OVERVIEW RESULTS: Harmonized invertebrate grouping feature databases

Results harmonized grouping feature databases

We used information on invertebrate traits for the regions Europe, North America, Australia, and New Zealand to establish harmonized grouping feature databases. Trait information for Europe was obtained by the Freshwaterecology database and complemented by Tachet when information was missing (e.g. for the grouping feature size). Trait information for North America was obtained from Laura Twardochleb and complemented by Vieira et al. 2006 (Vieira et al. 2006). Philippe Usseglio-Polatera provided information on body form for European and North American taxa.

We selected seven grouping features that we harmonized into 26 traits. The grouping features were chosen based on the availability of trait information across databases and so that they describe different parts of the biology of an organism: life history (Voltinism), morphology (Respiration, Body form, Size), ecology (Locomotion, Feeding mode) and reproduction (Oviposition). We used fuzzy coded traits for establishing our harmonized databases unless data quality prohibited and then we used binary traits, i.e categorical and continuous traits were converted into binary traits. Data processing prior harmonization also involved amalgamation of duplicate taxa on species, genus or family-level and conversion of trait affinities to percentages per trait. We omitted taxa with a lower taxonomic resolution than family-level. The following results relate to the harmonized grouping feature databases.

1 Taxonomic coverage of the trait databases

Regarding the taxonomical coverage, the New Zealand database has, as expected, the smallest taxon pool (Table 1). By contrast, the largest taxon pool is spanned by the European trait database with 4225 taxa followed by the North American trait database that contained trait information on 3542 taxa. The Australian database contains 1404 taxa. The European, New Zealand, and North American databases have most taxa on the highest taxonomical resolution while the Australian database has a similar number of taxa on species and genus-level.

Table 1: Number of taxa per harmonized database and per taxonomic level. Numbers in parenthesis show relative frequencies in percentage.

Database	Nr. of taxa	Species	Genus	Family	Aquatic insects
Europe	4225	3951 (93.51)	257 (6.08)	17 (0.4)	3654 (86.49)
North America	3542	2418 (68.27)	1074 (30.32)	50 (1.41)	3144 (88.76)
Australia	1402	564 (40.23)	578 (41.23)	260 (18.54)	1015 (72.4)
New Zealand	478	404 (84.52)	47 (9.83)	27 (5.65)	443 (92.68)

2 Completeness of trait information

The amount of entries with available information for the individual grouping features varied strongly for the European, North American and Australian databases (Table 2). By contrast, the New Zealand database contained complete trait information for most of the investigated grouping features (between 94 % and 100 %).

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

Database	Body form	Oviposition	Voltinism	Locomotion	Size	Respiration	Feeding mode
EU	7	16	24	33	11	56	65
NOA	26	12	47	51	75	44	61
AUS	5	48	51	42	78	70	99
NZ	100	94	100	99	100	100	99

3   Comparison of trait aggregation methods

Only taxa with complete trait profiles for the seven grouping features were used for this analysis. This reduced the number of taxa largely for European, North American and Australian databases (Europe = 294, North America = 106, Australia = 26, New Zealand = 440). Trait affinities ranged between 0 and 1. Hence, the maximum deviation possible is 1 or -1 (corresponds to 100 %).

   Traits of the harmonized grouping feature databases were aggregated to family-level using three approaches. I) taxa on species-level and genus-level were aggregated stepwise to the family-level by first aggregating 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 occurred the median was taken. With this method, each genus counts equally. Hereafter, we abbreviate this aggregation type as *stepwise_agg*. II) we directly aggregated taxa to family-level giving equal weight to every species. We denote this aggregation as *direct_agg*. For the *direct_agg* we tested aggregating with the median and the mean. We add *median* or *mean* to *direct_agg* to indicate when we use which method. III) taxa were aggregated using a weighted mean approach, denoted as *weighted_agg*. This method weights the genera according to the number of their species initially present in the databases. The weights were determined as the ratio of the number of species per genera that were initially present compared to the number of species per genera present after selecting only taxa with complete trait profiles.

3.1   Deviance in trait affinities between *stepwise_agg* and *direct_agg (median)*

The *stepwise_agg* and *direct_agg (median)* approaches yielded for the majority of taxa the same affinity (Table 3). The amount of cases where the two aggregation methods resulted in different affinities varied between approximately 0.5 % and 3 % per database. One case here is a taxa resolved on family-level with a specific trait.

Table 3: Percentage of cases for which the *stepwise_agg* and *direct_agg (median)* methods resulted in different trait affinities on family-level. A case is a factor combination of family and trait.

Database	Deviating cases [%]	Number of cases
Europe	1.77	2200
North America	0.54	1300
Australia	3.00	400
New Zealand	1.16	1900

   Deviances in trait affinities occurred for all grouping features except body form. Deviances ranged between -0.67 (family *Baetidae*, trait herbivore) and 0.6 (family *Libellulidae*, trait size medium). Standard deviations of the deviances varied between databases, i.e. only a small standard deviation in the Australian database (0.09) and relatively large in the North American database (0.43). The mean of the absolute deviances showed that deviances in the North American database, albeit seldom, were high (0.37) and close to zero in the Australian database (0.07) (Table 4, Figure SI7). Deviances in aggregated trait affinities occurred for all investigated grouping features.

   Repeating this analysis with a higher number of taxa using only the three grouping features feeding mode, respiration, and locomotion (Europe = 898, North America = 1033, Australia =

587, New Zealand = 470) resulted in a similar picture. The percentage of deviating cases between *direct_agg* and *stepwise_agg* ranged between 1.3 % (New Zealand) and 4 % (Europe).

As mentioned above, the highest deviance in trait affinity observed was for the family *Baetidae* for the trait herbivore. In this case *stepwise_agg* was zero and the *direct_agg (median)* 0.67. Manual examination showed that 7 taxa exist for this family in the European database, 5 of them at species-level and 2 at genus-level. None of the species belonged to the same genus. The affinities of the taxa were the following: (0.9, 0.7, 0, 0.67, 0, 0.75, 0.43). The steps of the *stepwise_agg* are: I) aggregation to genus-level using the median, which did not change the affinities. II) Subsequent aggregation to family-level using the mode, which resulted in the trait herbivore being assigned affinity 0. The *direct_agg* results in a more realistic affinity in this case (0.67).

In general, both methods seem to yield similar trait affinities. Furthermore, using the mode within the *stepwise_agg* seems to produce rather unrealistic affinities in some scenarios. Hence, we limit our analysis in the following sections to the *direct_agg* and *weighted_agg*.

Table 4: Number of cases per order where affinities between *stepwise_agg* and *direct_agg* differed per database. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in affinities (standard deviation in parenthesis).

	Australia (N = 12)	Europe (N = 39)	New Zealand (N = 22)	North America (N = 7)
Order				
Coleoptera	0 (0%)	0 (0%)	4 (18%)	0 (0%)
Diptera	0 (0%)	6 (15%)	5 (23%)	0 (0%)
Ephemeroptera	0 (0%)	4 (10%)	0 (0%)	2 (29%)
Odonata	0 (0%)	9 (23%)	0 (0%)	0 (0%)
Onychura	0 (0%)	0 (0%)	0 (0%)	2 (29%)
Plecoptera	8 (67%)	8 (21%)	2 (9.1%)	0 (0%)
Spinicaudata	0 (0%)	0 (0%)	0 (0%)	2 (29%)
Trichoptera	4 (33%)	12 (31%)	11 (50%)	1 (14%)
Mean dev.	0.049 (0.09)	-0.037 (0.27)	-0.053 (0.17)	0.012 (0.43)
Mean abs. dev.	0.07 (0.07)	0.22 (0.16)	0.15 (0.10)	0.37 (0.17)

3.2 Comparison of trait affinity values *direct_agg* using median or mean

The comparison of the *direct_agg (median)* with the *direct_agg (mean)* yielded between 2.23 % (North America) to 17.05 % (Europe) deviating cases (Table 5). Deviations between trait affinities were in a range of -0.4 (family *Gripopterygidae*, trait size small) to 0.48 (family *Gripopterygidae*, trait size medium). Most deviations were small as the standard deviation of the deviances were close to 0.1 and the mean of the absolute deviances were close to 0 (Table 6, see also figure SI8). Deviances in aggregated trait affinities occurred for all investigated grouping features.

Table 5: Percentage of cases for which the comparison of *direct_agg (median)* to *direct_agg (mean)* resulted in different trait affinities.

Database	Deviating cases [%]	Number of cases
Europe	17.05	2200
North America	2.23	1300
Australia	6.75	400
New Zealand	7.16	1900

Table 6: Number of cases per order where affinities between *direct_agg (median)* and *direct_agg (mean)* differed per database. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in affinities (standard deviation in parenthesis). Where the taxa were unranked on order-level, the class is displayed.

	Australia ($N = 27$)	Europe ($N = 375$)	New Zealand ($N = 136$)	North America ($N = 29$)
Order or higher				
Amphipoda	0 (0%)	12 (3.2%)	0 (0%)	0 (0%)
Branchiopoda	0 (0%)	0 (0%)	0 (0%)	2 (6.9%)
Coleoptera	0 (0%)	58 (15%)	20 (15%)	0 (0%)
Diptera	0 (0%)	20 (5.3%)	24 (18%)	0 (0%)
Ephemeroptera	0 (0%)	41 (11%)	11 (8.1%)	4 (14%)
Gastropoda	0 (0%)	0 (0%)	0 (0%)	3 (10%)
Hemiptera	0 (0%)	14 (3.7%)	0 (0%)	2 (6.9%)
Littorinimorpha	0 (0%)	0 (0%)	0 (0%)	2 (6.9%)
Mollusca	0 (0%)	0 (0%)	8 (5.9%)	0 (0%)
Odonata	0 (0%)	56 (15%)	5 (3.7%)	0 (0%)
Onychura	0 (0%)	0 (0%)	0 (0%)	2 (6.9%)
Plecoptera	7 (26%)	60 (16%)	24 (18%)	2 (6.9%)
Spinicaudata	0 (0%)	0 (0%)	0 (0%)	2 (6.9%)
Trichoptera	20 (74%)	114 (30%)	44 (32%)	10 (34%)
Mean dev.	-0.024 (0.08)	-0.008 (0.12)	-0.016 (0.15)	-0.009 (0.16)
Mean abs. dev.	0.06 (0.05)	0.09 (0.07)	0.11 (0.11)	0.14 (0.08)

3.3 Deviances in trait affinity values between *direct_agg* and *weighted_agg*

Comparison of the *direct_agg (median)* with the *weighted_agg* showed varying levels of cases with deviating trait affinities across databases (between 3.46 % deviating cases for the North American database up to 19.27 % in the European database, Table 5). Likewise, the number of deviating cases varied for the comparison of the *direct_agg (mean)* with *weighted_agg*, but was generally lower (between no deviating case for New Zealand and 11.59 % for the European database). Actually, there were a small number of deviating cases for the New Zealand database. However, the deviances were extremely small (below the 10th decimal place). Hence, we neglected these small deviations. The absence of deviating cases and the similar amount of deviating cases for the comparison *direct_agg (median)* and *weighted_agg* to the comparison *direct_agg (median)* with *direct_agg (mean)*

(Section 3.2) for New Zealand reflects the fact that most entries in this database contain complete trait profiles.

Deviances in trait affinities occurred for both comparisons for all investigated grouping features. Also, for both comparisons deviations between trait affinities were in range of -0.49 (European database, family *Leuctridae*, trait crawler and trait semivoltine) to 0.49 ((European database, family *Leuctridae*, trait burrower). Standard deviations of the deviances between *direct_agg (median)* and *weighted_agg* ranged from 0.09 (Australia) to 0.16 (North America). The mean of absolute deviances ranged between 0.068 (Australia) to 0.135 (North America) (Table 8). Standard deviations of the deviances between *direct_agg (mean)* and *weighted_agg* were smaller and ranged between 0.06 (Australia) and 0.11 (North America), while the mean of absolute deviances were close to zero for all databases (from 0.037 Australia to 0.078 North America) (Table 9, see also figure SI 9).

Table 7: Percentage of cases for which *direct_agg (median)* and *weighted_agg* resulted in different trait affinities and cases for which *direct_agg (mean)* and *weighted_agg* resulted in different trait affinities.

