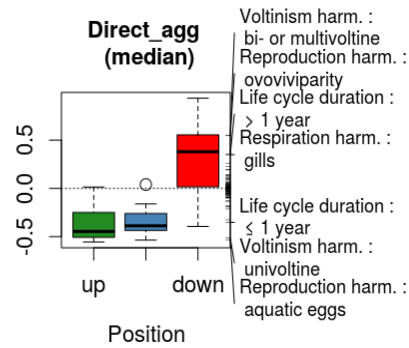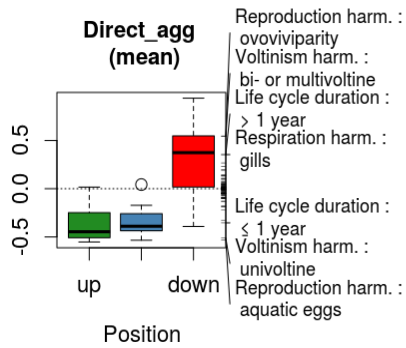
**Harmonized;
not aggregated**

Scores on constrained axis

0.5
0.0
-0.5

Reproduction harm. :
 ovoviviparity
Voltinism harm. :
 bi- or multivoltine
Life cycle duration :
 > 1 year

Life cycle duration :
 ≤ 1 year
Voltinism harm. :
 univoltine
Reproduction harm. :
 aquatic eggs

up   down

Position

Original

0.5
0.0
-0.5

Reproduction : ovoviviparity
Potential number of cycles per year : > 1
Respiration : gill
Life cycle duration : > 1 year
Feeding habits : shredder

Life cycle duration : ≤ 1 year
Reproduction : clutches, cemented or fixed
Potential number of cycles per year : 1

up   down

Position

**Stepwise_agg
(mean)**

Scores on constrained axis

0.5
0.0
-0.5

Voltinism harm. :
 bi- or multivoltine
Reproduction harm. :
 ovoviviparity
Respiration harm. :
 gills

Voltinism harm. :
 univoltine
Reproduction harm. :
 aquatic eggs

up   down

Position

**Stepwise_agg
(median)**

0.5
0.0
-0.5

Voltinism harm. :
 bi- or multivoltine
Reproduction harm. :
 ovoviviparity
Respiration harm. :
 gills

Voltinism harm. :
 univoltine
Reproduction harm. :
 aquatic eggs

up   down

Position

**Direct_agg
(mean)**

Scores on constrained axis

0.5
0.0
-0.5

Reproduction harm. :
 ovoviviparity
Voltinism harm. :
 bi- or multivoltine
Life cycle duration :
 > 1 year
Respiration harm. :
 gills

Life cycle duration :
 ≤ 1 year
Voltinism harm. :
 univoltine
Reproduction harm. :
 aquatic eggs

up   down

Position

**Direct_agg
(median)**

0.5
0.0
-0.5

Voltinism harm. :
 bi- or multivoltine
Reproduction harm. :
 ovoviviparity
Life cycle duration :
 > 1 year
Respiration harm. :
 gills

Life cycle duration :
 ≤ 1 year
Voltinism harm. :
 univoltine
Reproduction harm. :
 aquatic eggs

up   down

Position

Weighted_agg

Scores on constrained axis

0.5
-0.5

Reproduction harm. :
 ovoviviparity
Voltinism harm. :
 bi- or multivoltine
Life cycle duration :
 > 1 year
Respiration harm. :
 gills

Life cycle duration :
 ≤ 1 year
Voltinism harm. :
 univoltine
Reproduction harm. :
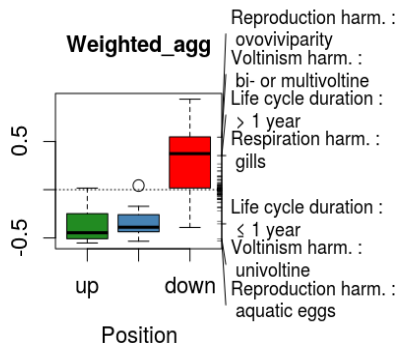 aquatic eggs

