

DRAFT: Harmonized macroinvertebrate trait database, Aggregation of traits, Trait definitions

1 Introduction

Understanding distributions of invertebrate communities and predicting effects of potential natural or anthropogenic stressors is a long time goal of freshwater research. Since the 1970s ecologists have started using organismal characteristics or traits to group aquatic insect species to understand their diversity. The first to do so was Cummins by assigning feeding mechanisms to aquatic insects (Cummins 1973). Since then many studies have been carried out trying to use invertebrate traits to understand ecosystem processes, as tool for biomonitoring, and the effects of multiple stressors on freshwater communities (Statzner and Bêche 2010; Menezes, Baird, and Soares 2010). As a result, comprehensive databases on freshwater invertebrate traits have been compiled over the years (see table 1 for examples).

Traits are defined as measurable characteristics and ecological preferences at the individual organism. They reflect an organisms adaptation to its habitat and provide a mechanistic link between a species and its environment. Trait based approaches in ecology have its theoretical foundation in the habitat template theory, that predicts where environmental conditions are similar trait composition should also converge, even across biogeographic boundaries (Southwood 1977). Several studies suggested that trait variability is lesser on larger geographical scales than taxonomic variability (Bonada, DoléDec, and Statzner 2007). Therefore, trait based approaches can be a suitable tool for comparing the effect of various environmental stressors to invertebrate communities on large scales or across regions. Consequently, studies have been carried that e.g. examine the relationship between climate change and freshwater community assemblages using species traits (Bhowmik and Schäfer 2015; Brown et al. 2018). Many of such studies combine information from several trait databases. For example, Brown et al. used invertebrate trait databases from Europe, North America and New Zealand to investigate the effect of decreasing glacier cover to river ecosystems (Brown et al. 2018). As mentioned earlier, invertebrate trait data have become increasingly available over the last decades, also for different regions.

However, researchers face various problems when they need to synthesize information from multiple invertebrate trait databases. I) The use of inconsistent terminology across studies (see Schmera et al. 2015 for a comprehensive discussion). For example, some studies used the term *trait* to describe a general organismal property like "generations per year" (Statzner, Hoppenhaus, et al. 1997, Usseglio-Polatera et al. 2000), while in other studies this term was related to categories like "bi/multivoltine" (Haybach et al. 2004, Vieira et al. 2006). Here, we follow the proposal of Schmera et al. (2015) and use the term *trait* for a morphological, physiological, or phenomenological feature measurable at the individual organism (e.g. tegument, gills, etc.) and the term *grouping feature* to describe a general property of related organismal traits (e.g. respiration). The *membership state*

gives the measurement scale (normally nominal or ratio scale) the trait is described on. II) Invertebrate trait databases from different regions are not standardized. Often, the same grouping features are categorized using different traits. This is complicated by the use of different measurement scales. For example, trait databases from North America use traditionally a binary coding (i.e. trait is expressed or not), whereas most other trait databases (e.g. Tachet, freshwaterecology, and New Zealand) use fuzzy coding (i.e. trait is expressed to a certain extent by the organism). Hence, transformation of nominal scale to ratio scale or vice versa is required. Furthermore, definitions and terminology applied to traits differ as well across databases. Recently, Schneider et al. advocated standard terminologies and ontologies for trait definitions after finding that trait databases for different organism groups (*inter alia* invertebrates) are rarely standardized and have low potential for data synthesis (Schneider et al. 2019). III) Taxonomical resolution between databases differs, further complicating the synthesis of trait data from multiple regions. Some trait databases have recorded information on mixed levels of taxonomical resolution like the North American trait database or the Australian trait database. In contrast, trait information in the freshwaterecology database is entirely recorded on species-level. Using trait information on varying taxonomical levels is only possible when traits are aggregated to the lowest taxonomical level that is shared by all used databases. However, IV) so far studies comparing different ways of aggregating traits are lacking.

Given the problems mentioned above extensive data processing is required before researchers can use multiple invertebrate trait databases for their work. In this paper we examine difficulties that ecologists face when synthesizing trait information from multiple databases. We explore the effect of different decisions researches have to make when working with invertebrate trait data from several sources, involving trait harmonization, handling different codings, normalization, and aggregation of traits. Therefore, we harmonized six grouping features of different macroinvertebrate trait databases from four regions and aggregated the trait information to family-level. We discuss the harmonization and show the effect of different ways of aggregating traits. We also present an overview of differences in trait definitions among databases. Finally, our paper compares the references for the trait information that were specified in the trait databases we used.

2 Methods

2.1 Description of harmonized trait databases

The harmonized databases are using the available information on aquatic invertebrate traits for the regions Europe, North America, Australia, and New Zealand. Due to the different number and identity of grouping features in each database, the following six grouping features were chosen for this study: locomotion, feeding mode, respiration, voltnism, size, and body form. Trait information for Australia and New Zealand were retrieved from a single database, respectively. For Europe we gathered trait information from the freshwaterecology trait database (<https://www.freshwaterecology.info/>) and complemented where possible missing information with the Tachet trait database (Usseglio-Polatera et al. 2000). North American invertebrate traits were retrieved from Laura Twardochleb and complemented where possible by trait information from Vieira et al (Vieira et al. 2006). From now on, if we use the term European or North American trait database we refer to the combined databases. We used all available information on invertebrates in the databases but restricted our analysis to those taxa that have complete trait profiles for the investigated grouping features. We are aware that imputation methods exist, which infer missing information for traits by interpolating from related traits (Penone et al. 2014). However, by

using only complete data we were able to evaluate the taxonomical coverage within the databases. We consider this a helpful information for researchers who strive to fill data gaps. Table 1 gives an overview of the used databases.

Table 1: Overview of trait databases.