Database	Deviating cases *direct_agg (median)* vs. *weighted_agg* [%]	Deviating cases *direct_agg (mean)* vs. *weighted_agg* [%]	Nr. of cases
Europe	19.27	11.59	2200
North America	3.46	2.62	1300
Australia	9.75	9.75	400
New Zealand	7.21	0	1900

Table 8: Number of cases per order where affinities between *direct_agg (median)* and *weighted_agg* differed per database. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in affinities (standard deviation in parenthesis). If taxa were unranked on order-level, the class is displayed.

	Australia N = 39	Europe N = 423	New Zealand N = 137	North America N = 45
Order or higher				
Amphipoda	0 (0%)	12 (2.8%)	0 (0%)	0 (0%)
Branchiopoda	0 (0%)	0 (0%)	0 (0%)	2 (4.4%)
Coleoptera	0 (0%)	58 (14%)	20 (15%)	0 (0%)
Diptera	0 (0%)	21 (5.0%)	24 (18%)	0 (0%)
Ephemeroptera	0 (0%)	60 (14%)	11 (8.0%)	8 (18%)
Gastropoda	0 (0%)	0 (0%)	0 (0%)	3 (6.7%)
Hemiptera	0 (0%)	14 (3.3%)	0 (0%)	3 (6.7%)
Littorinimorpha	0 (0%)	0 (0%)	0 (0%)	2 (4.4%)
Mollusca	0 (0%)	0 (0%)	8 (5.8%)	0 (0%)
Odonata	0 (0%)	57 (13%)	5 (3.6%)	0 (0%)
Onychura	0 (0%)	0 (0%)	0 (0%)	2 (4.4%)
Plecoptera	8 (21%)	73 (17%)	24 (18%)	4 (8.9%)
Spinicaudata	0 (0%)	0 (0%)	0 (0%)	2 (4.4%)
Trichoptera	31 (79%)	128 (30%)	45 (33%)	19 (42%)
Mean deviance	-0.017 (0.09)	-0.007 (0.13)	-0.016 (0.15)	-0.006 (0.16)
Mean abs. deviances	0.068 (0.07)	0.094 (0.09)	0.112 (0.11)	0.135 (0.08)

Table 9: Number of cases per order where affinities between *direct_agg (mean)* and *weighted_agg* differed per database. There were no deviating cases for the New Zealand database. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in affinities (standard deviation in parenthesis). If taxa were unranked on order-level, the class is displayed.

	Australia N = 39	Europe N = 255	North America N = 34
Order or higher			
Amphipoda	0 (0%)	1 (0.4%)	0 (0%)
Coleoptera	0 (0%)	18 (7.1%)	0 (0%)
Diptera	0 (0%)	2 (0.8%)	0 (0%)
Ephemeroptera	0 (0%)	51 (20%)	8 (24%)
Hemiptera	0 (0%)	14 (5.5%)	3 (8.8%)
Mollusca	0 (0%)	0 (0%)	0 (0%)
Odonata	0 (0%)	28 (11%)	0 (0%)
Plecoptera	8 (21%)	40 (16%)	4 (12%)
Trichoptera	31 (79%)	101 (40%)	19 (56%)
Mean deviance	0 (0.06)	0 (0.08)	0 (0.11)
Mean abs. deviances	0.037 (0.04)	0.041 (0.07)	0.078 (0.08)

4 Deviances in trait affinitites obtained by trait aggregation methods compared to traits assigned at family-level

Aggregated trait affinities using the *direct_agg (median)*, *direct_agg (mean)* and the *weighted_agg* were compared to trait affinities assigned at family-level by experts, which were available for the Australian and North American database for a limited subset of grouping features and taxa.

For the Australian database, we compared aggregated trait affinities with assigned trait affinities resolved at family-level for the grouping features feeding mode and size by using data from Chessman 2017 (Chessman 2018). We split our analysis into two parts: I) we compared *direct_agg (median)* and *direct_agg (mean)* to traits assigned at family-level using all available trait information from the Australian database for the two grouping features (total number of taxa = 1091). For this analysis, we could carry out the comparison to all taxa listed in Chessman 2017, which contained trait information for 220 families. Almost all species used from the Australian database had complete trait profiles for the used grouping features. II) Since the *weighted_agg* is basically the *direct_agg (mean)* for complete trait databases we compared the *weighted_agg* to traits assigned at family-level using the Australian database from section 3. As a result, we could compare results of the *weighted_agg* only with trait data of 16 families of the data from Chessman.

For the North American database, we compared aggregated trait affinities with assigned trait affinities on family-level for the grouping features feeding mode, respiration, size, voltinism, and locomotion from Pyne et al. Trait information in the Pyne database was on the nominal scale and was converted to binary traits prior to the comparison with aggregated trait affinities. Similar to the comparison of the Australian trait data, we divided our analysis into two parts. *Direct_agg (median)* and *direct_agg (mean)* were compared to traits assigned at family-level using all available trait information from the North American database for the five grouping features (total number of taxa = 785). Trait information on family-level in Pyne database was available for 94 families of which 89 were present in the aggregated North American database. Almost all species used from the North American database had complete trait profiles for the used grouping features. For the *weighted_agg* we used the North American database from section 3 and compared aggregated traits to traits assigned at family-level. As a result, we could compare results of the *weighted_agg* with trait data of 30 families of the data from Pyne.

4.1 Australia

4.1.1 *direct_agg*

The comparison of the *direct_agg (median)* to trait affinities assigned at family-level yielded in 16.45 % cases different trait affinities. Comparison of *direct_agg (mean)* to traits assigned at family-level yielded 22.57 % of deviating cases.

Most deviations occurred for both methods for the orders Coleoptera, Diptera and Trichoptera. Deviations in trait affinities between *direct_agg (median)* and at family-level assigned were in the maximum range possible of 1 to -1. These high deviations only occurred for the grouping feature size for 15 different families. Largest deviances observed for the grouping feature feeding mode were 0.66 and -0.66 (Table SI17). The standard deviation of the deviating trait affinities was 0.48 and the mean of absolute deviances 0.405 (Table 10).

Deviations in trait affinities between *direct_agg (mean)* and at family-level assigned trait affinities were in a range of 0.91 to to -0.98. Also here, these high deviances occurred in the grouping feature

size. The standard deviation of the deviations was 0.38 and the mean of absolute deviances 0.312 (Table 10).

Table 10: Number of cases per order where trait affinities between *direct_agg (median)* and *direct_agg (mean)* compared to at family-level assigned trait affinities by Chessman et al. differed. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in affinities (standard deviation in parenthesis).

	Dev. cases *direct_agg (median)* and family-lvl assigned traits $N = 258$	Dev. cases *direct_agg (mean)* and family-lvl assigned traits $N = 354$
Order		
Amphipoda	17 (6.6%)	21 (5.9%)
Architaenioglossa	2 (0.8%)	2 (0.6%)
Arhynchobdellida	1 (0.4%)	1 (0.3%)
Coleoptera	46 (18%)	54 (15%)
Collembola	6 (2.3%)	6 (1.7%)
Decapoda	8 (3.1%)	13 (3.7%)
Diptera	76 (29%)	88 (25%)
Ephemeroptera	12 (4.7%)	19 (5.4%)
Hemiptera	11 (4.3%)	14 (4.0%)
Isopoda	9 (3.5%)	17 (4.8%)
Lepidoptera	4 (1.6%)	5 (1.4%)
Lumbriculida	0 (0%)	2 (0.6%)
Mecoptera	2 (0.8%)	2 (0.6%)
Megaloptera	2 (0.8%)	2 (0.6%)
Neuroptera	2 (0.8%)	2 (0.6%)
Odonata	4 (1.6%)	6 (1.7%)
Plecoptera	7 (2.7%)	16 (4.5%)
Rhynchobdellida	0 (0%)	3 (0.8%)
Trichoptera	48 (19%)	76 (21%)
Unionida	0 (0%)	2 (0.6%)
Venerida	1 (0.4%)	3 (0.8%)
Mean deviances	-0.018 (0.48)	-0.001 (0.38)
Mean abs. deviances	0.405 (0.26)	0.312 (0.21)

4.1.2 *weighted_agg*

The comparison of the *weighted_agg* to trait affinities assigned at family-level by Chessman yielded in 41.4 % cases different trait affinities. Deviations ranged from 1 to -1. Deviations of 100 % only occurred for the grouping feature size (Figure SI10). The standard deviation of the deviating cases as well as the mean of the absolute deviances were similar across the orders with deviating cases (Table 11).

Table 11: Number of traits per family where trait affinities between *weighted_agg* compared to at family-level assigned trait affinities by Chessman 2017 differed. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in deviating affinities (standard deviation in parenthesis).

	Ephemeroptera, N = 7	Plecoptera, N = 7	Trichoptera, N = 39
Family			
Baetidae	5 (71%)	0 (0%)	0 (0%)
Caenidae	2 (29%)	0 (0%)	0 (0%)
Calamoceratidae	0 (0%)	0 (0%)	2 (5.1%)
Conoesucidae	0 (0%)	0 (0%)	2 (5.1%)
Ecnomidae	0 (0%)	0 (0%)	5 (13%)
Gripopterygidae	0 (0%)	7 (100%)	0 (0%)
Helicopsychidae	0 (0%)	0 (0%)	2 (5.1%)
Hydrobiosidae	0 (0%)	0 (0%)	4 (10%)
Hydropsychidae	0 (0%)	0 (0%)	5 (13%)
Hydroptilidae	0 (0%)	0 (0%)	3 (7.7%)
Leptoceridae	0 (0%)	0 (0%)	7 (18%)
Philopotamidae	0 (0%)	0 (0%)	4 (10%)
Philorheithridae	0 (0%)	0 (0%)	5 (13%)
Mean deviances	0 (0.44)	0 (0.49)	0 (0.46)
Mean abs. deviances	0.381 (0.16)	0.334 (0.33)	0.369 (0.27)

4.2 North America

4.2.1 *direct_agg*

Comparison of the *direct_agg (median)* to trait affinities assigned at family-level yielded in 18.04 % cases different trait affinities. Comparison of *direct_agg (mean)* to trait affinities assigned at family-level in 37.07 % cases.

Most of the deviations occurred for both methods for the orders Coleoptera, Diptera and Trichoptera and were in the maximum range possible of 1 to -1. For both methods these deviations occurred for all grouping features (Figure SI 18). The standard deviation of the deviating trait affinities for the comparison *direct_agg (median)* and at family-level assigned trait was 0.62 and the mean of absolute deviances 0.573. The standard deviation of the deviating trait affinities for the comparison *direct_agg (mean)* and at family-level assigned trait was 0.43 and the mean of absolute deviances 0.346 (Table 13).

Table 12: Number of cases per order where trait affinities between *direct_agg (median)* and *direct_agg (mean)* compared to at family-level assigned trait affinities by Pyne differed. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in affinities (standard deviation in parenthesis).

	Dev. cases *direct_agg (median)* and family-lvl assigned traits, $N = 273$	Dev. cases *direct_agg (mean)* and family-lvl assigned traits, $N = 561$
Order		
Coleoptera	46 (17%)	73 (13%)
Diptera	53 (19%)	109 (19%)
Ephemeroptera	64 (23%)	104 (19%)
Hemiptera	16 (5.9%)	34 (6.1%)
Megaloptera	4 (1.5%)	9 (1.6%)
Neuroptera	4 (1.5%)	4 (0.7%)
Odonata	12 (4.4%)	26 (4.6%)
Plecoptera	17 (6.2%)	57 (10%)
Trichoptera	57 (21%)	145 (26%)
Mean deviances	-0.029 (0.62)	-0.001 (0.43)
Mean abs. deviances	0.573 (0.24)	0.346 (0.26)

4.2.2 *weighted_agg*

The comparison of the *weighted_agg* to trait affinities assigned at family-level by Pyne yielded in 30.19 % cases different trait affinities. Deviations ranged from 1 to -1. Deviations of 100 % occurred for all investigated grouping features (Figures SI section6). The standard deviation of the deviance in affinity was between 0.53 (Trichoptera) and 0.81 (Ephemeroptera). The mean of the absolute deviances were ranged between 0.489 (Trichoptera) to 0.833 (Hemiptera) (Table 13). The high standard deviations and mean values of absolute deviances could be explained by the binary coding of the traits in the Pyne database.

Table 13: Number of traits per family where trait affinities between *weighted_agg* compared to at family-level assigned trait affinities by Pyne differed. Penultimate row shows the mean deviance in affinity for deviating cases (standard deviation in parenthesis). Last row shows the mean of absolute deviances in deviating affinities (standard deviation in parenthesis).

