DRAFT: Harmonized and trait aggregation paper

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 [1], might be beneficial in achieving this goal [2]. 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 [3, 4]. 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 [5, 6]. 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 [7]. 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 [8–11].

In the last decades, freshwater ecologists have compiled comprehensive invertebrate trait databases for various continents [12–17]. 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. [5] compared trait composition for Mediterranean and temperate regions in Europe using traits from Usseglio-Polatera et al. [12](typically referred to as Tachet database); Poff et al. [8] characterized trait composition across sites in the Western US using traits from Poff et al. [18], and Botwe et al. [19] investigated the effect of salinity on invertebrate traits in different sites in South Australia using trait data from Poff et al. [18] and Schäfer et al. [20]. 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., 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" [21]. 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. [22] harmonized 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 [23]. In the best case, grouping features would be classified into the same traits across databases or they could easily be harmonized using standardized terminology. However, a lack of standardized terminology of trait definitions and poor metadata quality in many trait databases are common issues throughout the field of trait-based ecology [6, 24]. 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 [25]. For example, in the study from Brown et al. [22] 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). A centralized database with standardized 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. [9, 26]). 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. [9, 18, 27, 28]. Studies have used different methods of trait aggregation, e.g. the mean [29], median [9] or the mode [27]. 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. [9] using trait data from the established harmonized European trait datasets and aggregated traits. By comparison with the original analysis, we investigated how harmonizing and aggregating trait data can modify the outcome of trait-environment relationships.

Methods

Selection of traits and harmonization of trait databases

We extracted information from 6 trait databases from Europe, North America, Australia, and New Zealand and harmonized 7 grouping features. Trait information for Europe was obtained from the freshwaterecology.info database [13] and the Tachet database [12]. 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. [30] and complemented by Vieira et al. [14]. Data on body form for European and North American taxa was based on expert knowledge [31]. For Australia and New Zealand, we used trait databases from Kefford et al. [16] and Philips and Smith respectively [15]. 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 harmonized them into 26 traits (Table 1). Harmonization 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 harmonized 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 standardized 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 harmonization 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 harmonized grouping features. The last column indicates traits that were amalgamated for harmonization (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. [14] database and was newly classified by expert knowledge into cylindrical and flattened [31].

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 [12]. Taxa exhibiting this trait have been assigned to predators or herbivores based on expert knowledge [32].

Trait aggregation

Traits of the harmonized 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).

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 scenario (25 species per family $*5$ levels of trait variability $*100$ replicates). All trait aggregation methods were applied to each simulated dataset resulting in 500 aggregated trait values per scenario. The results were compared based on the produced range of aggregated trait affinities between levels of trait variability and hypothetical taxonomic hierarchies. 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 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
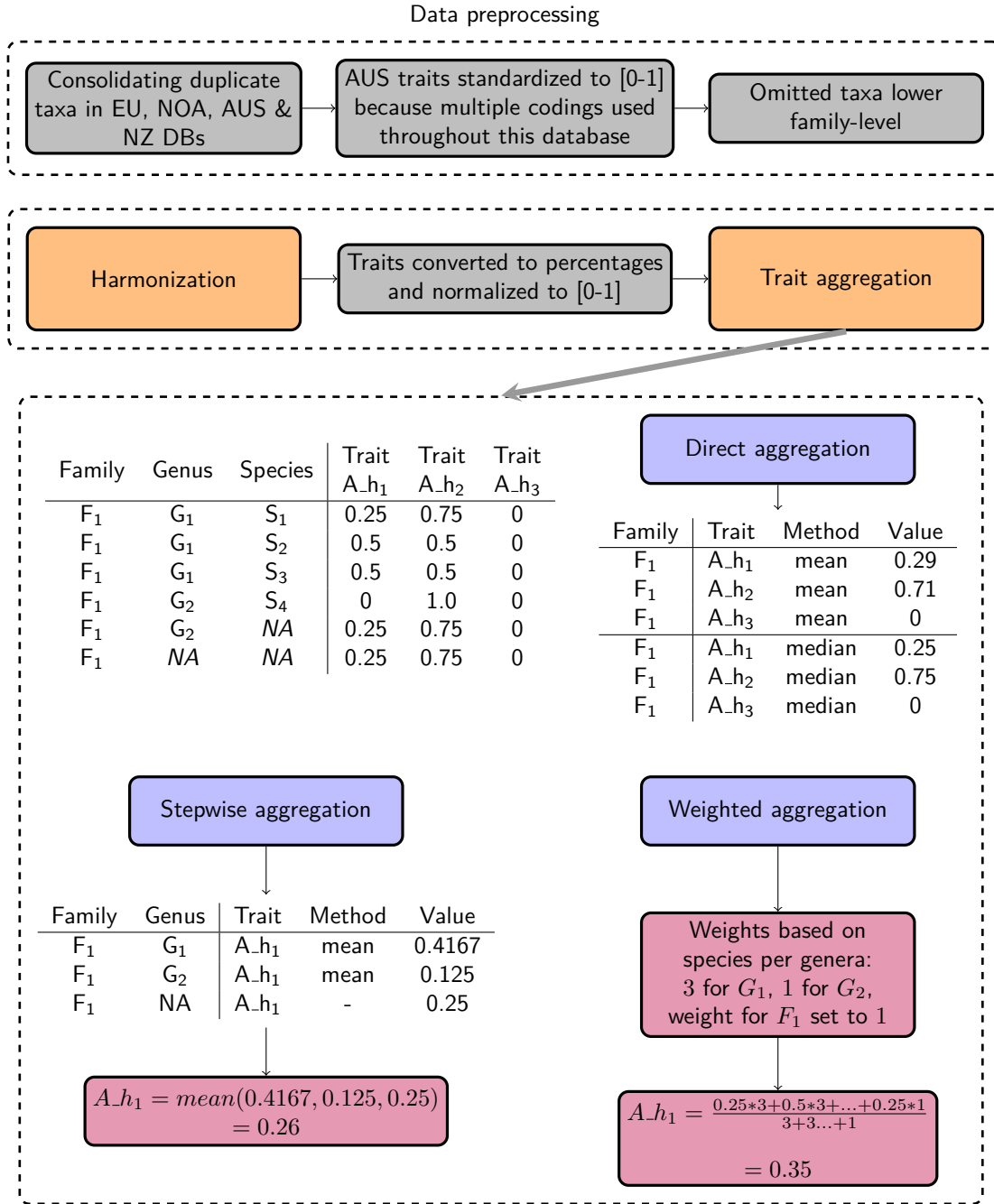
Data preprocessing

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

| Harmonization | → | 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.

using data from Chessman et al. [33]. In Chessman et al. [33] 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. [33]. 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.) [30] 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 harmonization and trait aggregation on trait-environment relationships

We repeated the analysis of Szöcs et al. [9] who studied the effect of anthropogenic salinization on invertebrates in the River Werra in Germany. For the re-analysis we used the established harmonized grouping features 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 [34]. 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. [9].

We substituted 6 of the grouping features from the original data with harmonized 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 harmonized 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. 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. [9] used. The 6 harmonized 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. [9] 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 harmonization 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 harmonized 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 harmonized 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 preprocessed European, North American, and Australian datasets (Table 3). By contrast, the New Zealand dataset contained complete trait information for most of the investigated grouping features (between 94 % and 100 %).