up   down

Position

Figure 7: 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 to unfold the full potential of trait-based approaches, e.g. for studying community trait responses to environmental gradients worldwide. Our attempt of harmonising invertebrate trait information from trait databases of different regions showed that grouping feature harmonisation is labour-intensive because grouping features are differentiated differently into traits between trait databases and various codings are used to describe traits. In addition, the same traits are sometimes defined differently in the investigated databases, 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 et al. 2011; Brink et al. 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 et al. 2017. The core of this harmonisation initiative is to provide standardised trait definitions and to draw connections to synonyms, related terms, and surrounding concepts by linking to other controlled vocabularies or ontologies in the field. Such initiatives could be a role model for freshwater ecologists to establish unambiguous terminologies for invertebrate traits. The existing freshwater invertebrate trait databases could be linked through standardised terminologies or ontologies, as suggested by Baird et al. Baird et al. 2011 almost 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 et al. 2019. Once standardised trait definitions are established, these will improve trait data sharing, trait data processing when working with multiple trait databases, and ultimately interpretation of the derived results.

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 trait 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 on species-level, considerable portions of trait data from North America and Australia were on genus or family-level (Table 2). Assigned traits on family-level might not reflect the real trait diversity within a taxon, for example for Chironomidae Serra et al. 2016, February 1. However, species identification can be complicated and time-consuming Marshall et al. 2006, November 1; Resh 2008, March, and therefore trait information on family-level has been widely used, e.g. in bioassessment Beketov et al. 2009, June 1.

Trait aggregation

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

methods and compared the results to traits 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 on species or genus-level to a point estimate on family-level suggests a precision that is not necessarily present, especially not for traits with a high variability or if trait information on species-level is missing. Some traits can vary strongly on more precise taxonomic levels than the family-level. For example, Monaghan and Soares Monaghan and Soares 2013, June found a high intra-family trait dispersion in the Tachet database for traits of the grouping features body size, flow preference, and reproduction. Furthermore, studies focusing on the lability of traits, i.e. how much traits are unconstrained 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 N. L. Poff et al. 2006; Wilkes et al. 2020, October 18. However, if trait aggregation is necessary our study can give guidance, under which circumstances the choice of aggregation method might 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) that the different weightings tested exerted a minor impact on the aggregated trait affinities. We obtained for both datasets more 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 in 47 % of the cases a difference to the assigned traits, 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). The differences can be explained by the binary coding used in the North American trait dataset and in the assigned traits. As a result, traits have affinities of either 1 or 0. Therefore, traits of a particular grouping feature have 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 will often result in a value of 0 or 1, while using the mean will often result in 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 entries. From the taxa that were compared from the North American trait dataset, 14 % were resolved on family-level and 62 % were resolved on genus-level.

52 % had 5 or less genera and 13 % contained just one genus (Figure S2). In the Australian dataset, 21 % of the taxa were resolved on family-level, 40 % were resolved on genus level, and 39 % were resolved on species-level. 68 % of the taxa contain 5 or less genera and 40 % just contained one genus (Figure S3). Hence, these results might change when more trait information on 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 yielded only slightly different results compared to the original study, i.e. some of the traits that responded in the original analysis did not respond in the re-analysis. Those traits that were non-responsive in the re-analysis were those closest to the criterion that defined when a trait was associated with either high or low salinity (Mahalanobis distance greater than the 97.5 % quantile of the Chi-square distribution, see method section), 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. One reason for the non-responsiveness of shredders in the re-analysis was likely the harmonisation procedure. During harmonisation the trait shredder has been amalgamated based on three traits (miner, xylophagus, and shredder). As a result, the trait affinities in the original data have a higher mean and standard deviation than in the aggregated and harmonised data (Table S2) suggesting that the signal in the original data has presumably been weakened by the harmonisation. This might be the reason that shredders showed a less strong reaction to the salinity gradient in the re-analysis. 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 analysis. The fact that the family-level aggregated traits used in the re-analysis showed only slightly different results compared to the original analysis is in line with the findings of others that family-level traits can be sensitive enough to detect environmental impacts Beketov et al. 2009, June 1.

Outlook

Although comprehensive invertebrate trait databases have been developed for several regions in the past, data synthesis is difficult due to discrepancies in trait definitions. By providing an overview of these definition discrepancies we hope to 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 on large 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 in the course of trait standardisation.