	Ephemeroptera $N = 36$	Hemiptera $N = 18$	Megaloptera $N = 4$	Odonata $N = 9$	Plecoptera $N = 20$	Trichoptera $N = 67$
Family						
Aeshnidae	0 (0%)	0 (0%)	0 (0%)	3 (33%)	0 (0%)	0 (0%)
Baetidae	4 (11%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Brachycentridae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (13%)
Chloroperlidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (20%)	0 (0%)
Corixidae	0 (0%)	6 (33%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Ephemeridae	8 (22%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Glossosomatidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (9.0%)
Gomphidae	0 (0%)	0 (0%)	0 (0%)	4 (44%)	0 (0%)	0 (0%)
Helicopsychidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (7.5%)
Heptageniidae	2 (5.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Hydropsychidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	11 (16%)
Isonychiidae	10 (28%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Lepidostomatidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (12%)
Leptoceridae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (13%)
Leptophlebiidae	2 (5.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Lestidae	0 (0%)	0 (0%)	0 (0%)	2 (22%)	0 (0%)	0 (0%)
Limnephilidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	7 (10%)
Naucoridae	0 (0%)	6 (33%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Nemouridae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (20%)	0 (0%)
Nepidae	0 (0%)	6 (33%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Perlidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (20%)	0 (0%)
Perlodidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (20%)	0 (0%)
Polymitarcyidae	4 (11%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Potamanthidae	6 (17%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pteronarcyidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (10%)	0 (0%)
Rhyacophilidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (6.0%)
Sialidae	0 (0%)	0 (0%)	4 (100%)	0 (0%)	0 (0%)	0 (0%)
Taeniopterygidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (10%)	0 (0%)
Uenoidae	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (12%)
Mean dev.	0 (0.81)	0 (0.89)	0 (0.91)	0 (0.80)	0 (0.87)	-0.007 (0.53)
Mean abs. dev.	0.732 (0.33)	0.833 (0.24)	0.75 (0.29)	0.72 (0.23)	0.807 (0.27)	0.489 (0.20)

5 Re-analysis of Szöcs et al. 2014 using harmonized grouping features

We replicated the data analysis of Szöcs et al. 2014 (Szöcs et al. 2014) using harmonized grouping features for *Body size, Feeding mode, Locomotion, Reproduction/Oviposition, Respiration, and Voltinism* (21 grouping features have been used in total). The harmonized grouping features are those that responded strongly to salinity in the study of Szöcs et. al. 2014, except for life cylce duration.

Overall, using the harmonized grouping features leads only to slightly different results. The sites with high salinity (downstream) are still characterized by multivoltine, oviparous, gill-respirating and shredder species. Only species with the trait life cycle duration > 1 year do not anymore characterize sites with high salinization. Also, life cycle duration $<= 1$ year is not anymore characterizing sites not impacted by salinity. Transition and sites upstream from the point source are still characterized by univoltine species and species that lay their eggs in aquatic environment. Below, we display the results of this analysis and the graphs from the original paper. An overview over the harmonization for the European trait databases can be found in the supporting information in section 6.
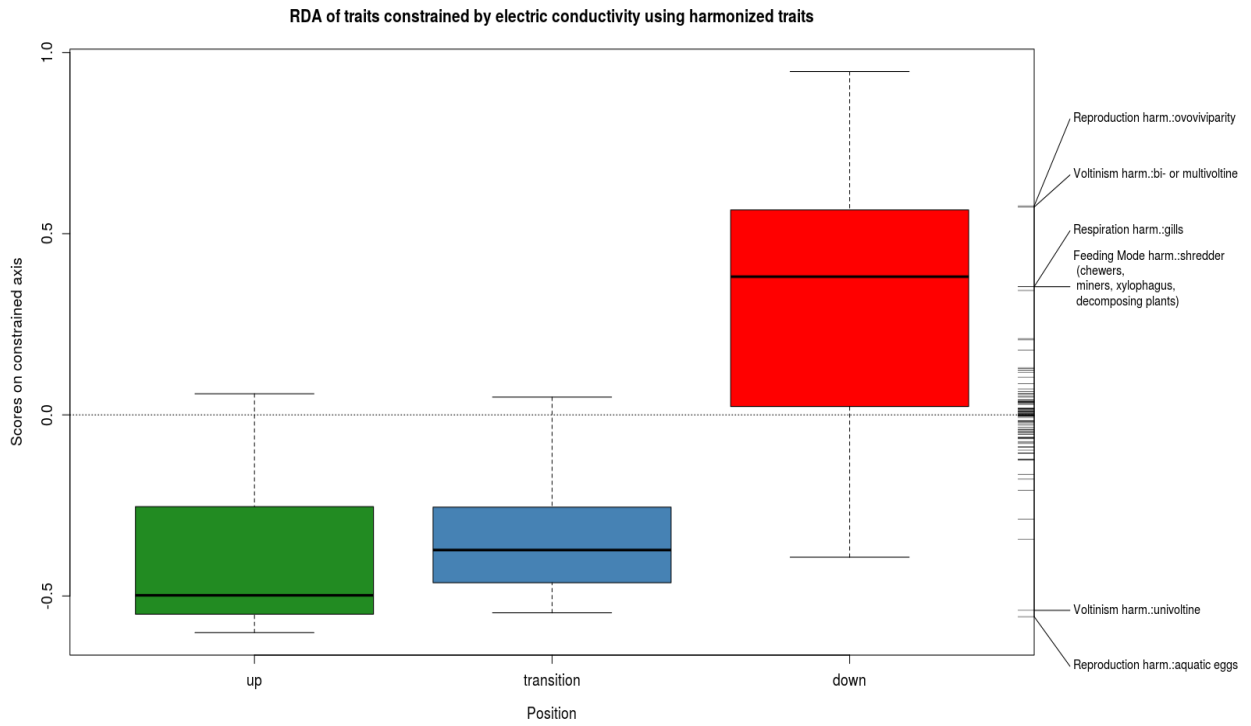
Figure 1: RDA of traits constrained by electric conductivity using harmonized grouping features. Boxplot of site scores along the conductivity axis (31.44% explained variance, p = 0.001, 1000 permutations). Rug on the left indicates trait scores on the conductivity axis. Only traits with a mahalanobis distance greater than 5.02 were labeled in accordance to the procedure in Szöcs et al. 2014.
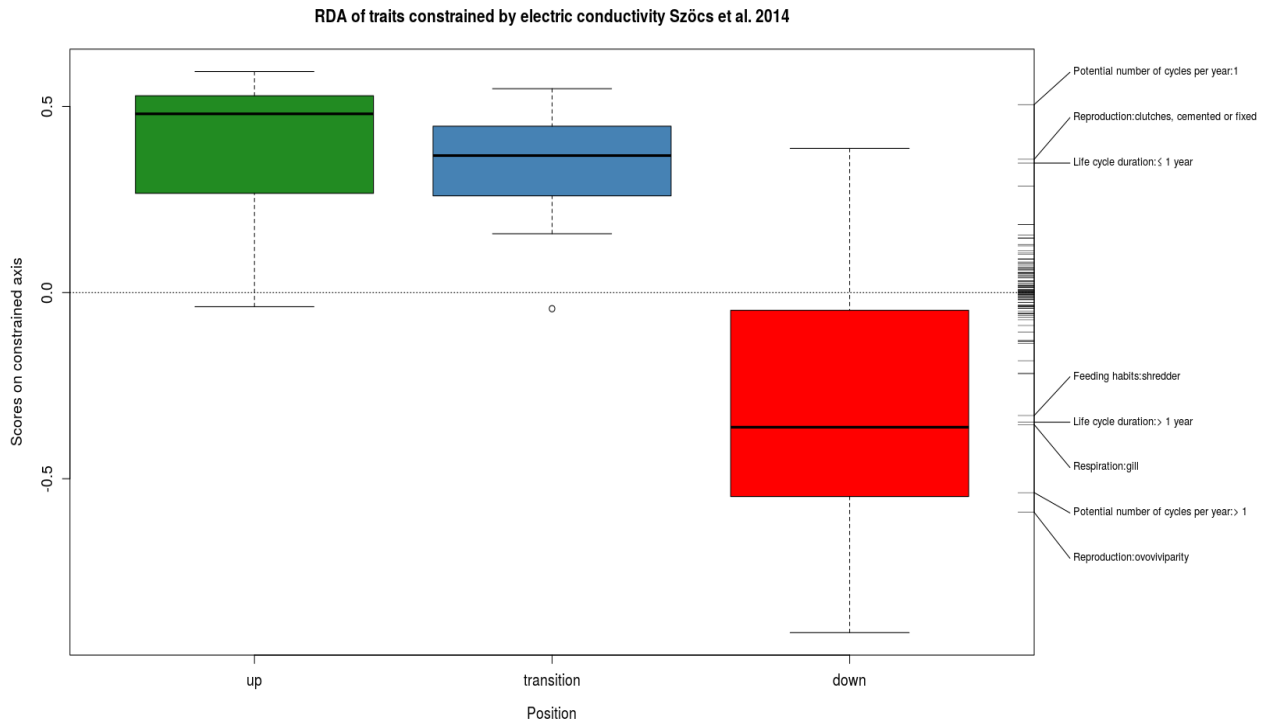
Figure 2: RDA of traits constrained by electric conductivity. Boxplot of site scores along the conductivity axis (30.09% explained variance, p = 0.001, 1000 permutations). Rug on the left indicates trait scores on the conductivity axis. Only traits with a mahalanobis distance greater than 5.02 were labeled.

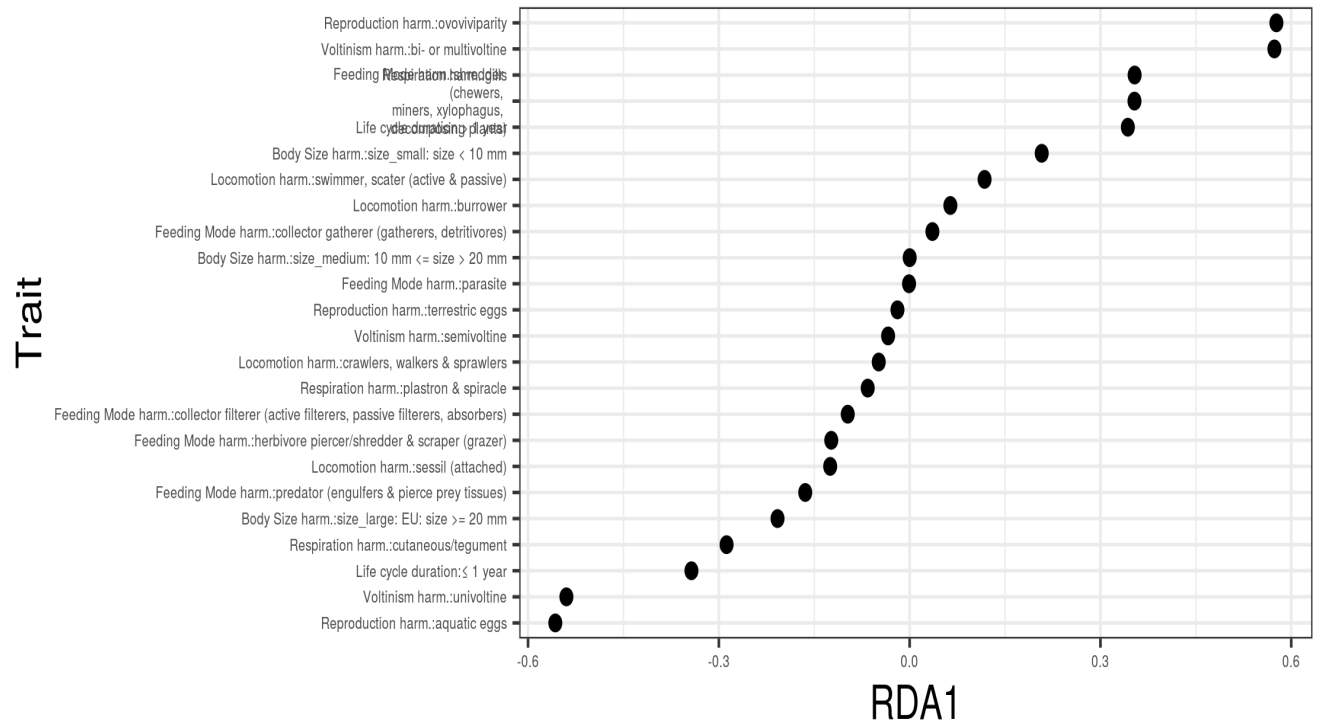# Trait distribution along first RDA axis