Table 3: Rounded percentage of entries that 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 [16]. 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	9	-	-
				binary		-	-
Tachet	7	3	8	5	8	7	-
	fuzzy [0 − 3]		fuzzy [0−5]	fuzzy [0 − 3]			-
North America (Twardochleb)	6	3	10	3	10	3	-
North America (Vieira)	8	3	9	8	10	3	5
			binary				
Australia	16 [a]	7	9	10	13 [b]	9	4
	binary; proportional [0 − 1]; fuzzy [0 − 3]		binary; fuzzy [0 − 3]	binary; proportional scale [0 − 1]; fuzzy [0 − 3]	categorical	binary; continuous; fuzzy [0 − 3]	fuzzy codes [0 − 3]
New Zealand	6	3	4	4	4	5	4
			fuzzy [0 − 3]				

[a] Some of the traits were similar (e.g. trait *Shredder*, *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.

The trait aggregation methods $direct\_agg_{mean}$, $direct\_agg_{median}$ and $weighted\_agg$ yielded similar ranges of aggregated trait affinities within each level of trait variability for the sim_base scenario. The $(stepwise\_agg_{median})$ and $(direct\_agg_{median})$ yielded mostly greater ranges of aggregated trait affinities for the sim_base scenario. As expected with increasing trait variability the produced range of trait affinities increased (Figure 2). 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.

Comparing the aggregated trait affinities, in total 15000 (500 aggregated traits $*10$ unique comparisons of trait affinities and 3 scenarios) for each simulated dataset showed that in most simulated datasets the differences between the aggregated traits were small. 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).

Overall, differences between the results of the aggregation methods for each simulated dataset increased with increasing trait variability. According to the simulation results, the different aggregation methods often result in similar trait affinities. If differences occurred, than mostly between aggregation methods using the median and aggregation methods using the mean.
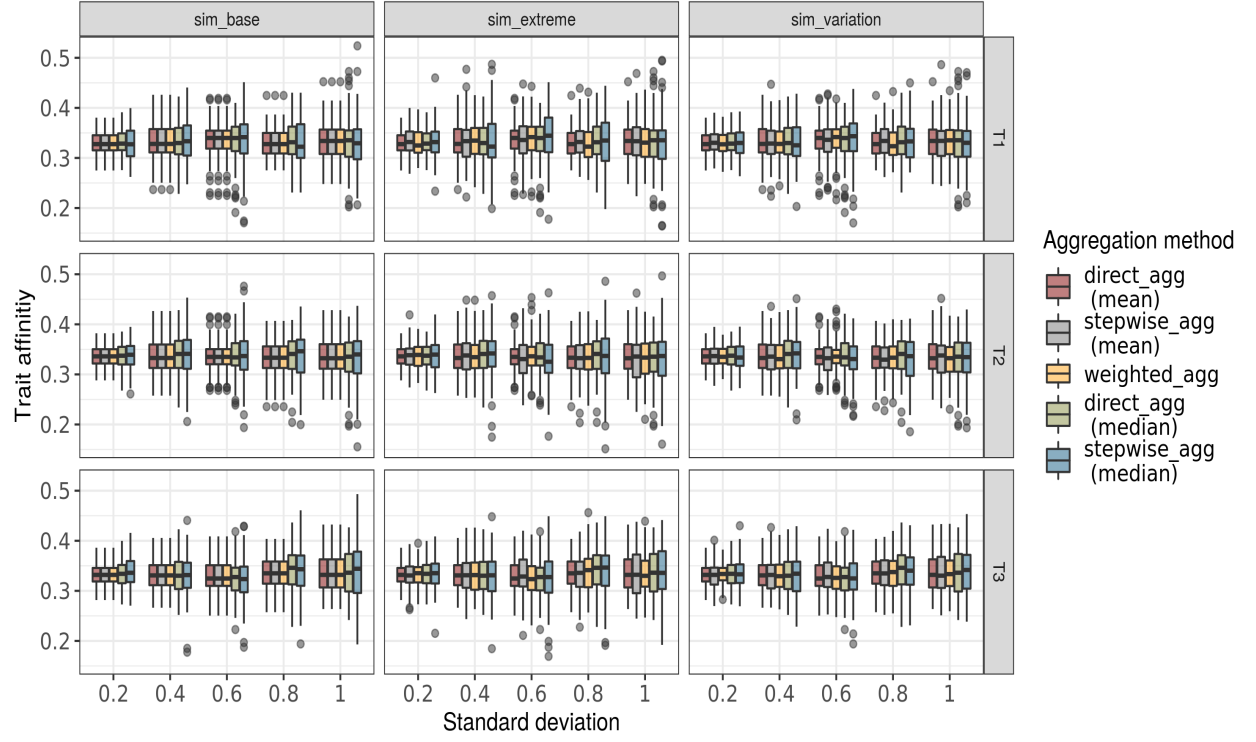
Figure 2: Ranges of aggregated trait affinities for the three examples of taxonomic hierarchies and simulated levels of trait variability. Boxplots depict results for each trait aggregation method of 100 simulations. T1, T2, and T3 are the simulated traits.
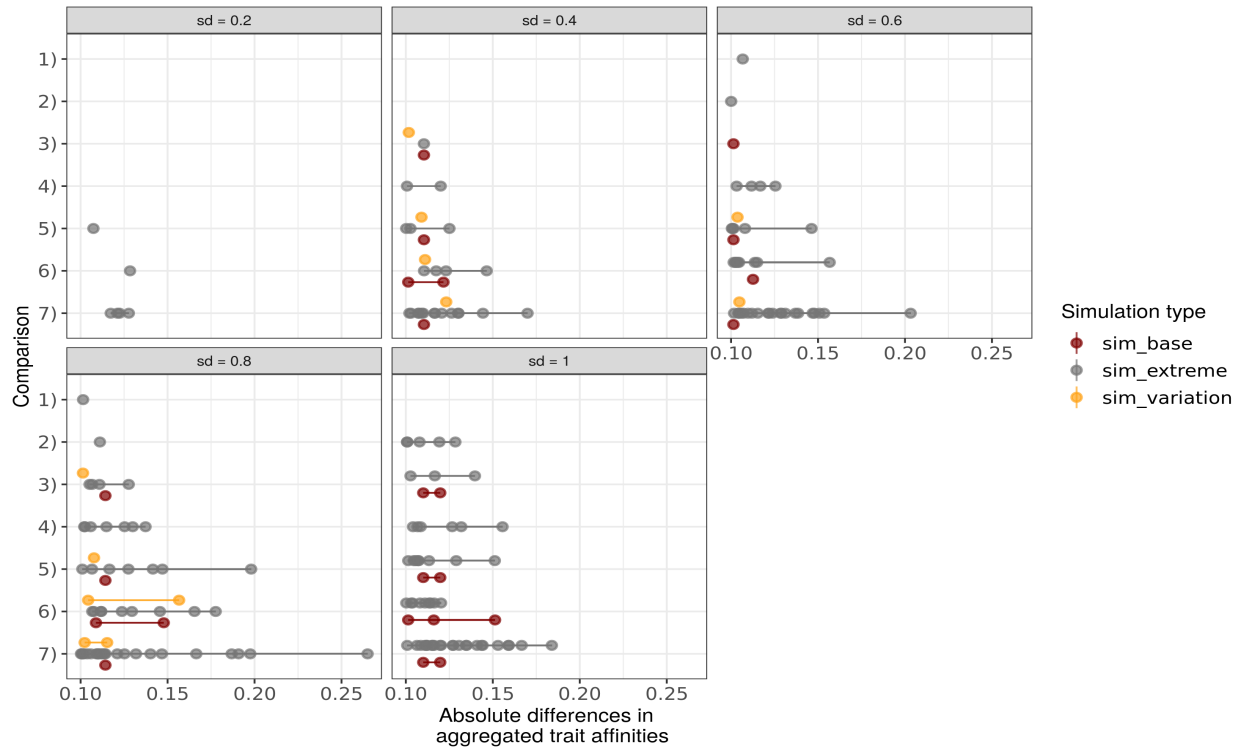
Figure 3: Comparison of the aggregated trait affinities produced by the different trait aggregation methods for every simulated dataset across all 3 simulated traits. Dots depict comparisons where absolute differences between aggregated trait affinities were greater than 0.1.