With the increasing use of computer vision in ecology, both species identification and trait quantification could be automated Milošević et al. 2020, April; Høye et al. 2020, July 4. This might result in more observational data and trait data on species-level in the future. However, if traits need to be aggregated we showed that trait aggregation is often in conclusion with expert assignments, specifically aggregation approaches using the median. Which aggregation method the objectively

better measure is cannot be answered, but when trait aggregation is necessary 1) median aggregation seems to be favourable, 2) highly variable traits and taxa where a few genera have a high number of species should be treated carefully, 3) coding of traits should be considered, and 4) weighting trait information based on the taxa present in the used trait database does not seem to have much influence. No matter which aggregation approach is chosen, for future studies in which researchers must aggregate traits, a measure of trait dispersion should be reported to indicate the uncertainty of the aggregated estimate.

References

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

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

Bäthe, J. & Coring, E. (2011, May 1). Biological effects of anthropogenic salt-load on the aquatic fauna: A synthesis of 17 years of biological survey on the rivers werra and weser. *Limnologica*. Salinisation of running waters, *41*(2), 125–133. doi:10.1016/j.limno.2010.07.005

Beketov, M. A., Foit, K., Schäfer, R. B., Schriever, C. A., Sacchi, A., Capri, E., . . . Liess, M. (2009, June 1). SPEAR indicates pesticide effects in streams – comparative use of species- and family-level biomonitoring data. *Environmental Pollution*, *157*(6), 1841–1848. doi:10.1016/j.envpol.2009.01.021

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

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

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

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

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

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

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

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

Heino, J., Schmera, D., & Erős, T. (2013). A macroecological perspective of trait patterns in stream communities. *Freshwater Biology*, *58*(8), 1539–1555. doi:10.1111/fwb.12164

Høye, T. T., Ärje, J., Bjerge, K., Hansen, O. L. P., Iosifidis, A., Leese, F., ... Raitoharju, J. (2020, July 4). Deep learning and computer vision will transform entomology. *bioRxiv*, 2020.07.03.187252. Publisher: Cold Spring Harbor Laboratory Section: New Results. doi:10.1101/2020.07.03.187252

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

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

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

Marshall, J., Steward, A., & Harch, B. (2006, November 1). Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an australian dryland river: The benefits and costs of using species abundance data. *Hydrobiologia*, *572*, 171–194. doi:10.1007/s10750-005-9007-0

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

Menezes, S., Baird, D. J., & Soares, A. M. V. M. (2010). Beyond taxonomy: A review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. *Journal of Applied Ecology*, *47*(4), 711–719. doi:10.1111/j.1365-2664.2010.01819.x

Milošević, D., Milosavljević, A., Predić, B., Medeiros, A. S., Savić-Zdravković, D., Stojković Piperac, M., ... Leese, F. (2020, April). Application of deep learning in aquatic bioassessment: Towards automated identification of non-biting midges. *Science of The Total Environment*, *711*, 135160. doi:10.1016/j.scitotenv.2019.135160

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

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

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

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

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

Poff, N. L., Olden, J. D., Vieira, N. K. M., Finn, D. S., Simmons, M. P., & Kondratieff, B. C. (2006). Functional trait niches of north american lotic insects: Traits-based ecological applications

in light of phylogenetic relationships. *Journal of the North American Benthological Society*, *25*(4), 730–755. doi:10.1899/0887-3593(2006)025[0730:FTNONA]2.0.CO;2

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

Schäfer, R. B., Kefford, B. J., Metzeling, L., Liess, M., Burgert, S., Marchant, R., . . . Nugegoda, D. (2011). A trait database of stream invertebrates for the ecological risk assessment of single and combined effects of salinity and pesticides in south-east australia. *Science of The Total Environment*, *409*(11), 2055–2063. doi:10.1016/j.scitotenv.2011.01.053

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

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

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

Serra, S. R. Q., Cobo, F., Graça, M. A. S., Dolédec, S., & Feio, M. J. (2016, February 1). Synthesising the trait information of european chironomidae (insecta: Diptera): Towards a new database. *Ecological Indicators*, *61*, 282–292. doi:10.1016/j.ecolind.2015.09.028

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

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

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

Twardochleb, L. A., Hilter, E., Pyne, M., Bills, P., & Zarnetske, P. L. (2020). Freshwater insect occurrences and traits for the contiguous united states, 2001 - 2018 ver 5. environmental data initiative.
https://doi.org/10.6073/pasta/8238ea9bc15840844b3a023b6b6ed158.

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

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

Usseglio-Polatera, P. (2020b). Personal information on body form traits. Personal information.