Figure 3: Trait scores on the first RDA axis for harmonized traits and traits of the grouping feature *life cycle duration.*
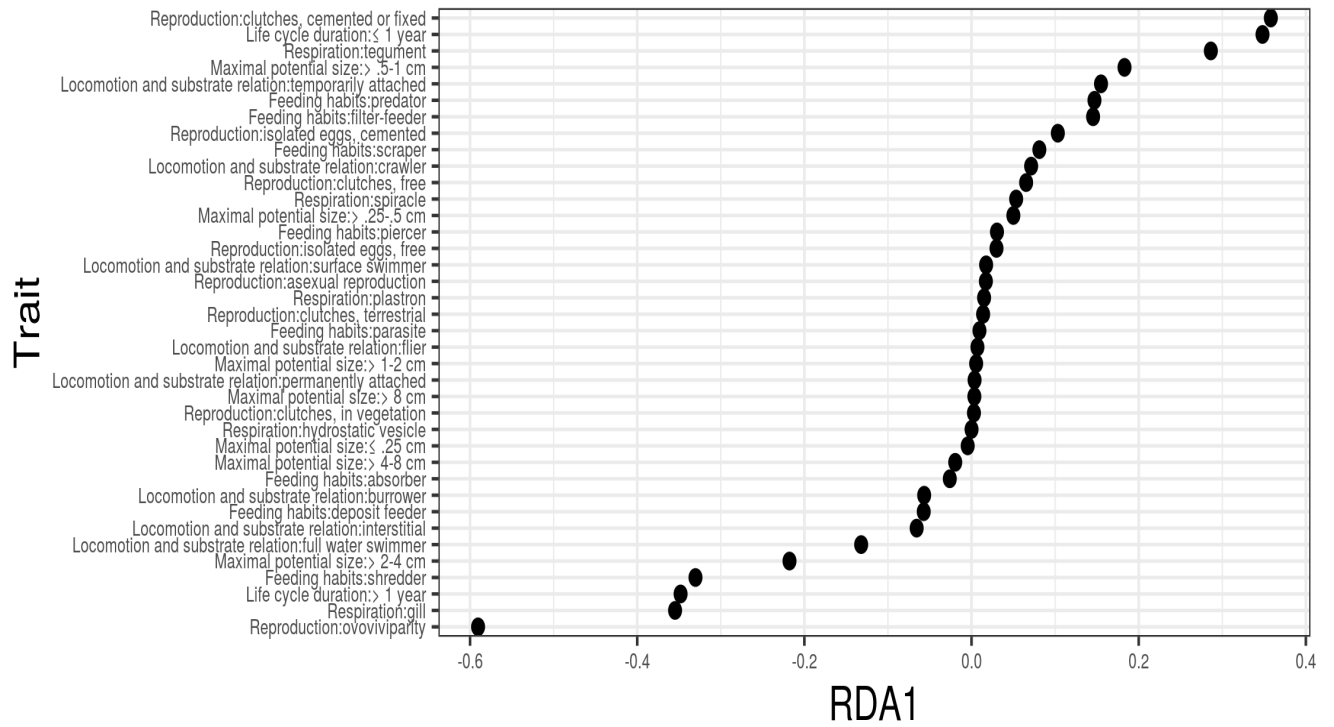
Figure 4: Trait scores on the first RDA axis for the traits responding to high salinity in Szöcs et al. 2014 .

Linear models of trait proportions

Linear models of trait proportions with harmonized traits:

Table 14: Results of linear models for the four selected harmonized traits and life cycle duration > 1 year. Trait proportions were logit transformed prior model building, estimates are on the logit scale. Although years were statistically not significant we kept this factor in the model to avoid temporal autocorrelation. Bold values indicate statistically significant effects ($p < 0.05$).

	Feeding mode: shredder	Life cycle duration: > 1 year	Voltinism: bi- or multivoltine	Reproduction: ovoviviparity	Respiration: gills
Intercept (= upstream)	**-1.041**	**-0.486**	0.375*	**-0.823**	0.092
Downstream	**0.926**	**0.605**	**1.376**	**1.684**	**0.854**
Downstream x 2008	-0.117	0.106	-0.235	-0.088	-0.317
Downstream x 2009	0.030	-0.056	0.001	0.245	0.180
Year 2008	-0.167	-0.115	0.033	-0.182	-0.151
Year 2009	0.175	0.086	-0.088	0.246	0.141

* p.value = 0.055

Linear models of trait proportions Szöcs et al. 2014:

Table 15: Results of linear models for the five selected traits for Szöcs et al. 2014. Trait proportions were logit transformed prior model building, estimates are on the logit scale. Although years were statistically not significant we kept this factor in the model to avoid temporal autocorrelation. Bold values indicate statistically significant effects ($p < 0.05$).

|	Feeding habits: shredder	Life cycle duration: > 1 year	Cycles per year: > 1	Reproduction: ovoviviparity	Respiration: gills
Intercept (= upstream)	**-0.853**	**-0.478**	**0.603**	**-0.838**	0.111
Downstream	**0.819**	**0.594**	**1.297**	**1.679**	**0.839**
Downstream x 2008	-0.155	0.102	-0.227	-0.070	-0.314
Downstream x 2009	0.073	-0.053	-0.020	0.248	0.176
Year 2008	-0.122	-0.112	0.026	-0.192	-0.154
Year 2009	0.167	0.084	-0.104	0.250	0.139

Trait proportions over time
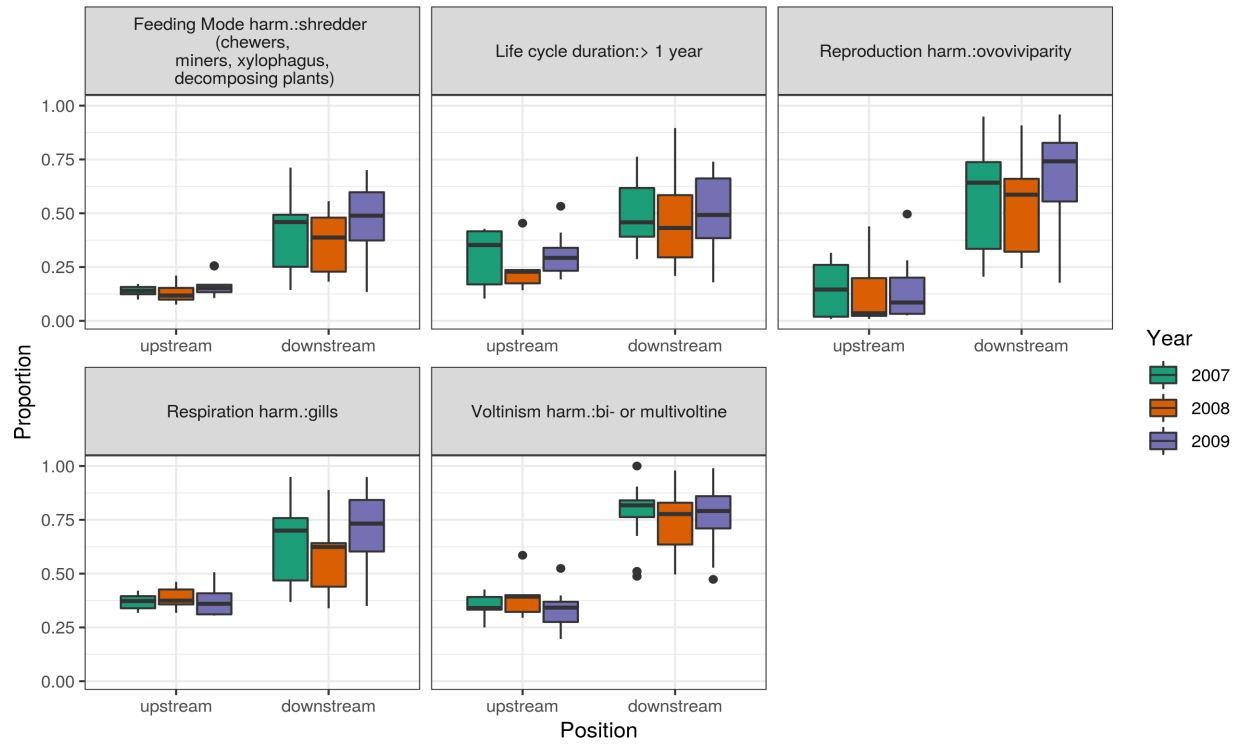
Figure 5: Proportions for the four harmonized traits that have been promoted by salinization and life cycle duration > 1 year for down- and upstream sites.
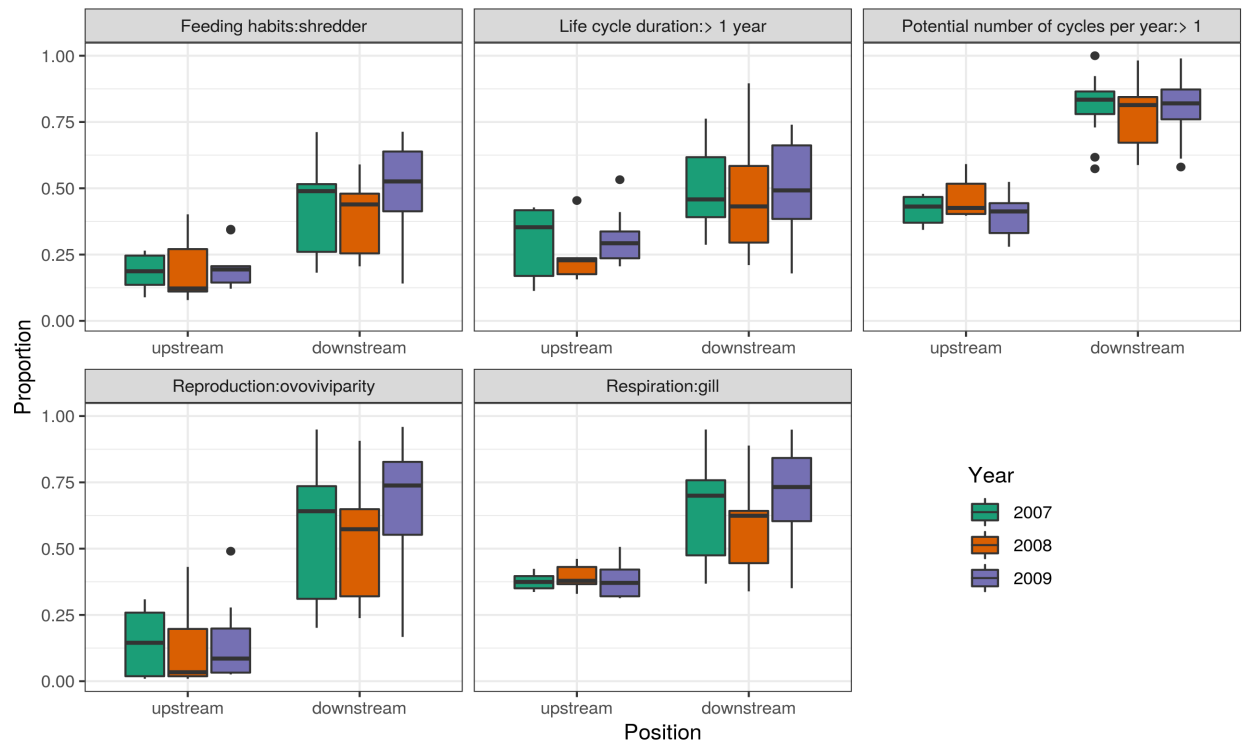
Figure 6: Proportions for five selected traits for down- and upstream sites (traits that have been promoted by salinization) from Szöcs et al. 2014.

6 Discrepancies in trait definitions

Table 16: Overview of discrepancies between trait definitions of the investigated trait databases. Grouping feature size was not available for the Freshwaterecology database. Grouping feature body form was not available for the Freshwaterecology, Tachet and North American (Twardochleb) databases.