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

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, trait aggregation methods using the median yielded fewer cases with differences compared to approaches using the mean. However, aggregation methods using the median 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).

A comparison of the aggregation methods with each other for the 4 datasets revealed that differences in aggregated trait affinities were largest between the $stepwise\_agg_{median}$ and $direct\_agg_{median}$ (Figure S1).

Table 5: Amount of differing cases, the minimum and maximum, and 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
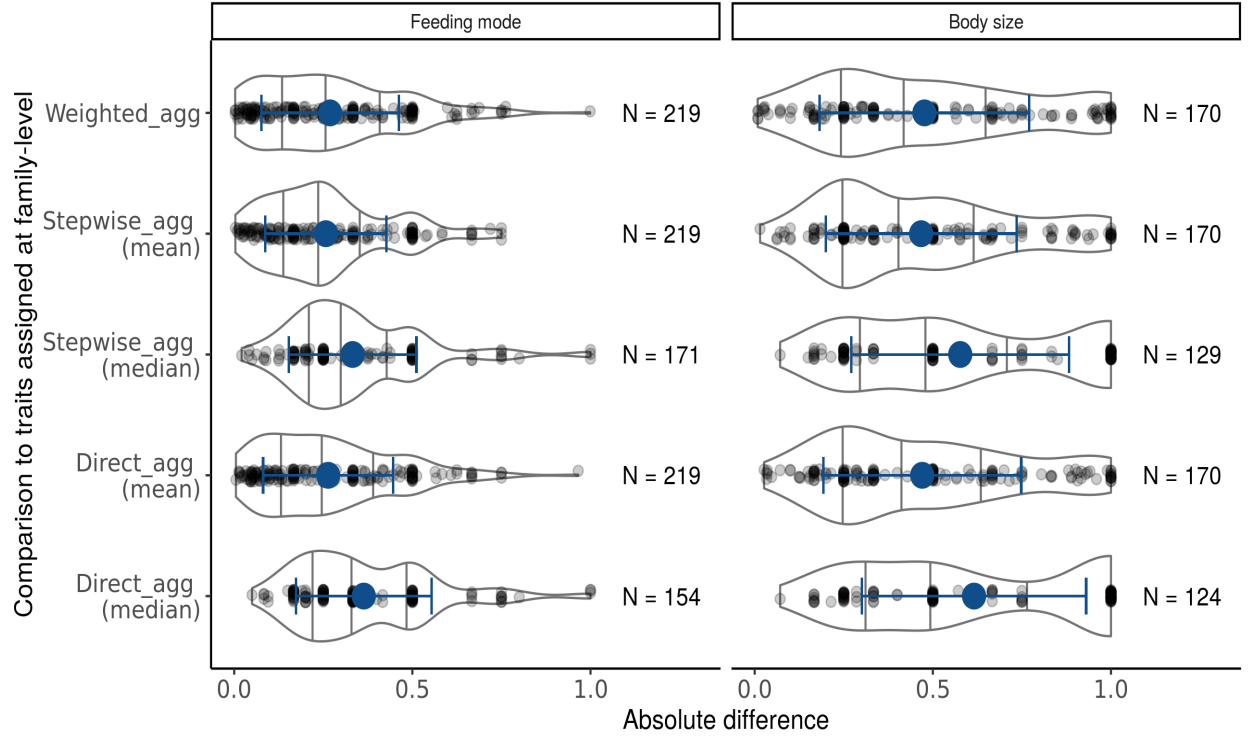
Figure 4: Display of the cases (factor combination investigated families and traits) where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the Australian dataset 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.
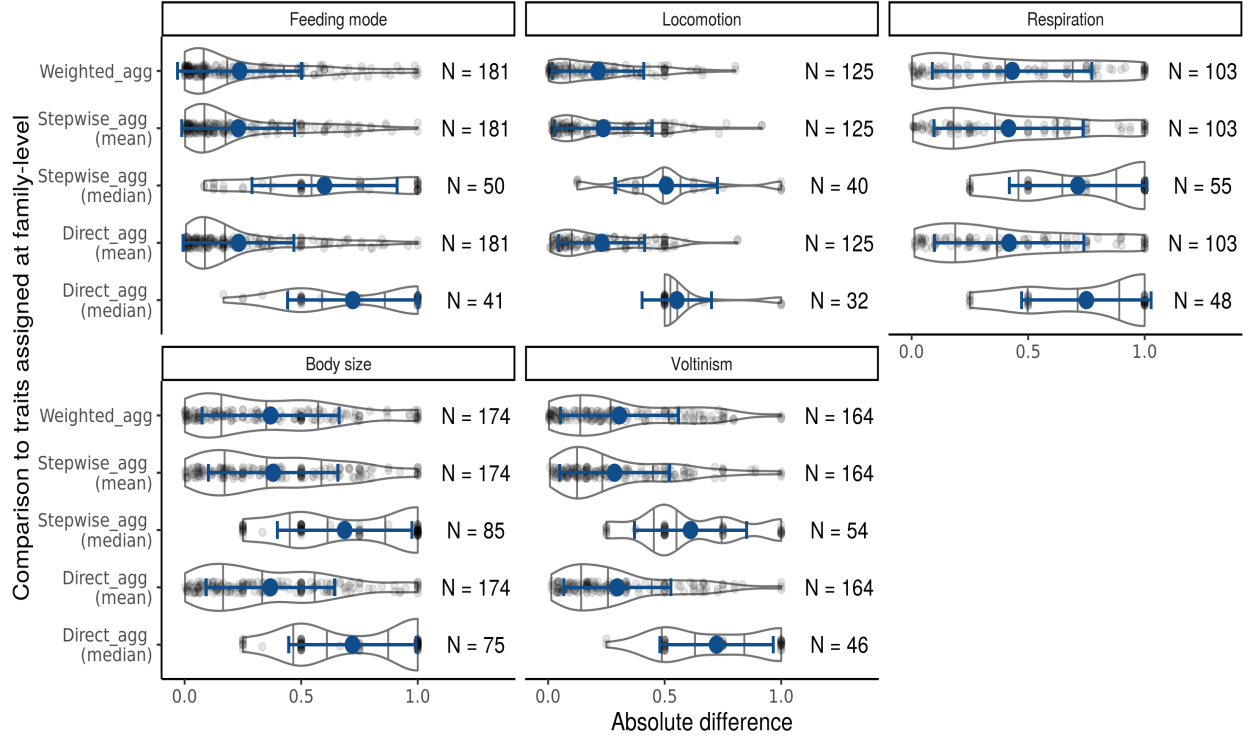
Figure 5: Display of the cases (factor combination investigated families and traits) where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the North America dataset 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 harmonized and aggregated grouping features