Region	Coding of trait states	Reference
Europe	Largely fuzzy	Schmidt-Kloiber and Hering 2015
Central Europe	Fuzzy coded	Usseglio-Polatera et al. 2000
North America	Largely binary	Vieira et al. 2006
North America	Largely binary	cite Laura Twardochleb
Australia	Binary & fuzzy coding	Kefford et al. 2019
New Zealand	Fuzzy coded	

2.2 Normalization & data conversion

Establishing a harmonized database required traits on the same measurement scale to enable comparability. However, in the used databases traits varied within and among databases in terms of measurement scale. Traits were either fuzzy coded, binary coded or to a small extent coded as continuous variable (e.g. size). Fuzzy codes capture the variation (i.e. temporal or spatial) in a trait and are represented by affinity scores which express the preference of a taxon to a certain trait. Usually, these scores range from 0 representing no affinity to an arbitrary maximum value which represents high affinity. This type of coding uses the ratio scale and not the ordinal scale. In fact, the intention of the developers of this coding system was to convert affinity scores to percentages per trait, e.g. for a grouping feature with four traits affinity scores of 5, 6, 2, 2 are equal to 1/3, 2/5, 2/15, 2/15 or 33.3 %, 40 %, 13.3 %, 13.3 % (Chevenet, Dolédec, and Chessel 1994). We used fuzzy coded traits for establishing our harmonized databases where possible, otherwise binary traits. Categorical and continuous traits across all used databases were converted into binary traits. Implicitly, we assumed for binary variables that a value of 1 for a particular trait corresponded to the maximum affinity of a taxon for that particular trait.

Two databases needed further data processing. Firstly, the Australian database which is a collection of seven trait databases. Thus, several grouping features occur multiple times but with traits that have different types of codings and ranges. For example, body size occurs as continuous variable, with traits that are fuzzy coded, and with traits that are binary coded. The same traits originating from different sub-databases in the Australian trait database were allocated. To enable allocation of fuzzy coded and binary traits we applied a range normalization by dividing each fuzzy coded trait by its potential maximum value. Consequently, the values of each trait were converted to a range from 0 to 1.

Secondly, the North American trait database which only contained traits on the nominal scale. We first converted nominal traits into binary traits. As a second approach we converted nominal traits for taxa on species and genus-level into "pseudo" fuzzy codes by calculating the percentage of occurrence for each trait per genera and grouping feature. Both approaches were compared regarding the trait values they yielded.

The Australian, Tachet, and the North American trait databases contained duplicate taxa entries - either on species, genus or family-level - which were amalgamated. The duplicate entries in the Australian trait database originated from the different sub-databases. As a result, many of the

duplicate entries complemented each other (e.g. for a given trait one duplicate contained a value, the other not). In those cases where several different values for a trait occurred, the average trait value was calculated. The North American database only contained 0 and 1 as entries. Hence, for duplicates the maximum value was taken. For the tachet trait database duplicate entries were allocated using the mean if they contained contrasting information on the same trait.

Finally, every trait within each grouping feature was standardized by the sum of all traits within this particular grouping feature to convert traits to percentages per trait.

2.3 Harmonization of traits and taxa

Harmonization of traits is the amalgamation of several similar traits into a single trait. It has to be undertaken when synthesizing a grouping feature from different sources which is not classified with the same traits. For example, in our study for the grouping feature locomotion the lowest number of traits that occur across all databases was in the New Zealand trait database ("Swimmer", "Burrower", "Crawler", "Sessil"). Hence, in all other databases locomotion traits have been allocated into these four traits (figures 1 and 2). Grouping features that differed in their traits among databases have been harmonized by condensing the traits in such a way that in the end the same grouping features in all databases consisted of the same traits. Thereby, traits were amalgamated based on ecological knowledge or expert judgment. Our approach of harmonizing the traits in the used databases is outlined in figures 1 and 2. Overall, the six grouping features were harmonized into 23 traits.

	EU	North America	AUS	NZ	Harmonized Traits		
Voltinism	Freshwater ecology Semivoltine Univoltine {Bivoltine Trivoltine Multivoltine flexible}	Tachet Semivoltine Monovoltine Polyvoltine	Laura Twardochleb Semivoltine Univoltine Bi/Multivoltine	Vieira et al. 2006 <1 Generation/y 1 Generation/y >1 Generation/y	Kefford et al. 2019 <1 Generation/y 1 Generation/y >1 Generation/y Up to 5 Generations/y Up to 10 Generations/y 1-2 Generations/y	Philips & Smith 2018 Semivoltine Univoltine Plurivoltine	Semivoltine Univoltine Bi/Multivoltine
Feeding Mode	{Shredder Miner Xylophagus} Gatherer {Active filterer Passive filterer} Grazer Predator Parasite Other	Shredder Deposit-feeder {Absorber Filter-feeder} Scraper Predator Parasite Piercer (plants or animals)	Shredder Collector-gatherer Collector-filterer Herbivore Predator Parasite Other	Shredder Collector-gatherer Collector-filterer {Scraper/grazer Piercer herbivore} Predator Parasite Other	(Shredder Shredder detritivore) {Detritivore Collector Collector gatherer Gatherer} {Filters Collector filterer} {Grazer Scraper Deposit grazers} {Predator Parasite Piercer}	Shredder Deposit-feeder Filter-feeder {Scraper Algal piercer} Predator * Collector, scraper, collector, shredder	Shredder Gatherer Filterer Herbivore Predator Parasite
Locomotion	{Swimming/scating Swimming/diving} burrowing/boring sprawling/walking (semi)sessil Other	{Surface swimmer Full water swimmer} Burrower Crawler {Temporarily attached Permanently attached} Interstitial Flier	{Swimmer Planktonic Skater} Burrower {Crawler Sprawler Climber Clinger} Attached	{Swimmer Planktonic Skater} Burrower {Sprawler Climber Clinger} Attached/fixed Other	{Swimmer Skater} Burrower {Crawler Sprawler Climber Clinger} Attached/fixed	Swimmer Crawler Burrower Attached	Swimmer Burrower Crawler Sessil

Figure 1: Proposed harmonization scheme for the grouping features voltinism, feeding mode and locomotion. Shown are all traits for the used grouping features in the investigated trait databases and the harmonized traits in the end. Traits in curly brackets were harmonized to one trait. Traits highlighted in Grey were omitted.

* Trait parasite was not available in New Zealand trait database.

	Europe	North America	AUS	NZ	Harmonized Traits
Respiration	Freshwater ecology Tachet	Laura Twardochleb Vieira et al. 2006	Kefford et al. 2019	Philipps & Smith 2018	
Size	{<= 0.25 cm > 0.25 – 0.5 cm > 0.5 - 1 cm} > 1 – 2 cm {2 – 4 cm 4 – 8 cm > 8 cm}	Small (< 9 mm) Medium (9 – 16 mm) Large (> 16 mm)	Small (< 9 mm) Medium (9 – 16 mm) Large (> 16 mm)	{Max size < 5 mm Max size 5 – 10 mm Small (< 9 mm)} (Max size 10 – 20 mm) Medium (9 – 16 mm) {Max size 20 – 40 mm Max size > 40 mm Large (> 16 mm)}	{Size < = 5 mm Size > 5 – 10 mm Size > 10 – 20 mm Size > 20 – 40 mm Size > 40 mm}
Body Form	* Streamlined Flattened Cylindrical Spherical	* Streamlined Flattened Cylindrical Spherical	Streamlined/fusiform Dorsoventrally flattened Tubular Round (humped) Bluff (blocky)**	Streamlined Flattened Cylindrical Spherical	Streamlined Flattened Cylindrical Spherical

Figure 2: Proposed harmonization scheme for the grouping features respiration, size and body form.