Database	Grouping feature	Nr. of traits	Coding	Definition differences	Notes
Freshwater-ecology	Feeding Mode	10	10 point assign-ment system	• Shredders: feed from fallen leaves, plant tis-sue, CPOM • Predator: eating from prey • Filter feeders: distinguish between passive and active filterers	Contains traits miners and xylophagus, which can be assigned to the shredders
Tachet	Feeding Mode	7	fuzzy $[0-3]$	• Shredders: eat coarse detritus, plant or an-imal material. • Predators: carvers, engulfers & swallowers • Filter-feeders: no distinction between active and passive filterers	Piercers (plant & ani-mals) are an additional trait

Region	Trait	No.	Scale	Description
North America (Twardochleb)	Feeding Mode	6	categorical	• Shredder: insects that shred decomposing vascular plant tissue • Predator: engulfers & piercers (prey tissues) • Filter-feeders: no distinction between active and passive	Trait *herbivore* includes among others insects that shred *living* aquatic plants. Gatherers denoted as deposit-feeders
North America (Vieira)	Feeding Mode	8	categorical	• Similar trais compared to North America (Twardochleb), but contains traits Scraper/grazer and Piercer herbivore instead of trait herbivore.	Grouping feature defined as feeding guild describing primary mode of food collection No precise trait definitions given
New Zealand	Feeding Mode	6	fuzzy [0 − 3]	• Filter-feeders: no distinction between active and passive filterers	Trait parasite is not assessed No precise definitions given
Australia	Feeding Mode	16[1]	binary (0 or 1); proportional scale [0 − 1]; fuzzy [0 − 3]	• Shredder: detrivore[2] • Predator: piercer & engulfer • Filter-feeder: no distinction between active and passive filterers	No precise definitions given. Trait *herbivore* includes among others the trait shredder.

Source	Trait	N	Type	Description	Notes
Freshwater-ecology	Voltinism	6	single category assignment system	• Semivoltine: defined as one generation in two years • Three traits are used to classify taxa with two or more generations per year. • Multivoltine: taxa are defined as having *more than three* generations per year.	Taxa are classified according to different floristic regions (arctic, boreal, ...) Contains as only database a trait for taxa with flexible number of life cycles per year
Tachet	Voltinism	3	fuzzy codes [0 − 3]	• Semivoltine: defined as one generation (life cycle) in *at least* two years • Multivoltine: defined as *at least* two successive generations per year	Semantic: multivoltine is termed polyvoltine
North America (Twardochleb)	Voltinism	3	categorical	• Semivoltine: defined as < 1 generations per year • Multivoltine: defined as > 1 generations per year	
North America (Vieira)	Voltinism	3	categorical	• Voltinism classification identical to North America (Twardochleb) database	

New Zealand	Voltinism	3	fuzzy codes [0 – 3]	• semivoltine: defined as < 1 reproductive cycles per year • multivoltine: defined as > 1 reproductive cycles per year.	Multivoltine is termed plurivoltine The authors note that this grouping feature varies with temperature and latitude
Australia	Voltinism	7	binary (0 or 1); proportional scale [0 – 1]; fuzzy [0 – 3]	• Multiple generations per year are captured in several traits (depending on the initial database from which they were taken): 1-2 generations per year, bi/multivoltine, up to 5 generations per year, up to 10 generations per year	

Database	Trait	#	Coding	Definitions	Additional traits
Freshwater-ecology	Locomotion	6	10 point assignment system	• Swimming: database distinguishes passive movement (floating or drifting) using the trait swimming/seating and active movement using the trait swimming/diving. • Burrowing/boring: defined for *soft and hard* substrate • Sprawling and walking combined into one trait	Contains a trait *other* for locomotion types like flying or jumping
Tachet	Locomotion	8	fuzzy codes [0 − 5]	• Swimming: distinguishes surface swimmers (over and under the water surface) and full water swimmers (e.g. Baetidae) • Burrowing: defined as burrowing within the first centimeters of the benthic fine sediment. • Crawler: defined as crawling over the bottom substrate. Traits sprawlers or walkers do not exist. • Sessil: captured in the two traits temporarily and permanently attached. Freshwatere-cology, both North American databases and the New Zealand database do not make this distinction.	Contains a trait *flier* encompassing active and passive fliers. Also contains a trait interstitial (endobenthic)

Source	Group	N	Type	Traits	Grouping feature
North America (Twardochleb)	Locomotion	10	categorical	• Swimming: defined as adapted for "fishlike" swimming. • Skating: separately defined as organisms that skate on the surface where they feed on organisms trapped in the surface film. • Burrowing: defined as inhabiting *fine* sediment of streams and lakes • Crawler: defined for crawling on bottom sediment *and* on floating leaves.	Grouping feature termed habit The grouping feature also comprises the traits clinger, sprawler, climber, planktonic and other.
North America (Vieira 2006)	Locomotion	9	categorical	• Contains identical traits as the North America (Twardochleb) database, except that it does not contain the trait crawler.	Grouping feature termed habit. Is based on how organisms deal with flow.
New Zealand	Locomotion	4	fuzzy codes	• Databases uses the four traits swimmer, crawler, burrower and attached.	Grouping feature termed attachment to substrate of aquatic stages (excluding eggs) No precise definitions given

Australia	Locomotion	9	binary (0 or 1); fuzzy [0 − 3]	• Swimming: distinguishes swimmer and skater • Crawling: database contains traits crawler, sprawler, climber and clinger • Sessil: distinguishes temporary and permanently attached	No precise definitions given Flying and swimming ability are assessed as well, but are considered part of the grouping feature dispersal.
Freshwater-ecology	Respiration	7	presence/absence assignment system	• Plastron and spiracle (aerial) are two separate traits	Contains the trait hydrostatic vesicle. Contains respiration using air stores of aquatic plants (tapping) and excursion/extension to the surface which can be allocated to the taxa with spiracles.
Tachet	Respiration	5	fuzzy codes [0 − 3]	• Spiracle trait definition includes respiration using air stores of aquatic plants	Contains the trait hydrostatic vesicle.
North America (Twardochleb)	Respiration	3	categorical	• Plastron and spiracle defined as one trait	

Region	Trait	No.	Coding	Details	Notes
North America (Vieira 2006)	Respiration	8	categorical	• Distinguishes temporary air store and tracheal gills • Distinguishes spiracular gills, plastron, atmospheric breathers, and plant breathers	No precise definition of traits given. Contains the trait Hemoglobin, which is not compatible with the respiration traits occurring in the other trait databases.
New Zealand	Respiration	4	fuzzy codes [0 − 3]	• Distinguishes plastron and spiracle (here termed aerial)	No precise definition given
Australia	Respiration	10	binary (0 or 1); proportional scale [0 − 1] ; fuzzy [0 − 3]	• Plastron and spiracle (here termed aerial) occur as two separate traits and as combined trait • Contains the traits air (plants), atmospheric, functional spiracles, plastron	No precise definitions given Contains the traits ”plastron and gills” and pneomostome which are not compatible with the respiration traits occurring in the other trait databases.
Freshwater-ecology	Reproduction	9	presence/absence assignment system	• Form and location of eggs described and mode of reproduction • Locations: freely in water, fixed, vegetation, riparian zone	Includes parasitic reproduction

	Trait group	No.	Coding type	Description	Grouping feature combines mode of reproduction and oviposition tactics
Tachet	Reproduction	8	fuzzy codes [0 − 3]	• Form and location of eggs described and mode of reproduction	
North America (Vieira)	Reproduction	10	categorical	• Traits describe the location of oviposition precisely (e.g. algal mats, bank soil, ...) • Ovoviviparous taxa are mentioned in a comment column	No precise definitions given Grouping feature termed oviposition behavior
New Zealand	Reproduction	4	fuzzy codes [0 − 3]	• Information captured in two grouping features: oviposition site using four traits and egg/egg mass describing if eggs are free, cemented, or in/on body. • Oviposition site defined more broadly than in the other databases (water surface, submerged, terrestrial, endophytic)	No precise definitions given
Australia	Reproduction	13[3]	categorical	• Contains traits that precisely describe oviposition location • Also contains traits broadly describing oviposition (aquatic eggs, terrestrial eggs, ovoviviparity) • Contains traits for mode of reproduction	No precise definitions given

	Trait		Type	Description	Notes
Tachet	Size	7	fuzzy codes $[0-3]$	• Describes the maximal potential size that *the last* aquatic stage of the taxon can reach. All other databases describe the maximal potential size without mentioning the aquatic stage. • Highest resolved size classification: $<= 0.25cm, > 0.25 - 0.5cm, 0.5 - 1cm, 1 - 2cm, 2 - 4cm, 4 - 8cm, > 8cm$ • Draws no threshold what a small, medium and large organism is	
North America (Twardochleb)	Size	3	categorical	• Defines small $(< 9mm)$, medium $(9 - 16mm)$ and large $(> 16mm)$.	Size classification uses not the same ranges as the size classifications in the Tachet and New Zealand trait databases.
North America (Vieira)	Size	3	categorical	• Identical to North America (Twardochleb)	
New Zealand	Size	5	fuzzy codes $[0-3]$	• Size traits similarly defined to Tachet, but without the smallest and largest size traits • Draws no threshold what a small, medium and large organism is	

Region	Trait	No.	Data type	Description	Comments
Australia	Size	9	Numeric, fuzzy [0 − 3], binary (0 or 1)	• Contains a size trait with numeric size values • Contains traits classifying size like Tachet (except highest and lowest) and like North American trait databases	
North America (Vieira)	Body Form	5	categorical	• Classifies body form using the traits streamlined, dorsoventrally flattened, tubular, round, and bluff (blocky)	No precise trait definitions given
New Zealand	Body Form	4	fuzzy codes [0 − 3]	• Distinguishes streamlined, flattened, cylindrical, spherical	No precise trait definitions given
Australia	Body Form	4	fuzzy codes [0 − 3]	• Identical to New Zealand trait database	No precise definitions given

[1] *Many traits were just semantically different.*
[2] *Traits from Botwe et al.*
[3] *Few traits not counted due to their ambiguous nature (rather comments).*

References

Chessman, B. C. (Mar. 1, 2018). "Dissolved-oxygen, current and temperature preferences of stream invertebrates estimated from field distributions: application to assemblage responses to drought". In: *Hydrobiologia* 809.1, pp. 141–153. ISSN: 1573-5117. DOI: 10.1007/s10750-017-3455-1. URL: https://doi.org/10.1007/s10750-017-3455-1 (visited on 09/08/2019).

Szöcs, Eduard et al. (Jan. 2014). "Effects of anthropogenic salinization on biological traits and community composition of stream macroinvertebrates". In: *Science of The Total Environment* 468-469, pp. 943–949. ISSN: 00489697. DOI: 10.1016/j.scitotenv.2013.08.058. URL: https://linkinghub.elsevier.com/retrieve/pii/S0048969713009728 (visited on 03/26/2020).

Vieira, Nicole K M et al. (2006). "A Database of Lotic Invertebrate Traits for North America". In: p. 19.

Supporting Information

Deviance in trait affinities between *stepwise_agg* and *direct_agg (median)*
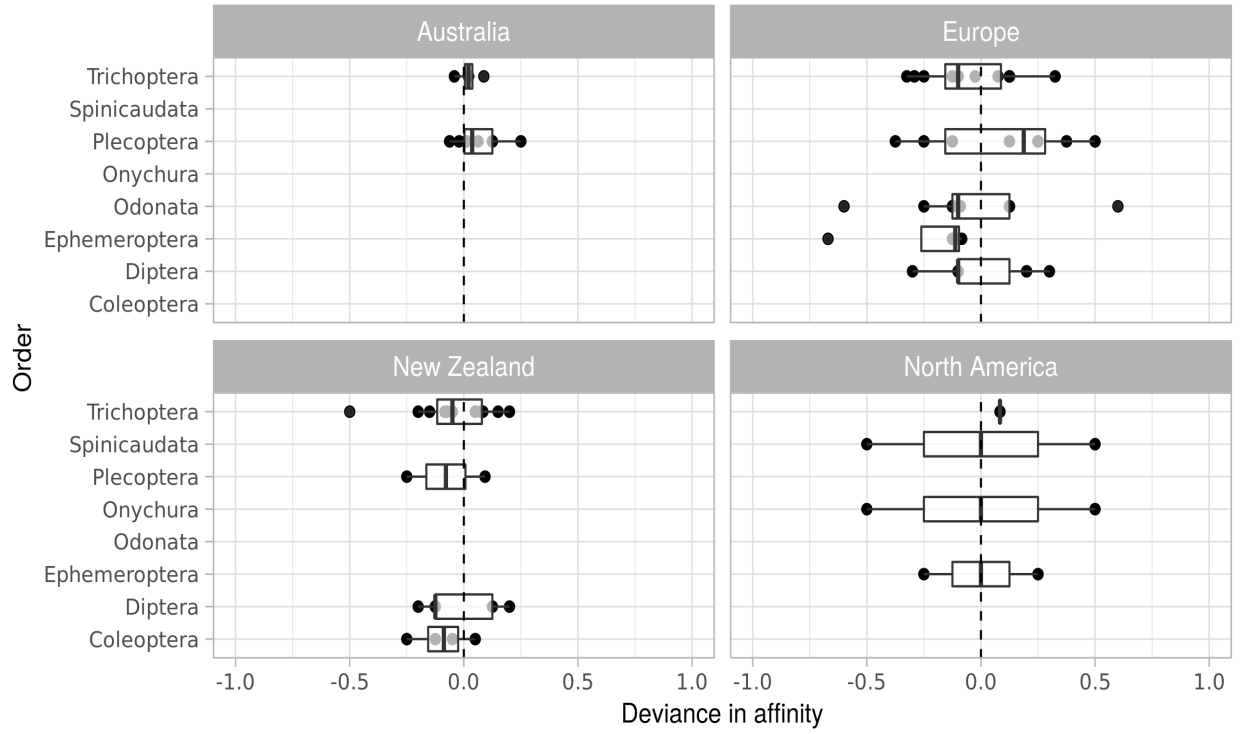

Figure 7: Affinity deviances for all deviating cases between the *stepwise_agg* and *direct_agg*, see section 3.1.