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

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

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

Wilkes, M. A., Edwards, F., Jones, J. I., Murphy, J. F., England, J., Friberg, N., ... Brown, L. E. (2020, October 18). Trait-based ecology at large scales: Assessing functional trait correlations, phylogenetic constraints and spatial variability using open data. *Global Change Biology*, gcb.15344. doi:10.1111/gcb.15344

Supporting Information

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 North America (Vieira) database contains the trait Bluff (blocky) which does not appear in the other databases.

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

31

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 *at least* two years"	"< 1 generation per year"	"< 1 generation per year"	"< 1 generation per year"	"< 1 reproductive cycle per year"
Multivoltine	"*Three* or more generations per year" [b]	"Able to complete *at least* two successive generations per year"	"> 1 generations per year"	"> 1 generations per year"	• 1-2 generations per year • bi/multivoltine • up to 5 generations per year • up to 10 generations per year	"> 1 reproductive cycles per year"

Locomotion swimming	• Passive movement like floating or drifting (trait swimming/scating) • Active movement (trait swimming/diving)	• Surface swimmers (over and under the water surface) • Full water swimmers (e.g. Baetidae).	Swimmer "Adapted for "fish-like" swimming"	Distinguishes swimmer and skater	Swimmers (water column)
Locomotion burrowing	"Burrowing in *soft* substrates or boring in *hard* substrates"	• Burrowing "within the first centimeters of the benthic fine sediment" • Differentiates also the trait interstitial (endobenthic)	Burrower "Inhabiting *fine* sediment of streams and lakes"	"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"	-	Sprawler Sprawling: "inhabiting the surface of floating leaves of vascular hydrophytes or fine sediments"	-	-

Trait						
Locomotion crawling	-	"Crawling over the bottom substrate"	Defined as crawling on the surface of floating leaves or fine sediments on the bottom	-	Database contains traits crawler, 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	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 [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.

Differences in trait affinities obtained by trait aggregation methods compared to traits assigned at family-level

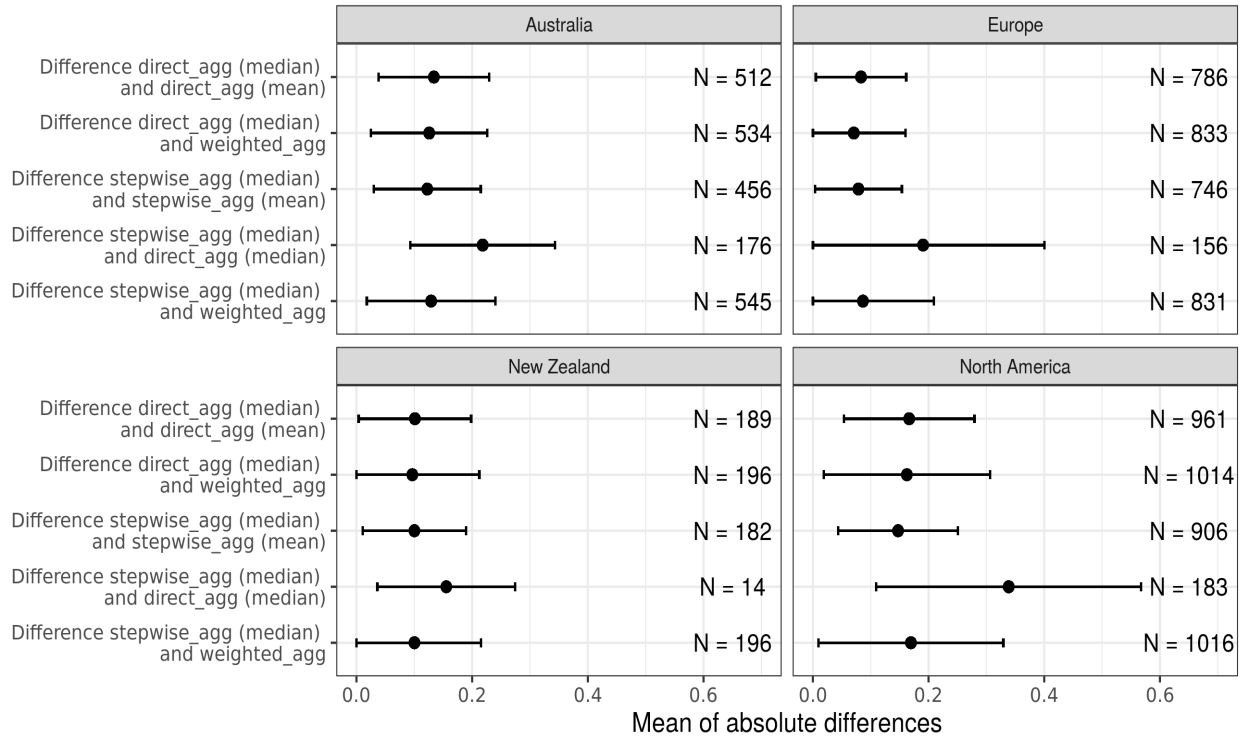**Comparison of the trait aggregation methods with each other**