* Body form information provided by Philippe Usseglio-Polatera.

** Bluff(blocky) taxa have been reclassified by Philippe Usseglio-Polatera using the traits streamlined, flattened, cylindrical and spherical.

Not only the categorization of grouping features into traits but also the definitions of individual traits varies between databases, complicating harmonization.

2.4 Aggregation of traits

Traits in the processed and harmonized trait databases were aggregated to family-level using three approaches. I) taxa on species-level and genus-level were stepwise aggregated to the family-level by initially allocating them at the genus-level using the median. Then all traits were aggregated to family-level by using the mode. In cases where it was not possible to take the mode, e.g. only distinct values, multiple duplicates, or multiple duplicates and distinct values occurred the mean was taken. Hereafter, we abbreviate this aggregation type as *stepwise_agg*. II) we directly aggregated taxa to family level using the median. We denote this aggregation as *direct_agg*. III) taxa were aggregated using a weighted approach, denoted as *weighted_agg*. The weights were determined as the ratio of how many taxa on species or genus-level were initially present per genera compared to how many taxa on species or genus-level were present after selecting only taxa with complete trait profiles. After determining the weights, taxa on species and genus-level were aggregated to

genus-level by multiplying their trait values by their respective weights. Then the weighted trait values were summed up per family for each trait.

The resulting aggregated trait values were compared to trait values assigned at family-level by experts. Trait assignments on family-level existed for the Australian database and the North American database, but only for a limited subset of grouping features and taxa. For the Australian database we could compare aggregated trait values with assigned trait values resolved at family-level for the grouping features feeding mode and size by using data from Chessman et al. For the North American database we could compare aggregated trait values with assigned trait values on family-level for the grouping features feeding mode, respiration, size, voltinism and locomotion. The trait information was obtained from the trait database by Pyne et al, which contains trait information for aquatic insects resolved on genus and family-level. Trait information on family-level was available for 94 families of which 61 (approximately 65 %) were present in the aggregated North American trait database. Trait information in the Pyne database was on the nominal scale and was converted to binary traits prior to the comparison with aggregated traits values.

3 Results

3.1 Taxonomical coverage

Regarding the taxonomical coverage the New Zealand database has, as expected, the smallest taxon pool. In total 492 taxa are covered by this database. Thereof, 404 taxa resolved on species-level, 47 taxa on genus-level and 27 taxa on family-level. The remaining entries are on a lower taxonomical resolution. 73 % of taxa in the New Zealand database belong to the group of aquatic insects. The largest taxon pool is spanned by the European trait database, with 4224 taxa of 76 different orders. 48 % of the taxa in this database belong to the group of aquatic insects. The European database is mostly on the highest taxonomical resolution possible, with 3953 taxa on species-level (approximately 93,6 %). 253 entries are resolved on genus-level and 18 entries on family-level. The Australian database has 1404 taxa of 64 orders. 52 % of the taxa covered are aquatic insects. 564 taxa are resolved on species-level, 578 on genus-level and 260 on family-level. The North American trait database contained trait information for 3542 taxa of 42 different orders, although 63 % of the taxa in the database belong to the aquatic insects. 2142 entries are on species-level, 1074 on genus-level and 50 on family-level.

3.2 Completeness of trait information

The percentage of entries with available information for the individual grouping features in the Australian, European and North American trait databases varied between 5 % and 99 %. By contrast, the New Zealand trait database contained complete trait information for 99 % to 100 % of their entries for the individual grouping features (2). The greatest data gap was for the grouping feature body form where information was only present for 7 % of entries in the Australian and European database, and 26 % of entries in the North American database. Selecting only taxa resolved at least at family-level and with complete trait profiles for the six grouping features lead to the omission of many taxa in all databases except for the New Zealand database. Out of the 21 orders that the New Zealand database covers, taxa of 20 orders remained after the selection process. For most orders, all families initially included were also included after selecting taxa with complete trait profiles (table 3). By contrast, in the Australian database, only taxa from five orders

(Ephemeroptera, Megaloptera, Odonata, Plecoptera, and Trichoptera) remained. Within these five orders, there was none where all families included in the database contained complete trait profiles. In the North American database, taxa of 18 orders had complete trait profiles, in the European database 9 orders.

Table 2: Percentage of entries that have information for the individual grouping features per database.

Database	Grouping feature	Trait covered [%]
Australia	Body form	5.00
Australia	Feeding mode	99.00
Australia	Locomotion	42.00
Australia	Respiration	70.00
Australia	Size	78.00
Australia	Voltinism	49.00
Europe	Body form	7.00
Europe	Feeding mode	65.00
Europe	Locomotion	33.00
Europe	Respiration	56.00
Europe	Size	11.00
Europe	Voltinism	24.00
New Zealand	Body form	100.00
New Zealand	Feeding mode	99.00
New Zealand	Locomotion	99.00
New Zealand	Respiration	100.00
New Zealand	Size	100.00
New Zealand	Voltinism	100.00
North America	Body form	26.00
North America	Feeding mode	61.00
North America	Locomotion	51.00
North America	Respiration	44.00
North America	Size	75.00
North America	Voltinism	47.00

Table 3: Proportion of families per order that remain after only taxa with complete trait profiles have been selected from the total number of distinctive families in the databases.