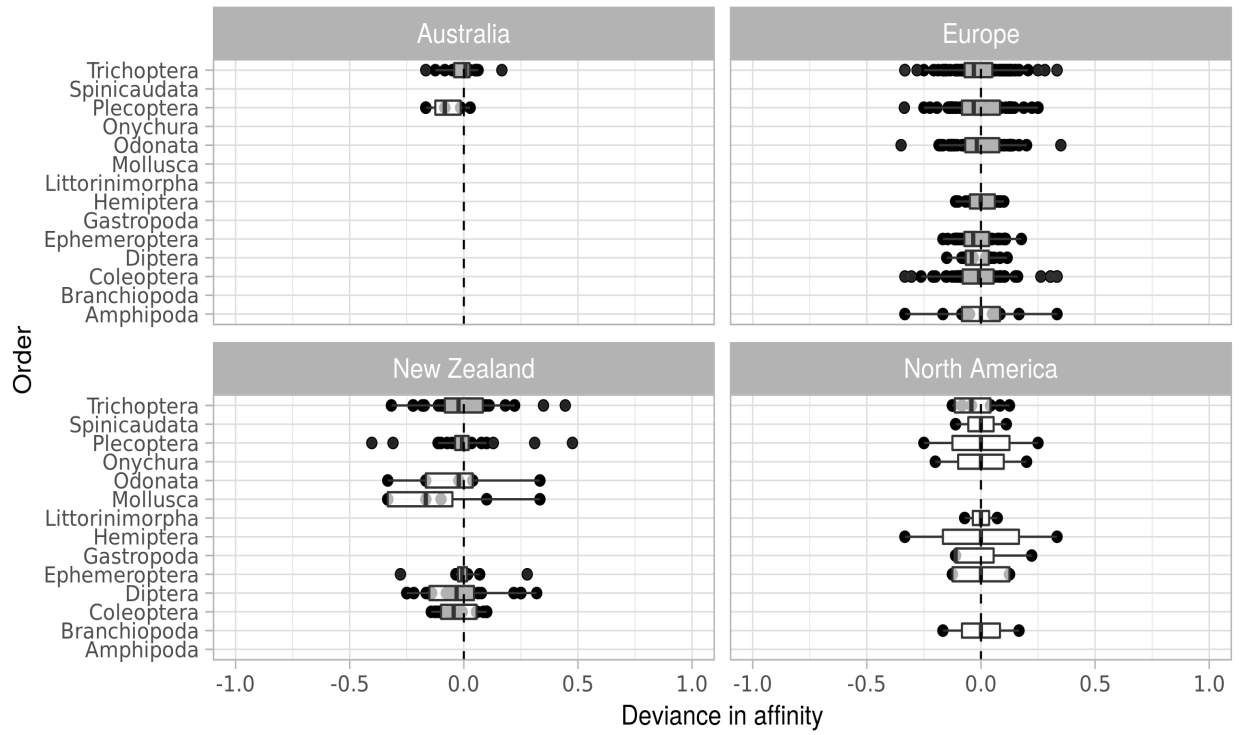# Comparison of trait affinity values *direct_agg* using median or mean

Figure 8: Affinity deviances for all deviating cases between the *direct_agg (median)* and *direct_agg (mean)*, see section 3.2.

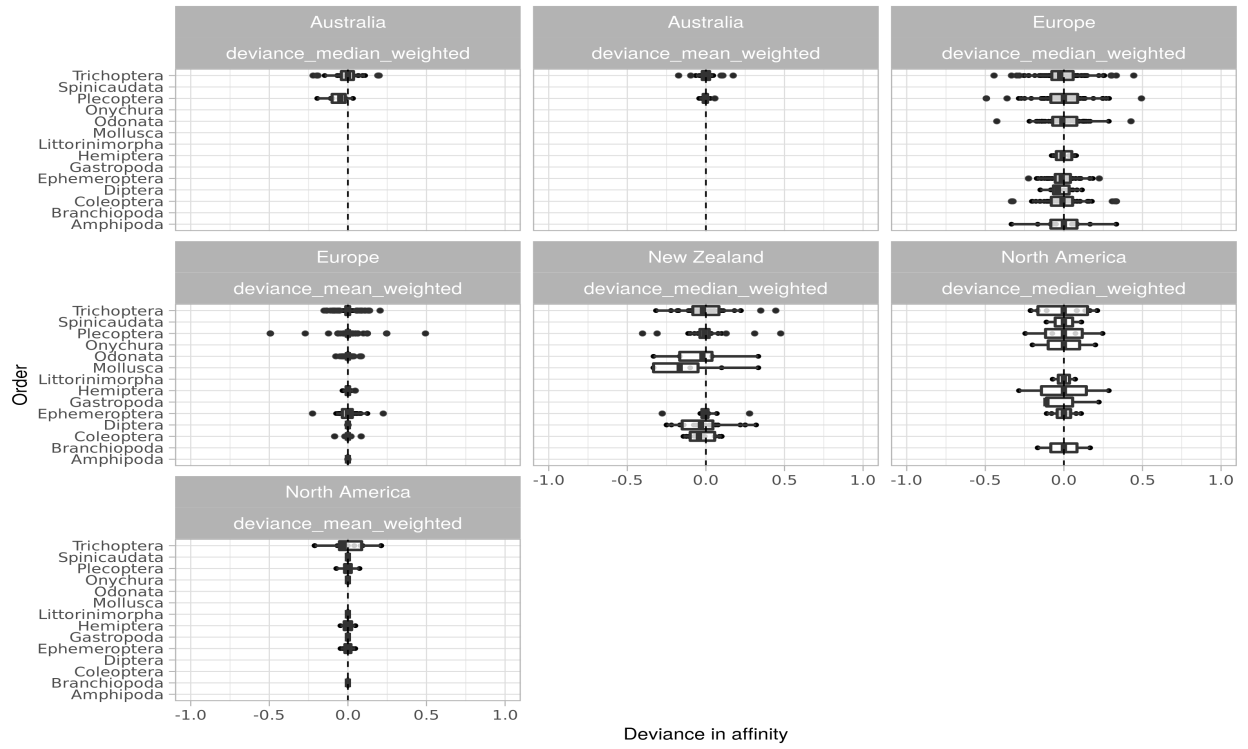**Deviances in trait affinities between *direct_agg* and *weighted_agg***

Figure 9: Affinity deviances for all deviating cases of comparisons between the *direct_agg (median)* and *direct_agg (mean)* with *weighted_agg*, see section 3.3.

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

Australia - *direct_agg*

Table 17: Cases where trait affinities aggregated through *direct_agg* differed more than 60 % from trait affinities assigned at family-level by Chessman 2017.

| Family	Order	Trait	Affinity Chessman	Affinity *direct_agg (median)*	Affinity *direct_agg (mean)*
Atyidae	Decapoda	feed_gatherer	0.33	1.00	0.77
Baetidae	Ephemeroptera	size_small	0.00	1.00	0.72
Baetidae	Ephemeroptera	size_medium	1.00	0.00	0.28
Blephariceridae	Diptera	size_small	0.00	1.00	0.60
Blephariceridae	Diptera	size_medium	1.00	0.00	0.40
Chironomidae	Diptera	size_small	0.00	1.00	0.73
Chironomidae	Diptera	size_large	1.00	0.00	0.02
Coenagrionidae	Odonata	size_medium	0.00	1.00	0.65
Coenagrionidae	Odonata	size_large	1.00	0.00	0.35
Conoesucidae	Trichoptera	size_small	0.00	0.75	0.58
Conoesucidae	Trichoptera	size_medium	1.00	0.25	0.42
Corixidae	Hemiptera	size_small	0.00	1.00	0.91
Corixidae	Hemiptera	size_medium	1.00	0.00	0.09
Corophiidae	Amphipoda	size_medium	1.00	0.25	0.25
Crambidae	Lepidoptera	size_medium	0.00	1.00	0.75
Crambidae	Lepidoptera	size_large	1.00	0.00	0.25
Dytiscidae	Coleoptera	size_medium	0.00	1.00	0.62
Dytiscidae	Coleoptera	size_large	1.00	0.00	0.06
Elmidae	Coleoptera	size_small	0.00	1.00	0.75
Elmidae	Coleoptera	size_medium	1.00	0.00	0.25
Ephydridae	Diptera	size_medium	1.00	0.33	0.33
Gerridae	Hemiptera	size_medium	0.00	1.00	0.64
Gerridae	Hemiptera	size_large	1.00	0.00	0.07
Gripopterygidae	Plecoptera	size_large	1.00	0.00	0.33
Gyrinidae	Coleoptera	size_small	0.00	1.00	0.73
Gyrinidae	Coleoptera	size_large	1.00	0.00	0.13
Hydrophilidae	Coleoptera	size_medium	0.00	1.00	0.84
Hydrophilidae	Coleoptera	size_large	1.00	0.00	0.04
Leptoceridae	Trichoptera	size_medium	0.00	1.00	0.82
Leptoceridae	Trichoptera	size_large	1.00	0.00	0.10
Leptophlebiidae	Ephemeroptera	size_large	1.00	0.00	0.02
Ochteridae	Hemiptera	size_small	0.00	0.75	0.75
Ochteridae	Hemiptera	size_medium	1.00	0.25	0.25
Tabanidae	Diptera	feed_predator	1.00	0.33	0.33
Thaumaleidae	Diptera	feed_herbivore	1.00	0.33	0.33
Thaumaleidae	Diptera	size_medium	1.00	0.33	0.33
Viviparidae	Architaenioglossa	size_medium	1.00	0.00	0.17
Viviparidae	Architaenioglossa	size_large	0.00	1.00	0.83

Australia - *weighted_agg*
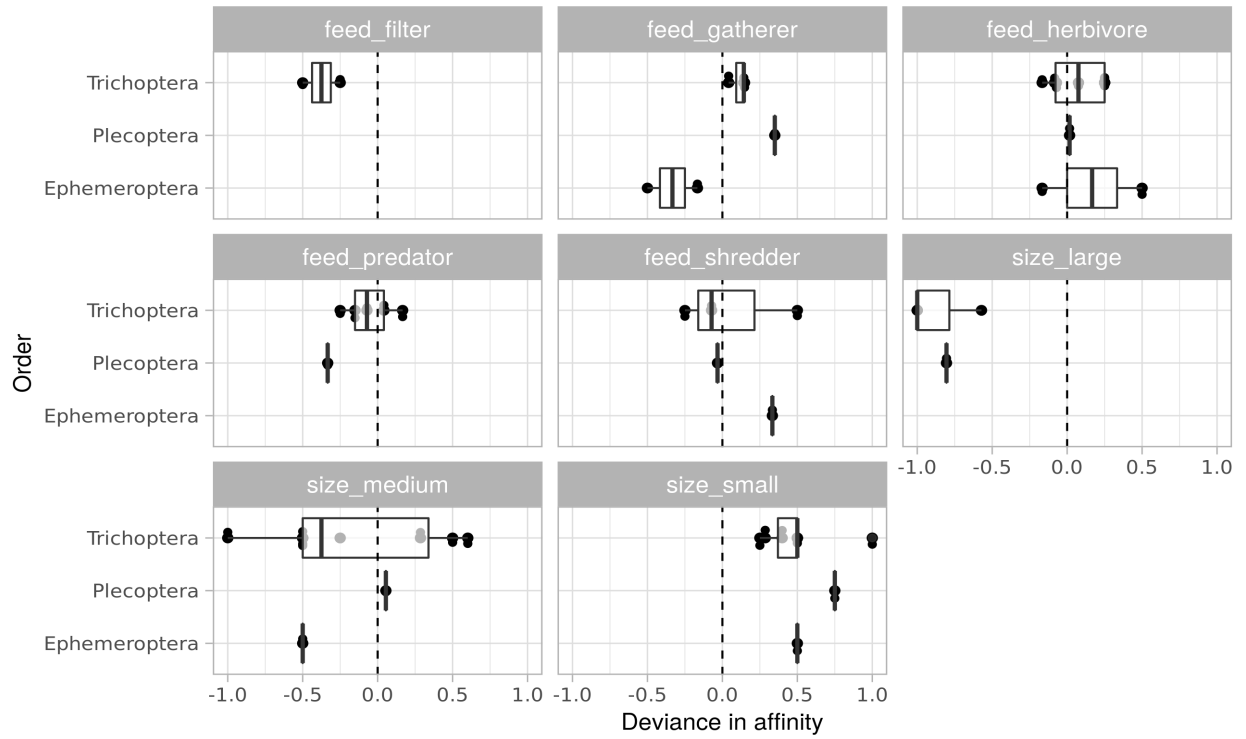
Figure 10: Affinity deviances for all deviating cases between *weighted_agg*, see also section 3.2.