Figure S1: Comparison of trait aggregation methods when aggregating over all traits for all datasets. Displayed are means of absolute differences in trait affinities with standard deviations (truncated at 0). Compared aggregation methods are displayed on the y-axis. N indicates the number of cases where differences occurred. Total number of cases: Australia 2223, Europe 3352, New Zealand 777, and North America 4080.

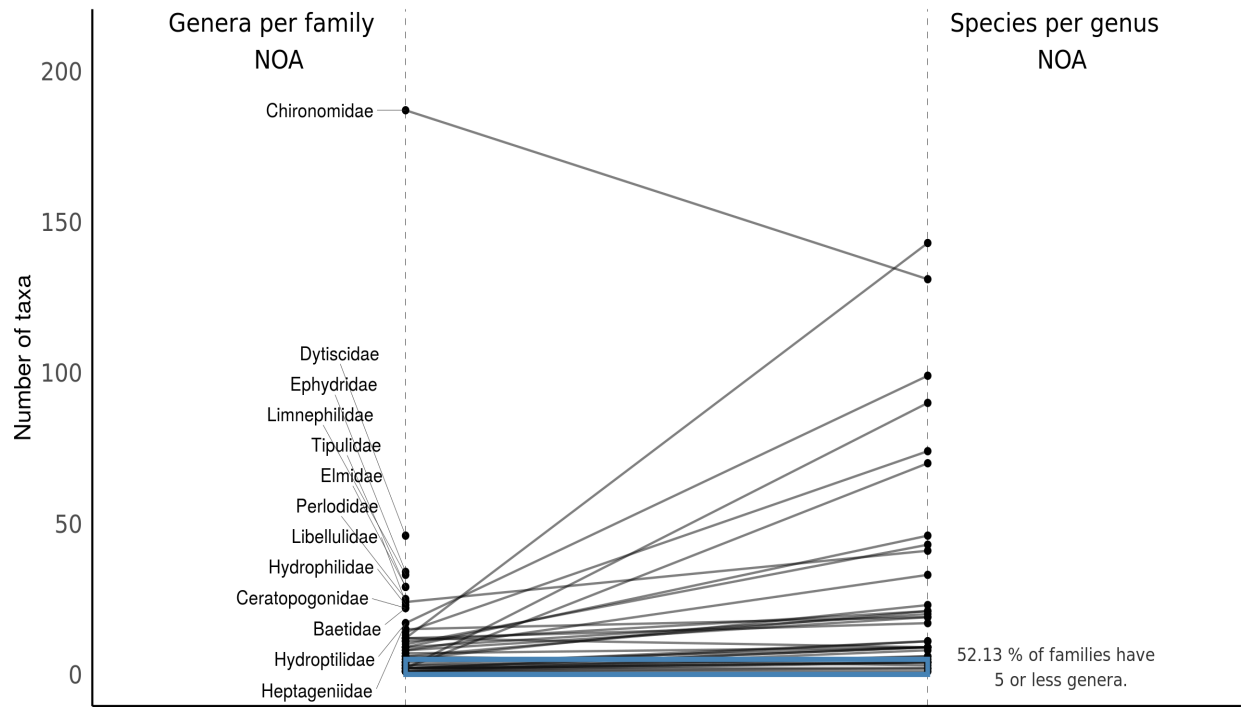**Taxonomic hierarchy in the trait datasets used for comparisons with assigned traits at family-level**