Order	Australia	New Zealand	Europe	North America
Amphipoda		100.00	37.50	10.00
Anthoathecata		100.00		
Architaenioglossa				100.00
Arhynchobdellida				100.00
Branchiopoda				100.00
Coleoptera		88.89	83.33	15.62
Cycloneritida				100.00
Decapoda		100.00		
Diptera		100.00	46.15	21.62
Ephemeroptera	50.00	100.00	66.67	73.91
Hemiptera		50.00	63.64	47.06
Hexapoda		100.00		
Lepidoptera		100.00		
Littorinimorpha				66.67
Mecoptera		100.00		
Megaloptera	50.00	100.00		50.00
Mollusca		100.00		
Mysida		100.00		
Nemertea		100.00		
Neuroptera		100.00		
Odonata	6.90	100.00	100.00	50.00
Oligochaeta		100.00		
Onychura				100.00
Plecoptera	25.00	100.00	100.00	88.89
Rhynchobdellida		100.00		50.00
Spinicaudata				100.00
Tanaidacea		100.00		
Trichoptera	52.17	100.00	95.45	71.43
Venerida			100.00	33.33

3.3 Taxonomical overlap between trait databases

3.4 Deviance in trait values

The *stepwise_agg* and *direct_agg* approaches yielded for the majority of taxa the same trait values (table 4) across all grouping features. The amount of cases for which the two aggregation methods resulted in different trait values varied between approximately 1 % and 3.5 %. One case here is a taxa resolved on family-level with a specific trait.

There was no family that had diverging trait values produced by the aggregation methods in all databases, although for Leptophlebiidae (order Ephemeroptera) the aggregation methods produced different trait values in the European, North American and Australian trait databases.

Table 4: Percentage of cases where the *stepwise_agg* and *direct_agg* methods resulted in different trait values on family-level.

Database	Deviating cases [%]	Number of cases
Australia	3.52	483
Europe	2.17	2024
New Zealand	1.03	2134
North America	2.66	2139

3.5 Comparison aggregated trait values and trait values assigned at family-level

3.5.1 Australia

3.5.2 North America

Overall, approximately 21 % of trait values evaluated at family-level by Pyne et al. were different to the trait values obtained by *direct_agg* in the North American trait database for families both databases share. Compared to the *stepwise_agg*, 22 % of trait values from Pyne et al. were different. Trait values for body *size medium* were most often deviating between the *direct_agg* and traits assigned at family-level (45.9 % of all families, total number of taxa on family-level: 61), followed by voltinism *univoltine* (36 %), and feeding mode *gatherer* (32.8 %). Since both aggregation methods mostly resulted in similar trait values, differences in trait values between *stepwise_agg* and traits assigned at family-level were similar to the differences for *direct_agg*. The most deviating trait was again body *size medium* (47.5 % of all families), followed by voltinism *univoltine* (37.7 %), and *size small* (29.5 %). Among all compared traits there were always differences in aggregated trait values and assigned values at family-level for at least families from one order. Tables S5 and S6 give an overview over how many families were differently evaluated by the aggregation methods compared to Pyne's database for all investigated traits. Regarding deviating cases per order, families from Ephemeroptera had most often different assigned trait values compared to the trait values obtained by the *direct_agg* (24.37 % of cases, table S7). For the *stepwise_agg* families from Trichoptera (24.51 % of cases, table S8) were deviating from the assigned trait values. For both aggregation methods families from Odonata differed the least often to the assigned trait values (16.18 % respectively).

Trait values ranged between 0 and 1. Hence, the maximum deviation possible is 1 (100 %). Regarding the cases with deviations in trait values between *direct_agg* and at family-level assigned traits, 26 % of the cases showed a deviation of 100 %, i.e. the value obtained by *direct_agg* was 0 or 1 while the value on family-level assigned was 1 or 0 (Figures S3 to S12.). Maximum deviations occurred within all orders except for Dipterans. Most often the trait *size medium* (8 times, 14 % of all maximum deviations) deviated by 100 %, followed by the trait *univoltine* (7 times, 12.3 % of all maximum deviations) and trait *size small* (7 times).

Comparison of the cases with deviations in trait values between *stepwise_agg* and at family-level assigned traits shows a similar picture. 25.8 % of the cases showed a deviation of 100 %. Maximum deviations occurred as well within all orders except for Dipterans. Most often the trait *size medium* (8 times, 13.3 % of all maximum deviations) deviated by 100 %, followed by the trait *size large* (7 times, 11.7 % of all maximum deviations), trait *size small* and trait *univoltine* (both 6 times, 10 % of all maximum deviations).

3.6 Discrepancies in trait definitions

3.6.1 Feeding Mode

The determination of the feeding mode of an organism is derived from the classification into functional-feeding guilds and is based on what organisms consume, their morphology of mouth-parts and their feeding behavior. The freshwaterecology database, the Tachet database and the North American trait database state trait explanations describing the food sources for each trait related to the feeding mode. In the North American the method of dietary intake is also roughly described (e.g. piercers pierce prey tissues and suck fluids). The other databases give no deeper description of how exactly traits of the grouping feature feeding mode are defined. As mentioned earlier, each databases uses different traits to describe the feeding mode (1). There are also gradual differences within the definitions between the traits that occur in all databases belonging to the grouping feature feeding mode. For example, in the North American trait database shredders are defined as insects that shred on decomposing vascular plant tissue. However, freshwaterecology and tachet database include both decomposing and living plant material as food source for shredders and also additionally coarse particulate organic material (CPOM). The North American trait database summarizes taxa that scrape, shred or pierce living aquatic plants with another trait termed herbivore.

The freshwaterecology database states as reference for their classification of the grouping feature feeding mode Moog et al. (Moog and Hartmann 1995), who in turn based their classification on Cummins (Cummins 1973), Cummins and Klug (Cummins and Klug 1979), and Merrit and Cummins. Also in the North American database, the classification of nutritional forms is derived from the publication of Cummins (Cummins 1973). For the New Zealand trait databases it is not explicitly stated how they derived their classification of the feeding mode.