The original RDA in Szöcs et al. [9] 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 harmonized grouping features resulted in less traits that distinguish upstream and downstream sites in comparison to the original analysis (Figure and Figure). 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 harmonized 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 harmonized dataset showed also results similar to the original analysis but with fewer traits distinguishing upstream and downstream sites (Figure and Figure). 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.
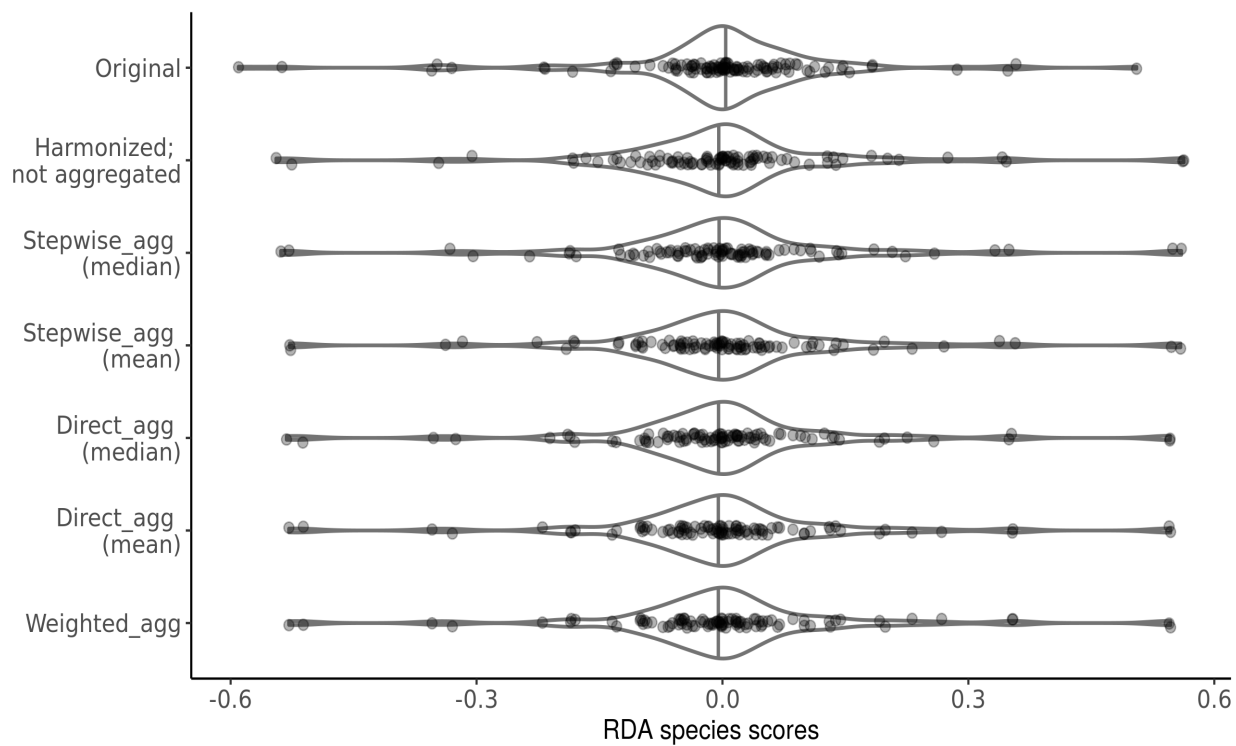
Figure 6: Species scores obtained by RDA from the original analysis [9], using harmonized grouping features, and using harmonized grouping features with traits aggregated to family-level. The dots represent the individual species scores for each analysed trait along the conductivity axis. The violin plot shows the density estimate of the species scores. Gray vertical lines indicate the median of the obtained species scores.
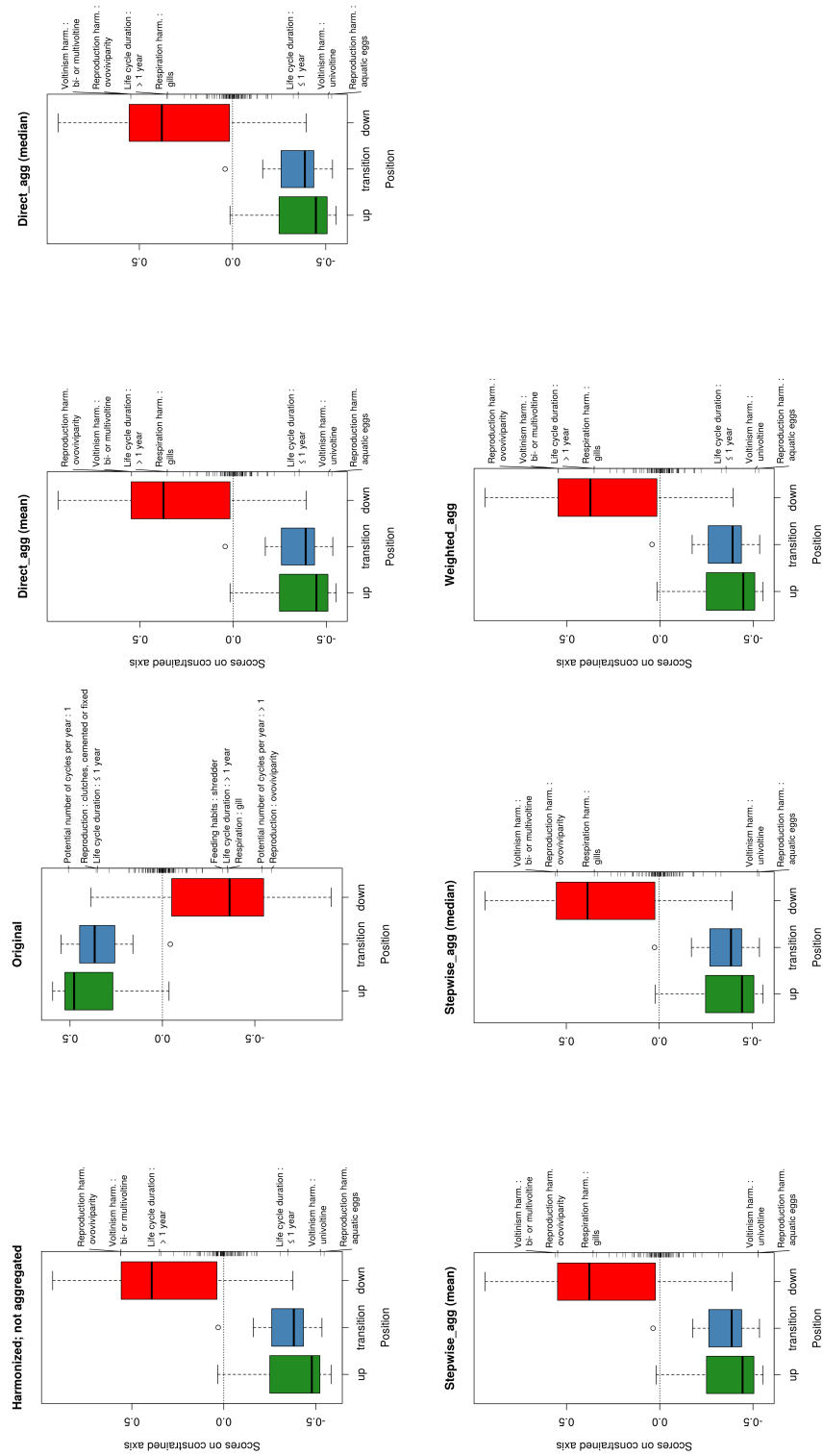
Figure 7: RDA of traits constrained by electric conductivity for the tested methods and the original study. Shown are boxplots of the site scores along the conductivity axis. The rug on the right side of each plot indicates species scores of the traits on the conductivity axis. Only traits with a mahalanobis distance greater than the 97.5% quantile of the Chi-square distribution (5.02) were labelled.