Figure S2: 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.
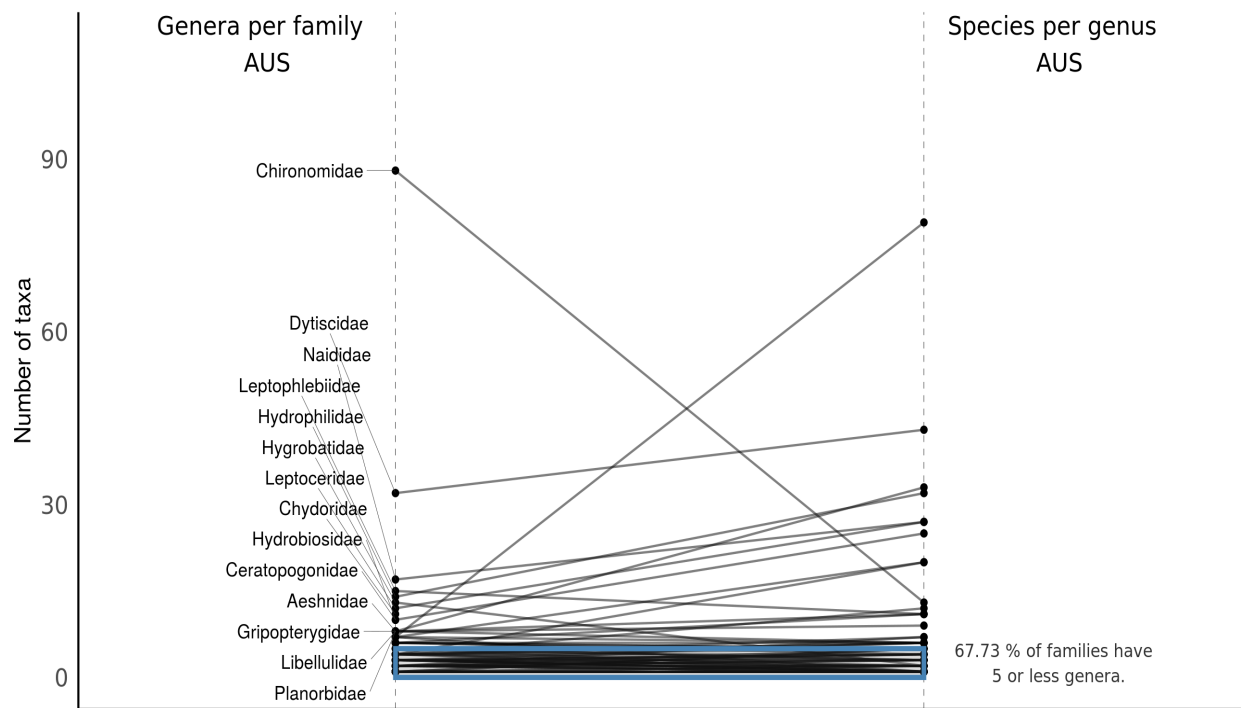
Figure S3: Number of genera per family and species per genus for the Australian trait dataset. For better visual display only families with more than 7 genera are displayed.

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

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

Dataset	Trait	Mean	Median	SD
Stepw_median	Shredder	0.20	0.14	0.25
Stepw_mean	Shredder	0.18	0.12	0.22
Direct_median	Shredder	0.21	0.14	0.25
Direct_mean	Shredder	0.19	0.14	0.22
Weighted	Shredder	0.19	0.14	0.22
Harmonized; not_aggregated	Shredder	0.18	0.12	0.24
Original	Shredder	0.25	0.14	0.32
Stepw_median	Gills	0.30	0.27	0.32
Stepw_mean	Gills	0.29	0.22	0.32
Direct_median	Gills	0.30	0.30	0.32
Direct_mean	Gills	0.30	0.30	0.32
Weighted	Gills	0.30	0.30	0.32
Harmonized; not_aggregated	Gills	0.30	0.25	0.32
Original	Gills	0.28	0.00	0.33
Stepw_median	Short life cycle	0.64	0.75	0.39
Stepw_mean	Short life cycle	0.64	0.79	0.39
Direct_median	Short life cycle	0.67	0.75	0.37
Direct_mean	Short life cycle	0.67	0.79	0.38
Weighted	Short life cycle	0.67	0.79	0.38
Harmonized; not_aggregated	Short life cycle	0.64	0.75	0.40
Original	Short life cycle	0.64	0.75	0.40
Stepw_median	Long life cylce	0.36	0.25	0.39
Stepw_mean	Long life cylce	0.36	0.21	0.39
Direct_median	Long life cylce	0.33	0.25	0.37
Direct_mean	Long life cylce	0.33	0.21	0.38
Weighted	Long life cylce	0.33	0.21	0.38
Harmonized; not_aggregated	Long life cylce	0.36	0.25	0.40
Original	Long life cylce	0.36	0.25	0.40