References

- Bhowmik, Avit Kumar and Ralf B. Schäfer (June 16, 2015). “Large Scale Relationship between Aquatic Insect Traits and Climate”. In: *PLoS ONE* 10.6. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0130025. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4469582/> (visited on 06/21/2019).
- Bonada, Núria, Sylvain Dolédec, and Bernhard Statzner (Aug. 2007). “Taxonomic and biological trait differences of stream macroinvertebrate communities between mediterranean and temperate regions: implications for future climatic scenarios”. In: *Global Change Biology* 13.8, pp. 1658–1671. ISSN: 1354-1013, 1365-2486. DOI: 10.1111/j.1365-2486.2007.01375.x. URL: <http://doi.wiley.com/10.1111/j.1365-2486.2007.01375.x> (visited on 06/25/2018).
- Brown, Lee E. et al. (Feb. 2018). “Functional diversity and community assembly of river invertebrates show globally consistent responses to decreasing glacier cover”. In: *Nature Ecology & Evolution* 2.2, pp. 325–333. ISSN: 2397-334X. DOI: 10.1038/s41559-017-0426-x. URL: <http://www.nature.com/articles/s41559-017-0426-x> (visited on 02/04/2019).
- Chevenet, Francois, Sylvain Dolédec, and Daniel Chessel (1994). “A fuzzy coding approach for the analysis of long-term ecological data”. In: *Freshwater Biology* 31.3, pp. 295–309. ISSN: 0046-5070. DOI: 10.1111/j.1365-2427.1994.tb01742.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2427.1994.tb01742.x> (visited on 04/18/2018).
- Cummins, Kenneth W. (Jan. 1973). “Trophic Relations of Aquatic Insects”. In: *Annual Review of Entomology* 18.1, pp. 183–206. ISSN: 0066-4170, 1545-4487. DOI: 10.1146/annurev.en.18.010173.001151. URL: <http://www.annualreviews.org/doi/10.1146/annurev.en.18.010173.001151> (visited on 03/16/2020).
- Cummins, Kenneth W. and Michael J. Klug (1979). “Feeding Ecology of Stream Invertebrates”. In: *Annual Review of Ecology and Systematics* 10.1. eprint: <https://doi.org/10.1146/annurev.es.10.110179.001051>, pp. 147–172. DOI: 10.1146/annurev.es.10.110179.001051. URL: <https://doi.org/10.1146/annurev.es.10.110179.001051> (visited on 03/18/2020).
- Haybach, Arne et al. (Dec. 1, 2004). “Use of biological traits for interpreting functional relationships in large rivers”. In: *Limnologica. New methods for assessing freshwaters in Germany* 34.4, pp. 451–459. ISSN: 0075-9511. DOI: 10.1016/S0075-9511(04)80012-4. URL: <http://www.sciencedirect.com/science/article/pii/S0075951104800124> (visited on 01/07/2020).
- Kefford, Ben J. et al. (2019). “An integrated database of stream macroinvertebrate traits for Australia: Concept and application”. In: *Ecological Indicators, Manuscript submitted*.
- Menezes, Salomé, Donald J. Baird, and Amadeu M. V. M. Soares (2010). “Beyond taxonomy: a review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring”. In: *Journal of Applied Ecology* 47.4, pp. 711–719. ISSN: 1365-2664. DOI: 10.1111/j.1365-2664.2010.01819.x. URL: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2664.2010.01819.x> (visited on 11/28/2019).
- Moog, Otto and Anne Hartmann (1995). “A Comprehensive Species Inventory of Austrian Aquatic Organisms with Ecological Notes”. In: p. 86.
- Penone, Caterina et al. (Sept. 2014). “Imputation of missing data in life-history trait datasets: which approach performs the best?” In: *Methods in Ecology and Evolution* 5.9. Ed. by Robert Freckleton, pp. 961–970. ISSN: 2041210X. DOI: 10.1111/2041-210X.12232. URL: <http://doi.wiley.com/10.1111/2041-210X.12232> (visited on 05/11/2018).
- Schmera, Dénes et al. (Sept. 2015). “A proposed unified terminology of species traits in stream ecology”. In: *Freshwater Science* 34.3, pp. 823–830. ISSN: 2161-9549, 2161-9565. DOI: 10.1086/

681623. URL: <https://www.journals.uchicago.edu/doi/10.1086/681623> (visited on 11/06/2019).
- Schmidt-Kloiber, Astrid and Daniel Hering (June 1, 2015). “www.freshwaterecology.info – An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences”. In: *Ecological Indicators* 53, pp. 271–282. ISSN: 1470-160X. DOI: 10.1016/j.ecolind.2015.02.007. URL: <http://www.sciencedirect.com/science/article/pii/S1470160X15000758> (visited on 06/21/2019).
- Schneider, Florian D. et al. (Sept. 16, 2019). “Towards an ecological trait-data standard”. In: *Methods in Ecology and Evolution*. Ed. by David Orme, pp. 2041–210X.13288. ISSN: 2041-210X, 2041-210X. DOI: 10.1111/2041-210X.13288. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13288> (visited on 10/10/2019).
- Southwood, T. R. E. (1977). “Habitat, the Templet for Ecological Strategies?” In: *Journal of Animal Ecology* 46.2, pp. 337–365. ISSN: 0021-8790. DOI: 10.2307/3817. URL: <https://www.jstor.org/stable/3817> (visited on 06/18/2019).
- Statzner, Bernhard and Leah A. Bêche (2010). “Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems?” In: *Freshwater Biology* 55 (s1), pp. 80–119. ISSN: 1365-2427. DOI: 10.1111/j.1365-2427.2009.02369.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2427.2009.02369.x> (visited on 11/28/2019).
- Statzner, Bernhard, Kerstin Hoppenhaus, et al. (1997). “Reproductive traits, habitat use and templet theory: a synthesis of world-wide data on aquatic insects”. In: *Freshwater Biology* 38.1, pp. 109–135. ISSN: 1365-2427. DOI: 10.1046/j.1365-2427.1997.00195.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2427.1997.00195.x> (visited on 01/07/2020).
- Usseglio-Polatera, Philippe et al. (Apr. 1, 2000). “Biomonitoring through biological traits of benthic macroinvertebrates: how to use species trait databases?” In: *Hydrobiologia* 422.0, pp. 153–162. ISSN: 1573-5117. DOI: 10.1023/A:1017042921298. URL: <https://doi.org/10.1023/A:1017042921298> (visited on 06/21/2019).
- Usseglio-Polatera, Philippe et al. (2000). “Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits”. In: *Freshwater Biology* 43.2, pp. 175–205. ISSN: 1365-2427. DOI: 10.1046/j.1365-2427.2000.00535.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2427.2000.00535.x> (visited on 02/27/2019).
- Vieira, Nicole K M et al. (2006). “A Database of Lotic Invertebrate Traits for North America”. In: p. 19.

4 Supporting Information

Table 5: Percentage of families differently evaluated by *direct_agg* and Pyne et al. for all compared traits.