North America - *direct_agg*

Table 18: Cases where trait affinities aggregated using *direct_agg* differed 100 % from trait affinities assigned at family-level by Pyne.

Family	Order	Trait	Affinity Pyne	Affinity *direct_agg (median)*	Affinity *direct_agg (mean)*
Cordulegastridae	Odonata	size_large	0.00	1.00	1.00
Cordulegastridae	Odonata	size_medium	1.00	0.00	0.00
Cordulegastridae	Odonata	volt_semi	0.00	1.00	1.00
Cordulegastridae	Odonata	volt_uni	1.00	0.00	0.00
Dytiscidae	Coleoptera	resp_pls_spi	1.00	0.00	0.00
Dytiscidae	Coleoptera	resp_teg	0.00	1.00	0.97
Dytiscidae	Coleoptera	volt_semi	1.00	0.00	0.12
Dytiscidae	Coleoptera	volt_uni	0.00	1.00	0.66
Ephemeridae	Ephemeroptera	size_large	0.00	1.00	0.86
Ephemeridae	Ephemeroptera	size_medium	1.00	0.00	0.14
Ephemeridae	Ephemeroptera	volt_semi	1.00	0.00	0.36
Ephemeridae	Ephemeroptera	volt_uni	0.00	1.00	0.64
Gerridae	Hemiptera	size_medium	1.00	0.00	0.30
Gerridae	Hemiptera	size_small	0.00	1.00	0.70
Gomphidae	Odonata	size_large	0.00	1.00	0.78
Gomphidae	Odonata	size_small	1.00	0.00	0.00
Gyrinidae	Coleoptera	resp_pls_spi	1.00	0.00	0.00
Helicopsychidae	Trichoptera	resp_gil	1.00	0.00	0.00
Helicopsychidae	Trichoptera	resp_teg	0.00	1.00	1.00
Helophoridae	Coleoptera	resp_pls_spi	1.00	0.00	0.00
Helophoridae	Coleoptera	resp_teg	0.00	1.00	1.00
Hydraenidae	Coleoptera	resp_pls_spi	1.00	0.00	0.00
Hydraenidae	Coleoptera	resp_teg	0.00	1.00	1.00
Hydraenidae	Coleoptera	volt_uni	1.00	0.00	0.00
Hydrophilidae	Coleoptera	resp_pls_spi	1.00	0.00	0.00
Hydrophilidae	Coleoptera	resp_teg	0.00	1.00	0.88
Hydrophilidae	Coleoptera	feed_gatherer	1.00	0.00	0.12
Isonychiidae	Ephemeroptera	size_small	1.00	0.00	0.00
Isonychiidae	Ephemeroptera	feed_filter	0.00	1.00	1.00
Isonychiidae	Ephemeroptera	feed_gatherer	1.00	0.00	0.00
Leptophlebiidae	Ephemeroptera	feed_filter	1.00	0.00	0.10
Leptophlebiidae	Ephemeroptera	feed_gatherer	0.00	1.00	0.90
Muscidae	Diptera	locom_burrow	0.00	1.00	1.00
Muscidae	Diptera	locom_crawl	1.00	0.00	0.00
Nepidae	Hemiptera	volt_bi_multi	1.00	0.00	0.00
Nepidae	Hemiptera	volt_uni	0.00	1.00	1.00
Odontoceridae	Trichoptera	feed_herbivore	1.00	0.00	0.33
Oligoneuriidae	Ephemeroptera	locom_burrow	0.00	1.00	1.00
Oligoneuriidae	Ephemeroptera	size_medium	0.00	1.00	1.00

Oligoneuriidae	Ephemeroptera	size_small	1.00	0.00	0.00
Oligoneuriidae	Ephemeroptera	locom_crawl	1.00	0.00	0.00
Perlidae	Plecoptera	volt_semi	0.00	1.00	0.64
Perlidae	Plecoptera	volt_uni	1.00	0.00	0.36
Potamanthidae	Ephemeroptera	locom_burrow	1.00	0.00	0.00
Potamanthidae	Ephemeroptera	feed_filter	1.00	0.00	0.33
Potamanthidae	Ephemeroptera	feed_gatherer	0.00	1.00	0.67
Potamanthidae	Ephemeroptera	locom_crawl	0.00	1.00	1.00
Psephenidae	Coleoptera	size_medium	1.00	0.00	0.40
Psephenidae	Coleoptera	size_small	0.00	1.00	0.60
Sciomyzidae	Diptera	volt_bi_multi	1.00	0.00	0.00
Sciomyzidae	Diptera	volt_semi	0.00	1.00	1.00
Sialidae	Megaloptera	size_large	0.00	1.00	1.00
Sialidae	Megaloptera	size_medium	1.00	0.00	0.00
Tipulidae	Diptera	feed_shredder	1.00	0.00	0.32