Discussion

Synthesizing trait information from various 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 trait information from multiple trait databases showed that grouping feature harmonisation is labour-intensive because grouping features are differentiated differently into traits between trait databases. In addition, the same traits are sometimes defined differently in the investigated databases, requiring expert knowledge to amalgamate these traits (e.g. the trait piercer in the Tachet database, Table 1). Thus, the field of freshwater ecology needs terminological standards to facilitate data synthesis. Harmonised definitions and concepts of traits have been for example developed for plants with the *Thesaurus Of Plant characteristics* (TOP) initiative [35]. The core of this harmonisation initiative is not only to provide standardized trait definitions but also 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 science to establish unambiguous terminologies for invertebrate traits. The existing freshwater invertebrate trait databases could then be linked through standardised terminologies or ontologies, as suggested by Baird et al. almost a decade ago [6]. By following the recently proposed *Ecological Trait-data Standard Vocabulary* (ETS) providers of invertebrate trait data could connect their traits to such ontologies [23]. Once standardised trait definitions are established, these will improve trait data sharing, trait data processing when working with multiple trait databases, and ultimately improve interpretation of derived results.

Our analysis aimed to use most of the available trait information for different regions to establish harmonised grouping feature datasets. Although many taxa are coverd in the trait databases, the availability of trait information varies strongly across grouping features and taxonomic resolution differs between databases.

As outlined in the introduction, various trait aggregation methods are used by researchers when trait data are on more precise taxonomic levels than observational data. 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 which 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 found a high intra-family trait dispersion in the Tachet database for traits of the grouping features body size, flow preference, and reproduction [36]. 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 [18, 37]. 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) aggregation of traits using $direct\_agg_{median}$ and $stepwise\_agg_{mean}$ (median aggregation methods) compared to $direct\_agg_{mean}$, $stepwise\_agg_{mean}$, and $weighted\_mean$ (mean aggregation methods) is 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 %, Table 5). However, when there were differences between median aggregation methods and assigned traits, these differences were greater compared to the mean aggregation methods, particularly for the North American trait dataset. The mean absolute differences for the median aggregations was twice as high (0.63 - 0.7) compared to the mean aggregations (0.3 - 0.31, Table 5). The differences can be explained by the binary coding used in the North American trait dataset and in the assigned traits for North America. 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 in many cases either to agreement with the assigned traits or the difference between median aggregated traits and assigned traits was often 1. The number of differing cases and the mean absolute differences to the assigned traits were similar within the mean aggregation and within the median aggregation methods, which confirms our second finding, the small influence of weighting on the aggregation methods. Furthermore, the distributions of absolute trait affinity differences to assigned traits obtained were relatively similar for the mean aggregation methods and the median aggregation methods in both tested datasets (Figure 4 and Figure 5). 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 to the assigned traits at family-level, 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).

Although the simulation results showed only small differences between the results of the aggregation methods the choice of the aggregation method for small numbers of genera per family can be important when one genus has a much larger number of species compared to other genera within a particular family.

Maybe include: How often does such a taxonomic data occur in the trait datasets?

In the re-analysis where we investigated how using harmonised and aggregated traits might change the results of trait-environment relationships we found similar results compared the original study. Traits that responded only in the original analysis but were non-response 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 methods), namely feeding mode shredder, respiration gills, and life cycle duration traits. For example, 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 dataset have a higher mean and standard deviation than in the aggregated and harmonized datasets (Table SS2). Hence, the signal in the original data for the trait shredder has presumably been weakened by the harmonization, which might be the reason that shredders showed a less strong reaction to the salinity gradient.

21

References

[1] B. J. McGill, B. J. Enquist, E. Weiher, and M. Westoby, "Rebuilding community ecology from functional traits," *Trends in Ecology & Evolution*, vol. 21, no. 4, pp. 178–185, 2006, ISSN: 0169-5347. DOI: `10.1016/j.tree.2006.02.002`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0169534706000334`.

[2] Heino Jani, Schmera Dénes, and Erős Tibor, "A macroecological perspective of trait patterns in stream communities," *Freshwater Biology*, vol. 58, no. 8, pp. 1539–1555, May 28, 2013, ISSN: 0046-5070. DOI: `10.1111/fwb.12164`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/full/10.1111/fwb.12164` (visited on 04/14/2018).

[3] T. R. E. Southwood, "Habitat, the templet for ecological strategies?" *Journal of Animal Ecology*, vol. 46, no. 2, pp. 337–365, 1977, ISSN: 0021-8790. DOI: `10.2307/3817`. [Online]. Available: `https://www.jstor.org/stable/3817`.

[4] W. C. E. P. Verberk, C. G. E. van Noordwijk, and A. G. Hildrew, "Delivering on a promise: Integrating species traits to transform descriptive community ecology into a predictive science," *Freshwater Science*, vol. 32, no. 2, pp. 531–547, 2013, ISSN: 2161-9549. DOI: `10.1899/12-092.1`. [Online]. Available: `https://www.journals.uchicago.edu/doi/10.1899/12-092.1`.

[5] N. Bonada, S. Dolédec, and B. Statzner, "Taxonomic and biological trait differences of stream macroinvertebrate communities between mediterranean and temperate regions: Implications for future climatic scenarios," *Global Change Biology*, vol. 13, no. 8, pp. 1658–1671, 2007, ISSN: 1354-1013, 1365-2486. DOI: `10.1111/j.1365-2486.2007.01375.x`. [Online]. Available: `http://doi.wiley.com/10.1111/j.1365-2486.2007.01375.x`.

[6] D. J. Baird, C. J. O. Baker, R. B. Brua, M. Hajibabaei, K. McNicol, T. J. Pascoe, and D. d. Zwart, "Toward a knowledge infrastructure for traits-based ecological risk assessment," *Integrated Environmental Assessment and Management*, vol. 7, no. 2, pp. 209–215, 2011, _eprint: https://setac.onlinelibrary.wiley.com/doi/pdf/10.1002/ieam.129, ISSN: 1551-3793. DOI: `10.1002/ieam.129`. [Online]. Available: `https://setac.onlinelibrary.wiley.com/doi/abs/10.1002/ieam.129`.

[7] O. Moog and A. Hartmann, "A comprehensive species inventory of austrian aquatic organisms with ecological notes," p. 86, 1995.

[8] N. L. Poff, M. I. Pyne, B. P. Bledsoe, C. C. Cuhaciyan, and D. M. Carlisle, "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*, vol. 29, no. 4, pp. 1441–1458, 2010, ISSN: 0887-3593. DOI: `10.1899/10-030.1`. [Online]. Available: `https://www.journals.uchicago.edu/doi/abs/10.1899/10-030.1`.