Trait	Families differently evaluated [%]
size_medium	45.90
volt_uni	36.07
feed_gatherer	32.79
size_small	31.15
size_large	22.95
locom_crawl	22.95
volt_bi_multi	22.95
resp_teg	22.95
resp_gil	21.31
feed_filter	16.39
feed_herbivore	16.39
volt_semi	14.75
locom_swim	13.11
feed_shredder	13.11
locom_burrow	9.84
resp_pls_spi	8.20
feed_predator	8.20

Table 6: Percentage of families differently evaluated by *stepwise_agg* and Pyne et al. for all compared traits.

Trait	Families differently evaluated [%]
size_medium	47.54
volt_uni	37.70
size_small	29.51
resp_gil	29.51
resp_teg	29.51
feed_gatherer	29.51
size_large	26.23
locom_crawl	24.59
volt_bi_multi	22.95
volt_semi	16.39
feed_filter	16.39
feed_herbivore	14.75
locom_swim	13.11
feed_shredder	13.11
locom_burrow	9.84
resp_pls_spi	9.84
feed_predator	9.84

Table 7: Percentage of families differently evaluated by *direct_agg* and Pyne et al. for all compared orders.

Order	Families differently evaluated direct_agg [%]
Ephemeroptera	24.37
Coleoptera	23.53
Megaloptera	23.53
Trichoptera	21.24
Diptera	20.59
Plecoptera	19.85
Hemiptera	16.47
Odonata	16.18

Table 8: Percentage of families differently evaluated by *stepwise_agg* and Pyne et al. for all compared orders.

Order	Families differently evaluated stepwise_agg [%]
Trichoptera	24.51
Ephemeroptera	23.95
Coleoptera	23.53
Megaloptera	23.53
Hemiptera	21.18
Diptera	20.59
Plecoptera	19.85
Odonata	16.18

Figure 3: Deviance of trait values obtained with *direct_agg* compared to at family level assigned traits by Pyne et al.

N denotes the number of families per order.

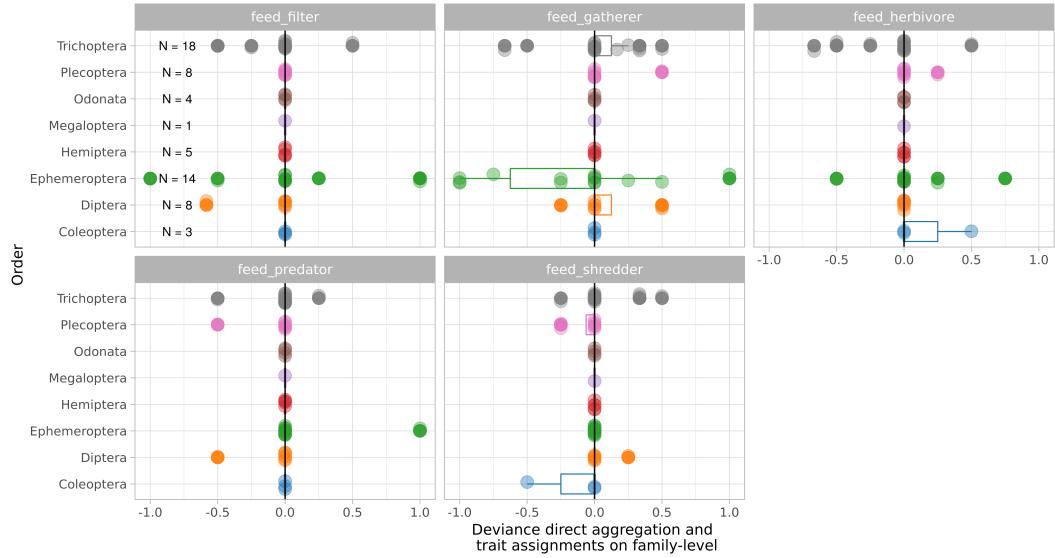


Figure 4: .

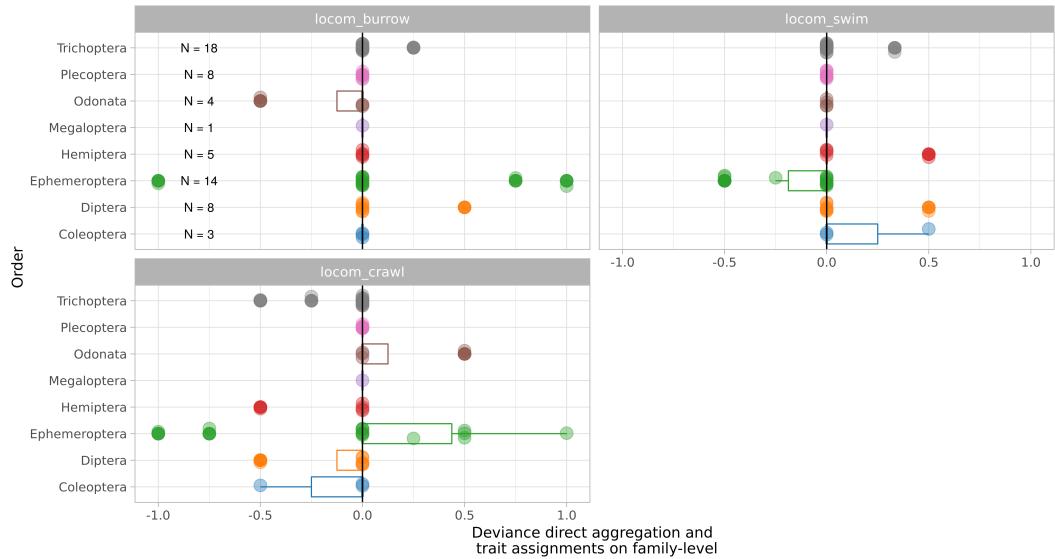


Figure 5: .