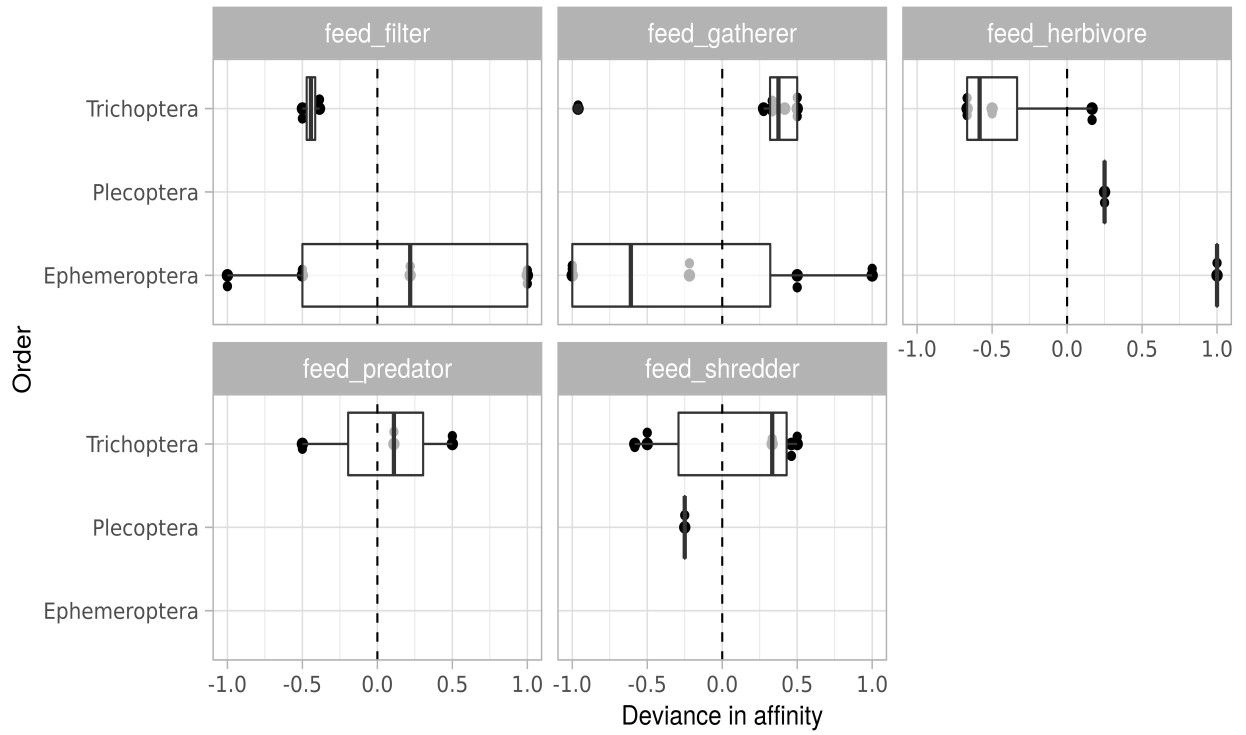**North America - *weighted_agg***

Figure 11: Affinity deviances for all deviating cases between the *weighted_agg* and traits assigned at family-level by Pyne for the grouping feature feeding mode. See also section 4.2.2.
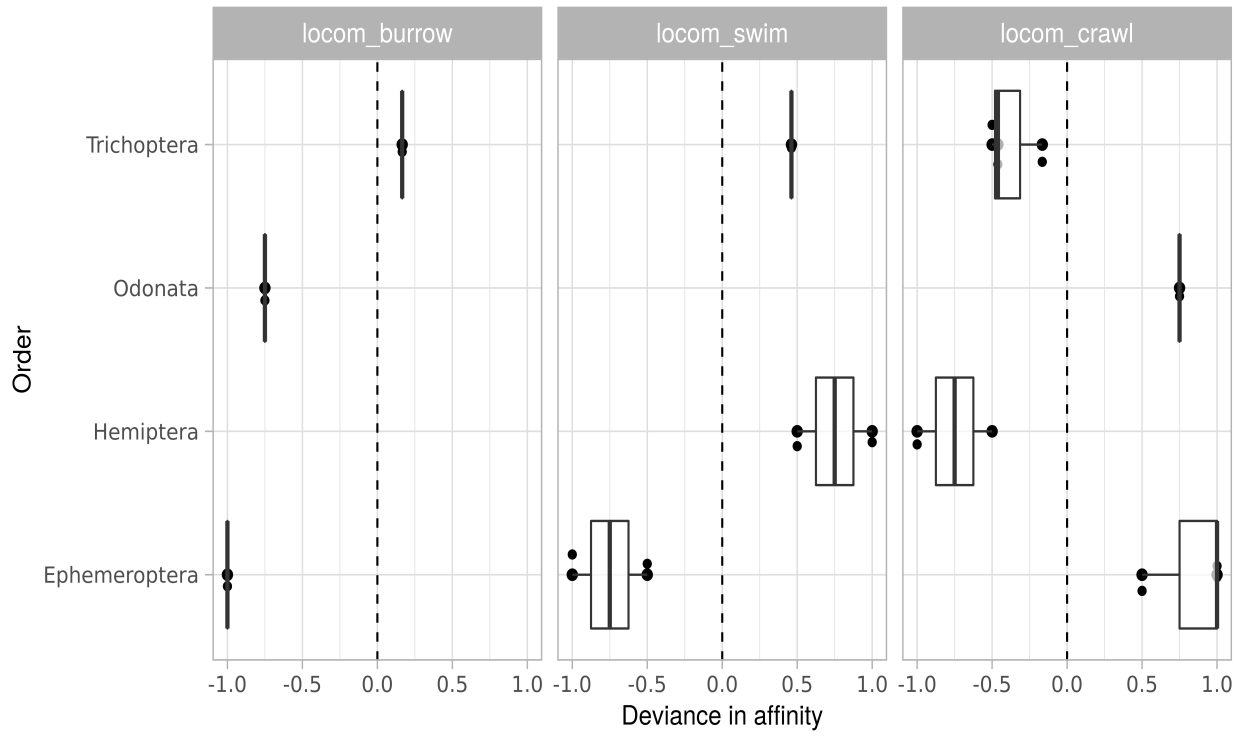
Figure 12: Affinity deviances for all deviating cases between the *weighted_agg* and traits assigned at family-level by Pyne for the grouping feature locomotion. See also section 4.2.2.
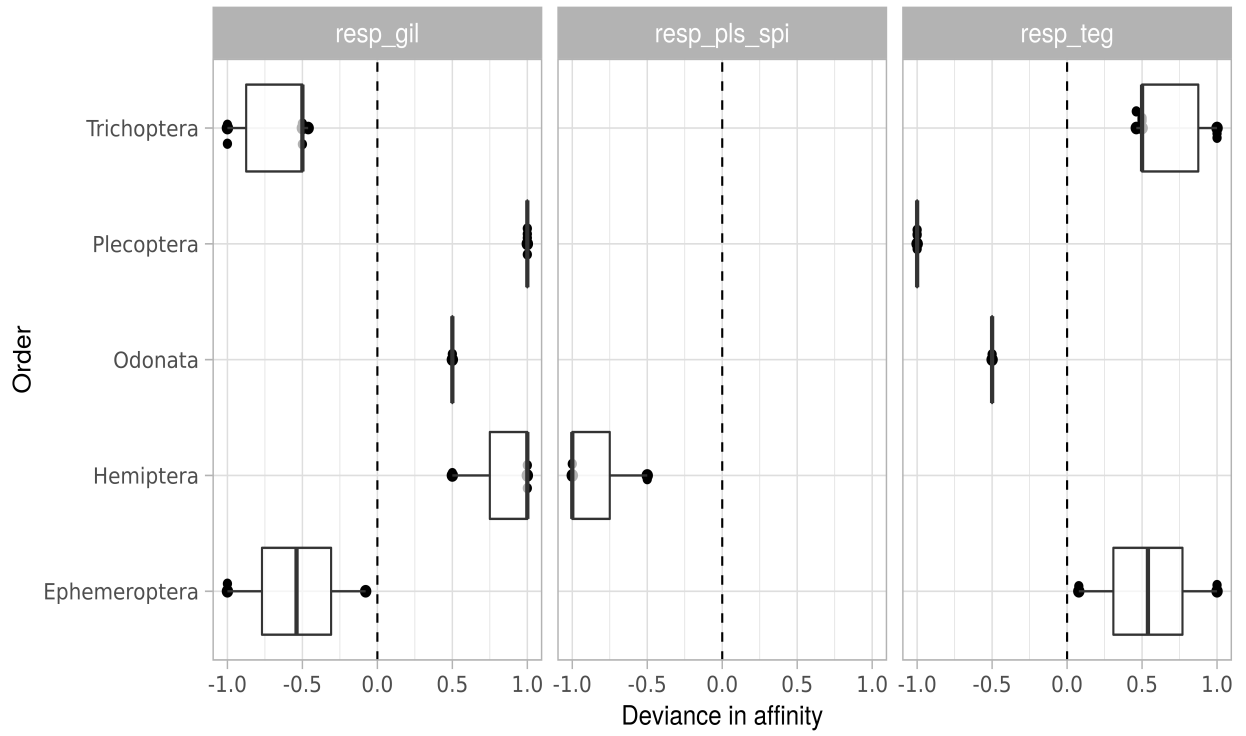
Figure 13: Affinity deviances for all deviating cases between the *weighted_agg* and traits assigned at family-level by Pyne for the grouping feature respiration. See also section 4.2.2.
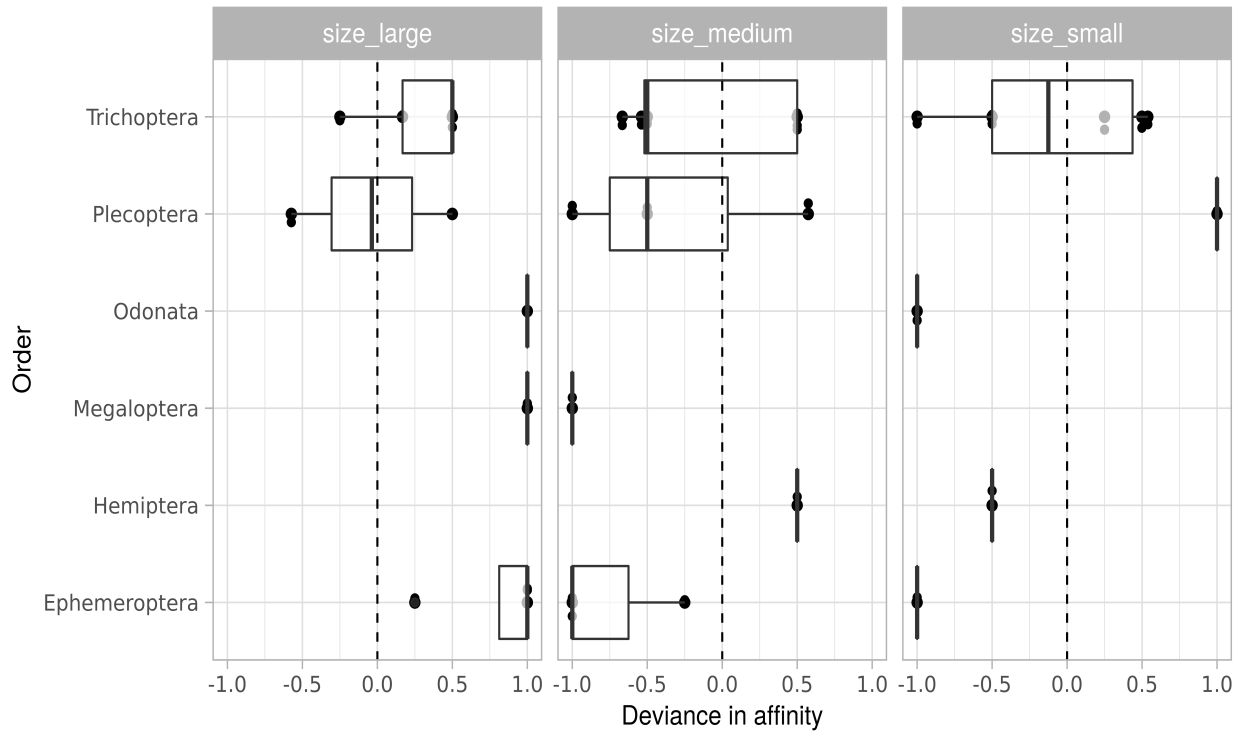
Figure 14: Affinity deviances for all deviating cases between the *weighted_agg* and traits assigned at family-level by Pyne for the grouping feature size. See also section 4.2.2.
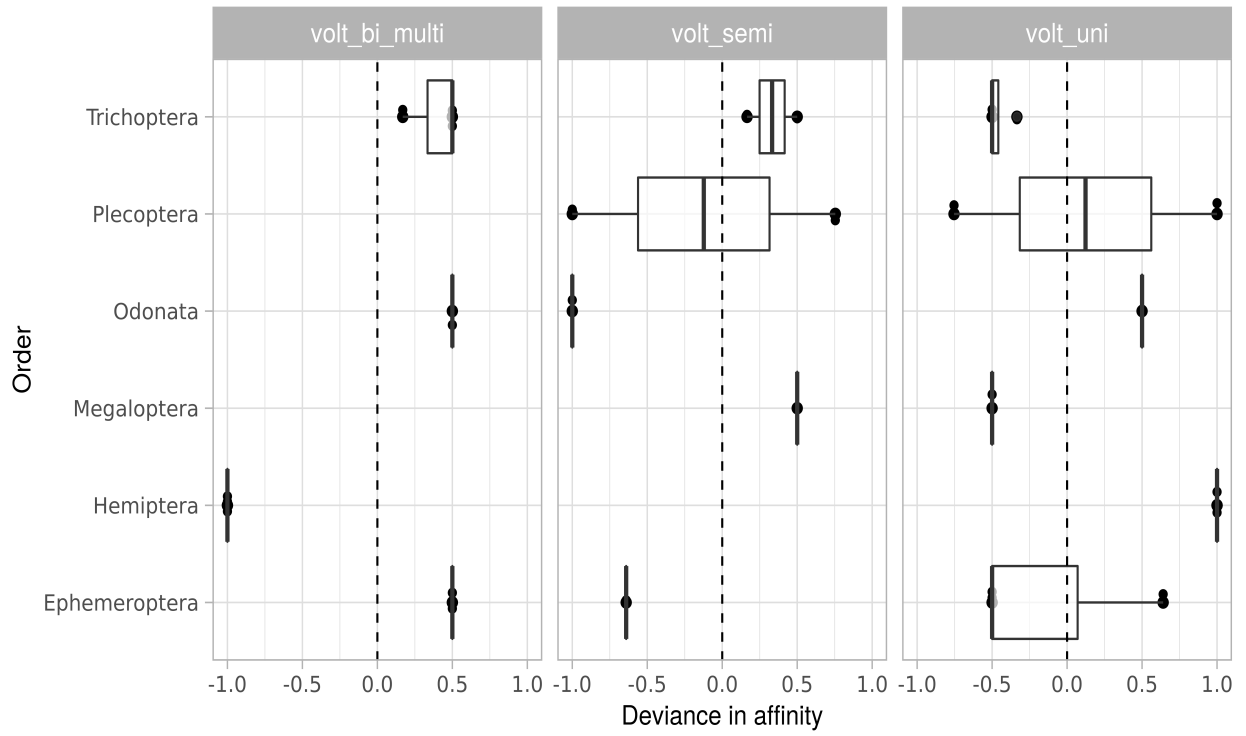
Figure 15: Affinity deviances for all deviating cases between the *weighted_agg* and traits assigned at family-level by Pyne for the grouping feature voltinism. See also section 4.2.2.

Harmonization of the European trait databases

Table 19: Representation of traits per grouping feature for the used trait databases. The color coding indicates traits that have been harmonized. Cyan colored traits have not been used because they either represented ambiguous traits or traits that were not compatible with the traits of the other databases. Harmonization was done by assigning the maximum affinity of the allocated traits for the respective taxa to the harmonized trait.

Grouping feature	Freshwater ecology	Tachet	Harmonized traits
Voltinism	Semivoltine	Semivoltine	Semivoltine
Voltinism	Univoltine	Univoltine	Univoltine
Voltinism	Bivoltine	Polyvoltine	Bi/Multivoltine
Voltinism	Trivoltine		
Voltinism	Multivoltine		
Voltinism	Flexible		
Feeding Mode	Shredder	Shredder	Shredder
Feeding Mode	Miner	Deposit-feeder	Gatherer
Feeding Mode	Xylophagus	Absorber	Filterer
Feeding Mode	Gatherer	Filter-feeder	Herbivore
Feeding Mode	Active filterer	Scraper	Predator
Feeding Mode	Passive filterer	Predator	Parasite
Feeding Mode	Grazer	Parasite	
Feeding Mode	Predator	Piercer (plants or animals)	
Feeding Mode	Parasite		
Feeding Mode	Other		
Locomotion	Swimming/scating	Surface swimmer	Swimmer
Locomotion	Swimming/diving	Full water swimmer	Burrower
Locomotion	Burrowing/boring	Burrower	Crawler
Locomotion	Sprawling/walking	Crawler	Sessil
Locomotion	(semi) sessil	Temporarily attachted	
Locomotion	Other	Permanently attached	
Locomotion		Interstitial	
Locomotion		Flier	
Respiration	Tegument	Tegument	Tegument
Respiration	Gill	Gill	Gills
Respiration	Plastron	Plastron	Plastron, spiracle
Respiration	Spiracle (aerial)	Spiracle (aerial)	
Respiration	Hydrostatic vesicle	Hydrostatic vesicle (aerial)	
Respiration	Tapping (air stores of aq. plants)		
Respiration	Excursion/Extension (to surface)		
Body size		<= 0.25cm	Small (< 1 cm)
Body size		> 0.25 - 0.5cm	Medium (>= 1cm - 2 cm)
Body size		> 0.5- 1cm	Large (>= 2 cm)
Body size		> 1 – 2 cm	

Body size		2 − 4 cm	
Body size		4 − 8 cm	
Body size		> 8 cm	
Reproduction	ovovivipar	ovoviviparity	ovoviviparity
Reproduction	free isolated eggs	isolated eggs, free	aquatic eggs
Reproduction	cemented isolated eggs	isolated eggs, cemented	terrestrial eggs
Reproduction	fixed clutches	clutches, cemented or fixed	
Reproduction	free clutches	clutches, free	
Reproduction	clutches in vegetation	clutches, in vegetation	
Reproduction	terrestrial clutches	clutches, terrestrial	
Reproduction	asexual	asexual reproduction	
Reproduction	parasitic		