[9] E. Szöcs, E. Coring, J. Bäthe, and R. B. Schäfer, "Effects of anthropogenic salinization on biological traits and community composition of stream macroinvertebrates," *Science of The Total Environment*, vol. 468-469, pp. 943–949, 2014, ISSN: 00489697. DOI: `10.1016/j.scitotenv.2013.08.058`. [Online]. Available: `https://linkinghub.elsevier.com/retrieve/pii/S0048969713009728`.

[10] A. K. Bhowmik and R. B. Schäfer, "Large scale relationship between aquatic insect traits and climate," *PLoS ONE*, vol. 10, no. 6, 2015, ISSN: 1932-6203. DOI: `10.1371/journal.pone.0130025`. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4469582/`.

[11] S. Menezes, D. J. Baird, and A. M. V. M. Soares, "Beyond taxonomy: A review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring," *Journal of Applied Ecology*, vol. 47, no. 4, pp. 711–719, 2010, ISSN: 1365-2664. DOI: `10.1111/j.1365-2664.2010.01819.x`. [Online]. Available: `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2664.2010.01819.x`.

[12] P. Usseglio-Polatera, M. Bournaud, P. Richoux, and H. Tachet, "Biomonitoring through biological traits of benthic macroinvertebrates: How to use species trait databases?" *Hydrobiologia*, vol. 422, no. 0, pp. 153–162, 2000, ISSN: 1573-5117. DOI: `10.1023/A:1017042921298`. [Online]. Available: `https://doi.org/10.1023/A:1017042921298`.

[13] A. Schmidt-Kloiber and D. Hering, "Www.freshwaterecology.info – an online tool that unifies, standardises and codifies more than 20,000 european freshwater organisms and their ecological preferences," *Ecological Indicators*, vol. 53, pp. 271–282, 2015, ISSN: 1470-160X. DOI: `10.1016/j.ecolind.2015.02.007`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S1470160X15000758`.

[14] N. K. M. Vieira, N. L. Poff, D. M. Carlisle, S. R. M. II, M. L. Koski, and B. C. Kondratieff, "A database of lotic invertebrate traits for north america," p. 19, 2006.

[15] N. Philips and B. Smith, "New zealand freshwater macroinvertebrate trait database," *NIWA*, 2018.

[16] B. J. Kefford, P. K. Botwe, A. J. Brooks, S. Kunz, R. Marchant, S. Maxwell, L. Metzeling, R. B. Schäfer, and R. M. Thompson, "An integrated database of stream macroinvertebrate traits for australia: Concept and application," *Ecological Indicators*, vol. 114, p. 106 280, 2020, ISSN: 1470160X. DOI: `10.1016/j.ecolind.2020.106280`. [Online]. Available: `https://linkinghub.elsevier.com/retrieve/pii/S1470160X2030217X`.

[17] S. Tomanova, E. Goitia, and J. Helešic, "Trophic levels and functional feeding groups of macroinvertebrates in neotropical streams," *Hydrobiologia*, vol. 556, no. 1, pp. 251–264, 2006, ISSN: 1573-5117. DOI: `10.1007/s10750-005-1255-5`. [Online]. Available: `https://doi.org/10.1007/s10750-005-1255-5`.

[18] N. L. Poff, J. D. Olden, N. K. M. Vieira, D. S. Finn, M. P. Simmons, and B. C. Kondratieff, "Functional trait niches of north american lotic insects: Traits-based ecological applications in light of phylogenetic relationships," *Journal of the North American Benthological Society*, vol. 25, no. 4, pp. 730–755, 2006, ISSN: 0887-3593. DOI: `10.1899/0887-3593(2006)025[0730:FTNONA]2.0.CO;2`. [Online]. Available: `https://www.journals.uchicago.edu/doi/10.1899/0887-3593%282006%29025%5B0730%3AFTNONA%5D2.0.CO%3B2`.

[19] P. K. Botwe, S. Carver, R. Magierowski, P. McEvoy, P. Goonan, C. Madden, and L. A. Barmuta, "Effects of salinity and flow interactions on macroinvertebrate traits in temporary streams," *Ecological Indicators*, vol. 89, pp. 74–83, 2018, ISSN: 1470160X. DOI: `10.1016/j.ecolind.2018.01.036`. [Online]. Available: `https://linkinghub.elsevier.com/retrieve/pii/S1470160X18300384`.

[20] R. B. Schäfer, B. J. Kefford, L. Metzeling, M. Liess, S. Burgert, R. Marchant, V. Pettigrove, P. Goonan, and D. Nugegoda, "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*, vol. 409, no. 11, pp. 2055–2063, 2011, ISSN: 00489697. DOI: `10.1016/j.scitotenv.2011.01.053`. [Online]. Available: `http://linkinghub.elsevier.com/retrieve/pii/S0048969711001124`.

[21] D. Schmera, J. Podani, J. Heino, T. Erős, and N. L. Poff, "A proposed unified terminology of species traits in stream ecology," *Freshwater Science*, vol. 34, no. 3, pp. 823–830, 2015, ISSN: 2161-9549, 2161-9565. DOI: 10.1086/681623. [Online]. Available: https://www.journals.uchicago.edu/doi/10.1086/681623.

[22] L. E. Brown, K. Khamis, M. Wilkes, P. Blaen, J. E. Brittain, J. L. Carrivick, S. Fell, N. Friberg, L. Füreder, G. M. Gislason, S. Hainie, D. M. Hannah, W. H. M. James, V. Lencioni, J. S. Olafsson, C. T. Robinson, S. J. Saltveit, C. Thompson, and A. M. Milner, "Functional diversity and community assembly of river invertebrates show globally consistent responses to decreasing glacier cover," *Nature Ecology & Evolution*, vol. 2, no. 2, pp. 325–333, 2018, ISSN: 2397-334X. DOI: 10.1038/s41559-017-0426-x. [Online]. Available: http://www.nature.com/articles/s41559-017-0426-x.

[23] F. D. Schneider, D. Fichtmueller, M. M. Gossner, A. Güntsch, M. Jochum, B. König-Ries, G. Le Provost, P. Manning, A. Ostrowski, C. Penone, and N. K. Simons, "Towards an ecological trait-data standard," *Methods in Ecology and Evolution*, D. Orme, Ed., pp. 2041–210X.13288, 2019, ISSN: 2041-210X, 2041-210X. DOI: 10.1111/2041-210X.13288. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13288.

[24] W. D. Kissling, R. Walls, A. Bowser, M. O. Jones, J. Kattge, D. Agosti, J. Amengual, A. Basset, P. M. van Bodegom, J. H. C. Cornelissen, E. G. Denny, S. Deudero, W. Egloff, S. C. Elmendorf, E. Alonso García, K. D. Jones, O. R. Jones, S. Lavorel, D. Lear, L. M. Navarro, S. Pawar, R. Pirzl, N. Rüger, S. Sal, R. Salguero-Gómez, D. Schigel, K.-S. Schulz, A. Skidmore, and R. P. Guralnick, "Towards global data products of essential biodiversity variables on species traits," *Nature Ecology & Evolution*, vol. 2, no. 10, pp. 1531–1540, Oct. 2018, Number: 10 Publisher: Nature Publishing Group, ISSN: 2397-334X. DOI: 10.1038/s41559-018-0667-3. [Online]. Available: https://www.nature.com/articles/s41559-018-0667-3 (visited on 11/30/2020).