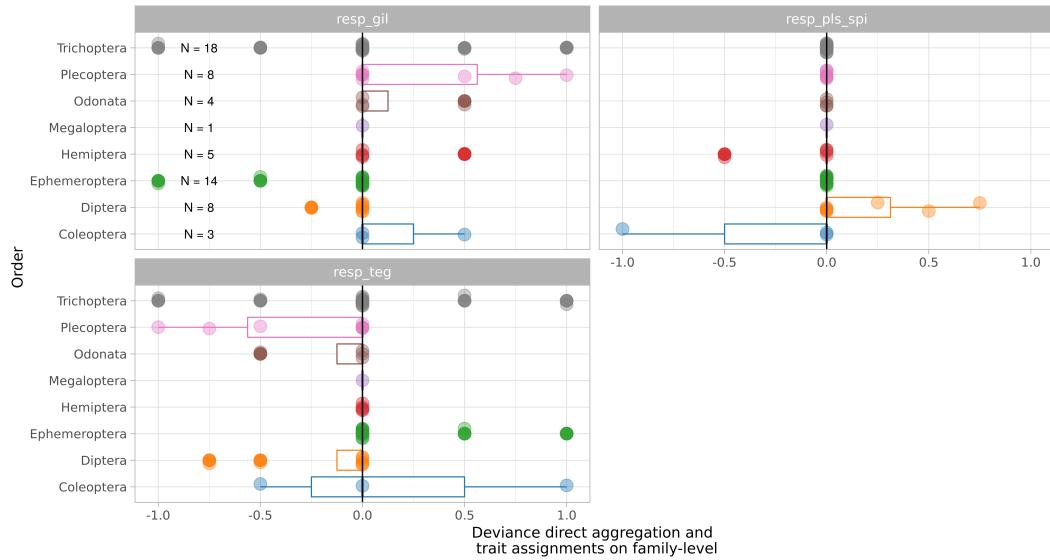


Figure 6: .

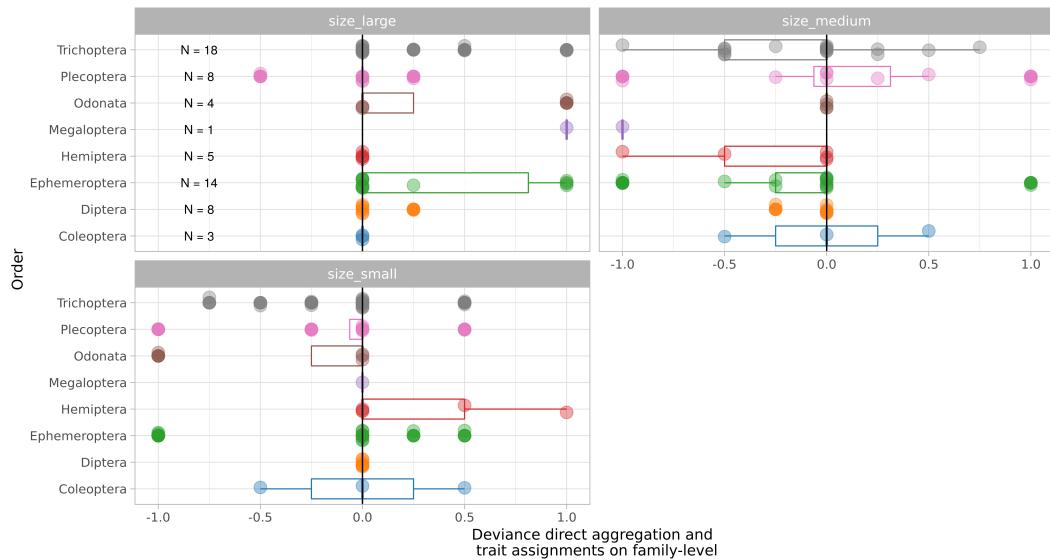


Figure 7: .

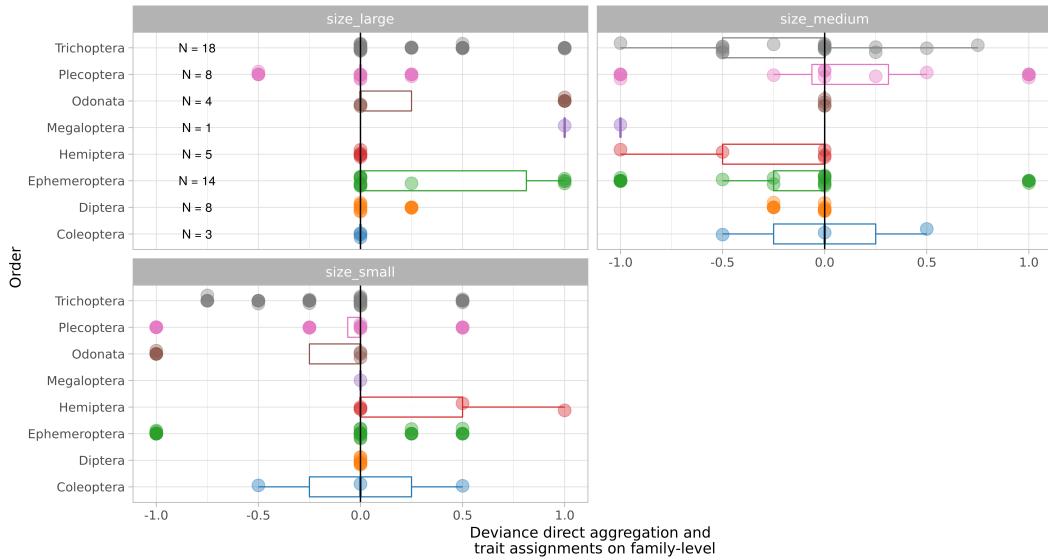


Figure 8: .

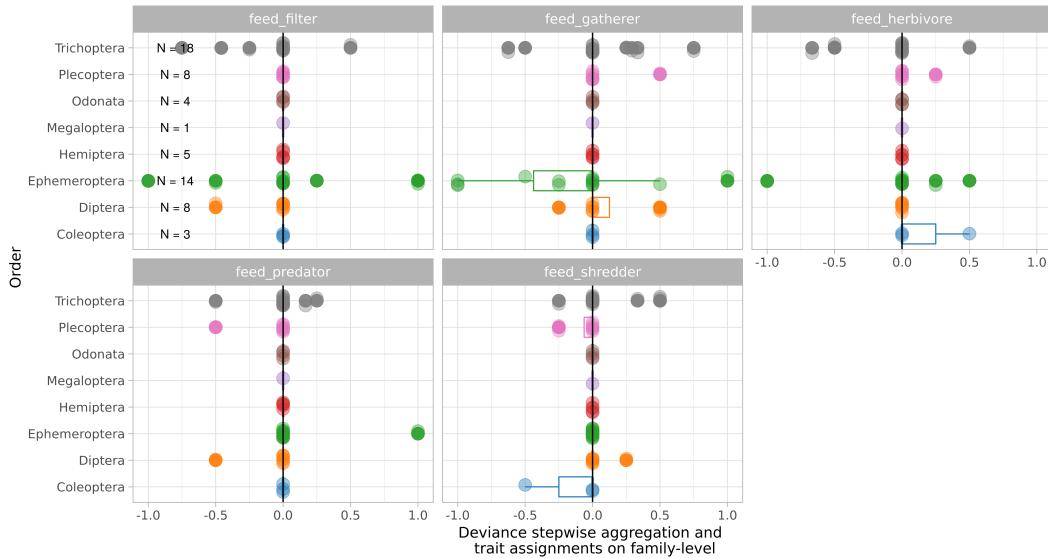


Figure 9: .

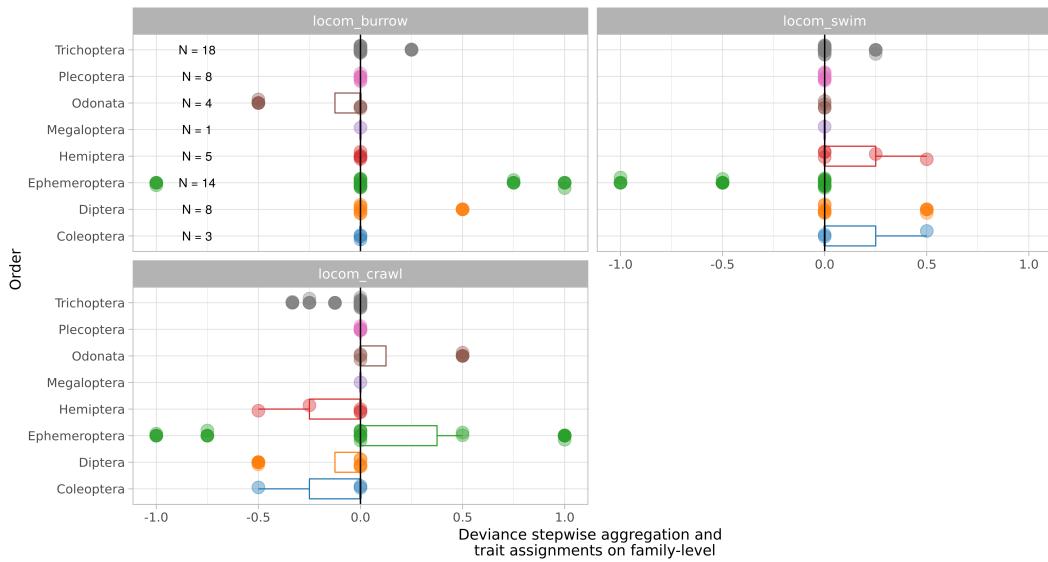


Figure 10: .

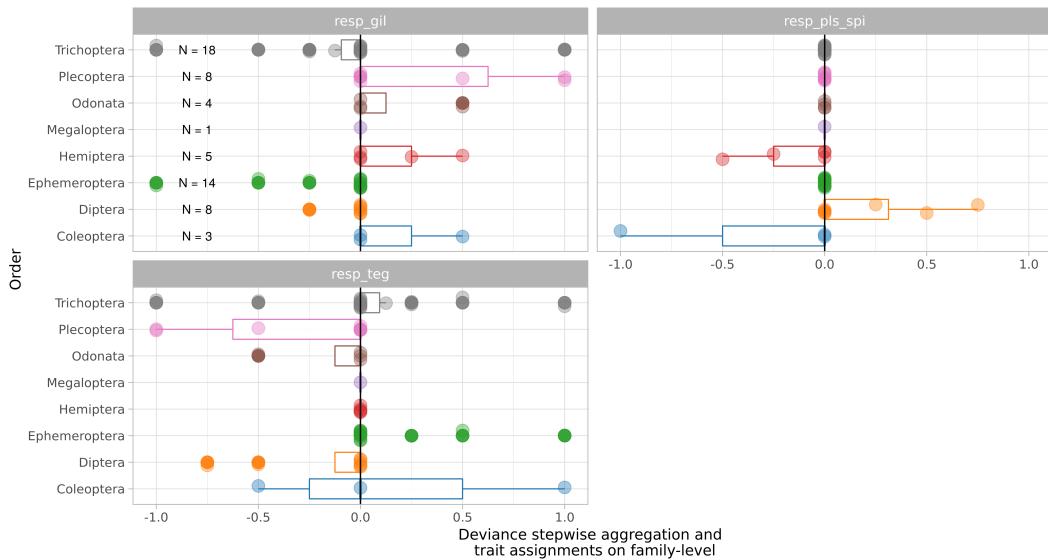


Figure 11: .

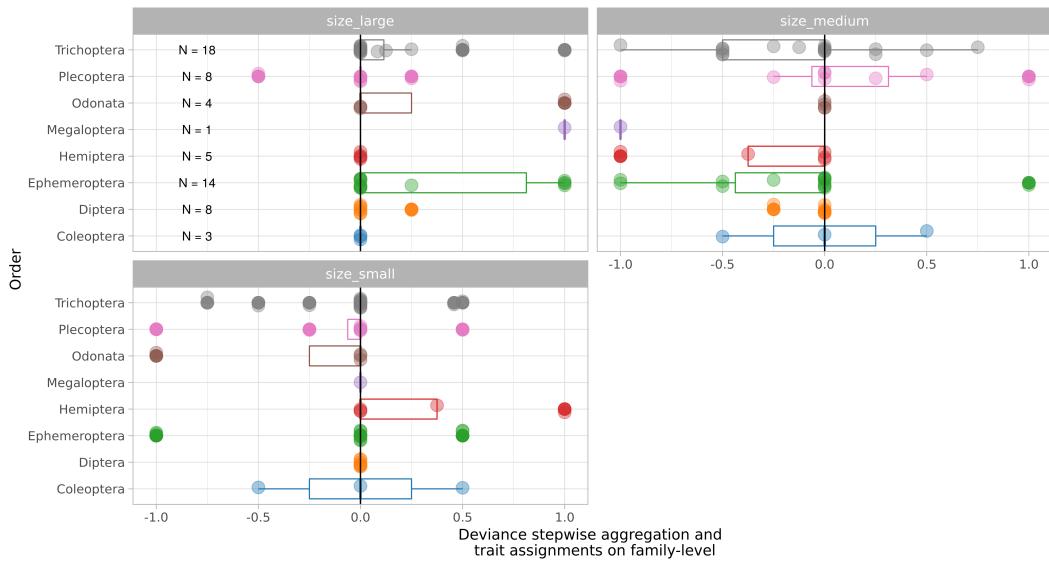


Figure 12: .