[25] J. M. Culp, D. G. Armanini, M. J. Dunbar, J. M. Orlofske, N. L. Poff, A. I. Pollard, A. G. Yates, and G. C. Hose, "Incorporating traits in aquatic biomonitoring to enhance causal diagnosis and prediction," *Integrated Environmental Assessment and Management*, vol. 7, no. 2, pp. 187–197, 2011, _eprint: https://setac.onlinelibrary.wiley.com/doi/pdf/10.1002/ieam.128, ISSN: 1551-3793. DOI: 10.1002/ieam.128. [Online]. Available: https://setac.onlinelibrary.wiley.com/doi/abs/10.1002/ieam.128.

[26] K. Voß and R. Schäfer, "Taxonomic and functional diversity of stream invertebrates along an environmental stress gradient," *Ecological Indicators*, vol. 81, pp. 235–242, 2017, ISSN: 1470160X. DOI: 10.1016/j.ecolind.2017.05.072. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1470160X17303412.

[27] Pilière A. F. H., Verberk W. C. E. P., Gräwe M., Breure A. M., Dyer S. D., Posthuma L., Zwart D., Huijbregts M. A. J., and Schipper A. M., "On the importance of trait interrelationships for understanding environmental responses of stream macroinvertebrates," *Freshwater Biology*, vol. 61, no. 2, pp. 181–194, 2016, ISSN: 0046-5070. DOI: 10.1111/fwb.12690. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/fwb.12690.

[28] T. W. H. Aspin, K. Khamis, T. J. Matthews, A. M. Milner, M. J. O'Callaghan, M. Trimmer, G. Woodward, and M. E. Ledger, "Extreme drought pushes stream invertebrate communities over functional thresholds," *Global Change Biology*, vol. 25, no. 1, pp. 230–244, 2019, ISSN:

1365-2486. DOI: `10.1111/gcb.14495`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.14495`.

[29] C. Magliozzi, P. Usseglio-Polatera, A. Meyer, and R. C. Grabowski, "Functional traits of hyporheic and benthic invertebrates reveal importance of wood-driven geomorphological processes in rivers," *Functional Ecology*, vol. 33, no. 9, J. Alahuhta, Ed., pp. 1758–1770, 2019, ISSN: 0269-8463, 1365-2435. DOI: `10.1111/1365-2435.13381`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2435.13381`.

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

[31] P. Usseglio-Polatera, *Personal information on body form traits.* Personal information, 2020.

[32] ——, *Classification of piercers as herbivores or predators.* Personal information, 2020.

[33] B. C. Chessman, "Dissolved-oxygen, current and temperature preferences of stream invertebrates estimated from field distributions: Application to assemblage responses to drought," *Hydrobiologia*, vol. 809, no. 1, pp. 141–153, 2018, ISSN: 1573-5117. DOI: `10.1007/s10750-017-3455-1`. [Online]. Available: `https://doi.org/10.1007/s10750-017-3455-1`.

[34] J. Bäthe and E. Coring, "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, vol. 41, no. 2, pp. 125–133, May 1, 2011, ISSN: 0075-9511. DOI: `10.1016/j.limno.2010.07.005`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0075951110000381` (visited on 10/16/2020).

[35] E. Garnier, U. Stahl, M.-A. Laporte, J. Kattge, I. Mougenot, I. Kühn, B. Laporte, B. Amiaud, F. S. Ahrestani, G. Bönisch, D. E. Bunker, J. H. C. Cornelissen, S. Díaz, B. J. Enquist, S. Gachet, P. Jaureguiberry, M. Kleyer, S. Lavorel, L. Maicher, N. Pérez-Harguindeguy, H. Poorter, M. Schildhauer, B. Shipley, C. Violle, E. Weiher, C. Wirth, I. J. Wright, and S. Klotz, "Towards a thesaurus of plant characteristics: An ecological contribution," *Journal of Ecology*, vol. 105, no. 2, pp. 298–309, 2017, _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2745.12698, ISSN: 1365-2745. DOI: `https://doi.org/10.1111/1365-2745.12698`. [Online]. Available: `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2745.12698` (visited on 11/30/2020).

[36] K. A. Monaghan and A. M. Soares, "Improving the accuracy of aquatic macroinvertebrate trait assignment — the importance of taxon-weighting," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 70, no. 6, J. Richardson, Ed., pp. 885–894, Jun. 2013, ISSN: 0706-652X, 1205-7533. DOI: `10.1139/cjfas-2012-0402`. [Online]. Available: `http://www.nrcresearchpress.com/doi/10.1139/cjfas-2012-0402` (visited on 11/16/2020).

[37] M. A. Wilkes, F. Edwards, J. I. Jones, J. F. Murphy, J. England, N. Friberg, D. Hering, N. L. Poff, P. Usseglio-Polatera, W. C. E. P. Verberk, J. Webb, and L. E. Brown, "Trait-based ecology at large scales: Assessing functional trait correlations, phylogenetic constraints and spatial variability using open data," *Global Change Biology*, gcb.15344, Oct. 18, 2020, ISSN: 1354-1013, 1365-2486. DOI: `10.1111/gcb.15344`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/10.1111/gcb.15344` (visited on 10/22/2020).

Supporting Information

Discrepancies in trait definitions

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

27

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
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”

28

Category	Description	Subtypes	Name	Definition		
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"	-	-

29

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, atmospheric 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

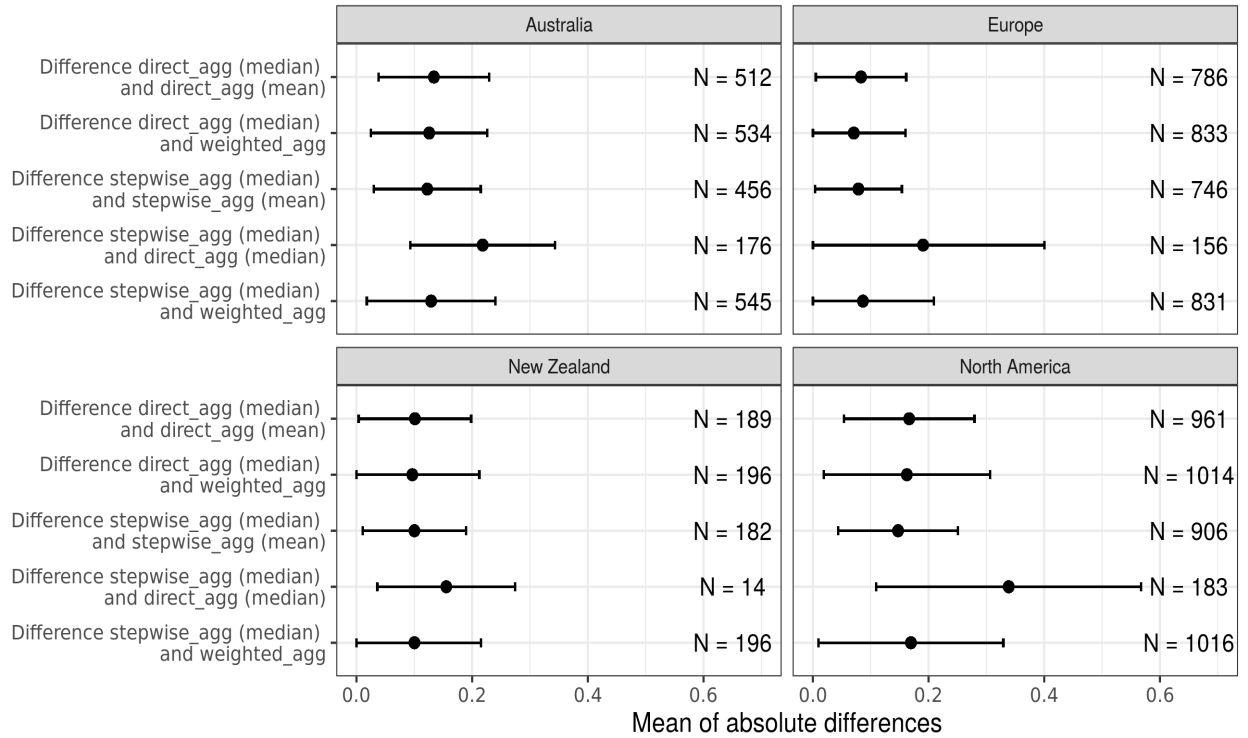**Comparison of the trait aggregation methods with each other**

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.

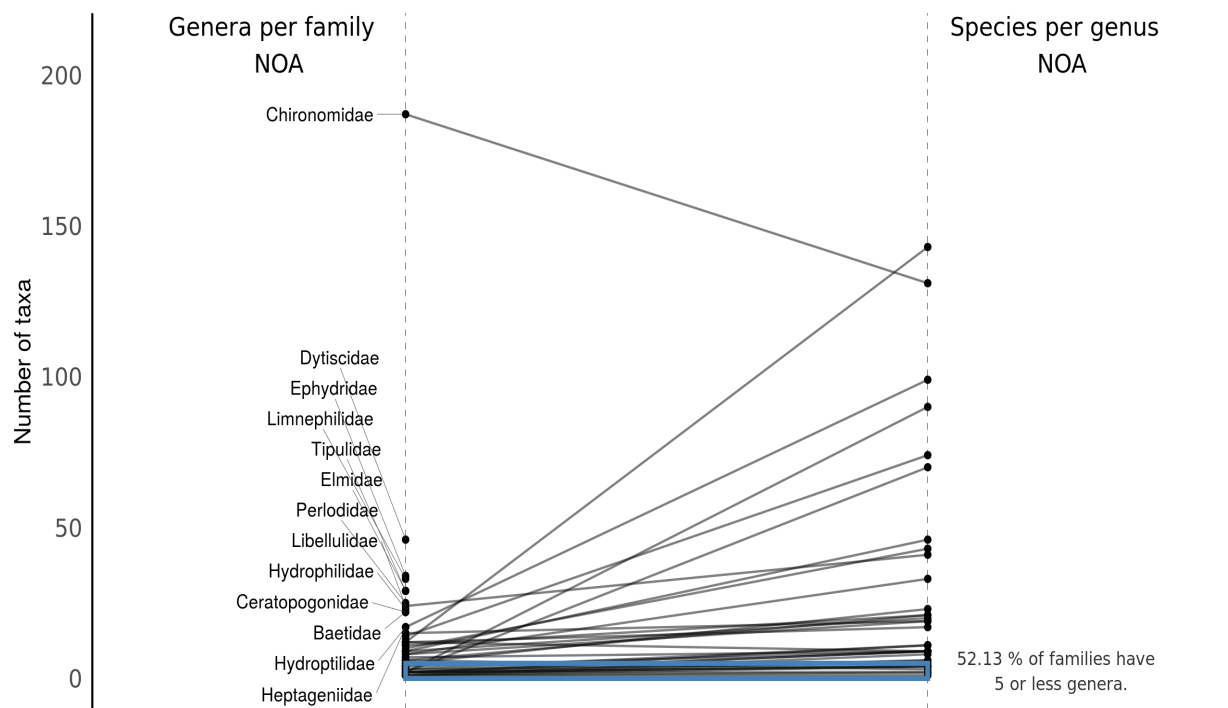**Taxonomic hierarchy in the trait datasets used for comparisons with assigned traits at family-level**

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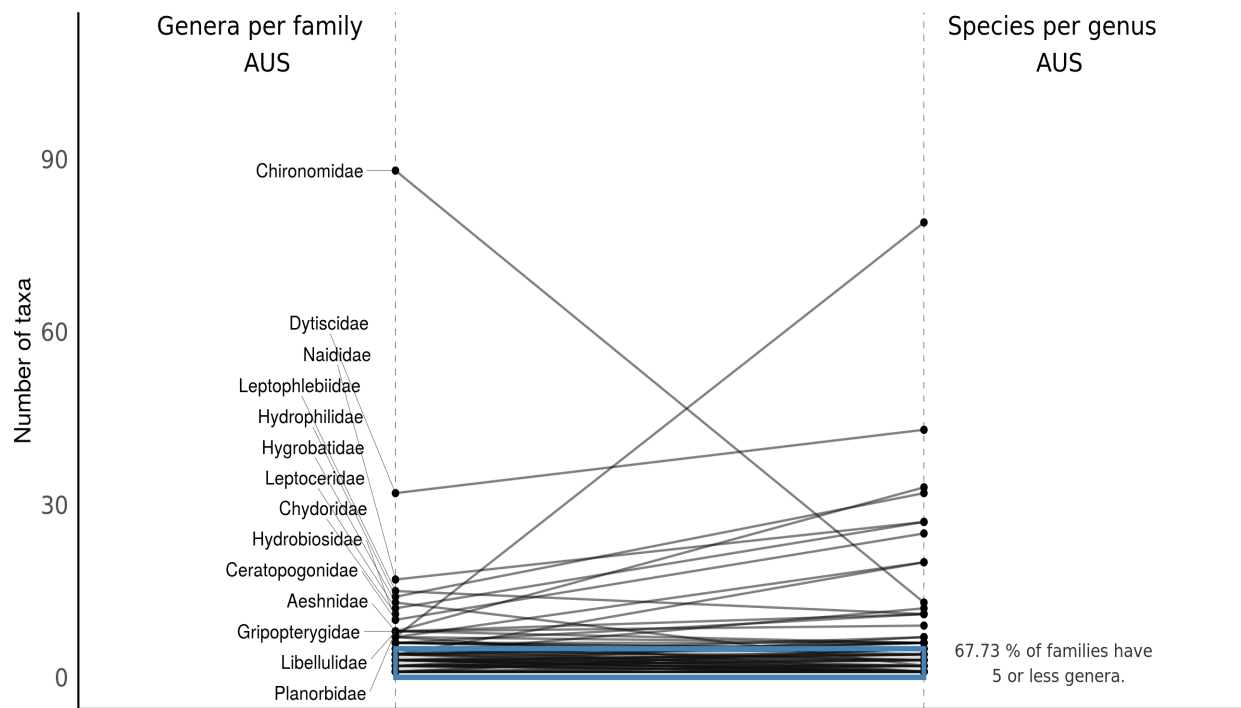
Figure S3: Number of genera per family and species per genus for the Australian trait dataset. For better visual display only families with more than 7 genera are displayed.

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