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

The percentage of differing cases of trait affinities obtained by the trait aggregation methods compared to trait affinities originally assigned at family-level varied between 16.2 % and 22.9 % for the Australian dataset. For the North American dataset, comparison of the trait aggregation methods to trait affinities assigned at family-level yielded between 15.3 % and 47 % differing cases (Table 1).

In general, trait aggregation methods using the median yielded fewer cases with differences compared to approaches using the mean. However, aggregation methods using the median produced greater differences for both datasets. Standard deviations of absolute differences were similar for all tested aggregation methods. For both datasets maximum differences of 1 occurred for all investigated grouping features (Figure 1 and Figure 2).

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

Table 1: Amount of differing cases, the minimum and maximum, and means and standard deviations of absolute differences between trait affinities assigned at family-level and aggregated trait affinities.

| Database	Comparison to traits at fam.-lvl.	Differing cases [%]	Min. differences	Max. differences	Mean abs. differences	SD abs. differences
Australia (Chessman)	direct_agg (median)	16.53	0.01	1.00	0.45	0.27
	direct_agg (mean)	23.24	< 0.01	0.99	0.34	0.23
	stepwise_agg (median)	17.90	0.01	1.00	0.42	0.26
	stepwise_agg (mean)	23.24	< 0.01	0.99	0.33	0.22
	weighted_agg	23.24	< 0.01	1.00	0.34	0.24
North America	direct_agg (median)	15.33	0.17	1.00	0.70	0.26
	direct_agg (mean)	47.00	< 0.01	1.00	0.30	0.26
	stepwise_agg (median)	18.00	0.08	1.00	0.63	0.28
	stepwise_agg (mean)	47.00	< 0.01	1.00	0.30	0.27
	weighted_agg	47.00	< 0.01	1.00	0.31	0.28
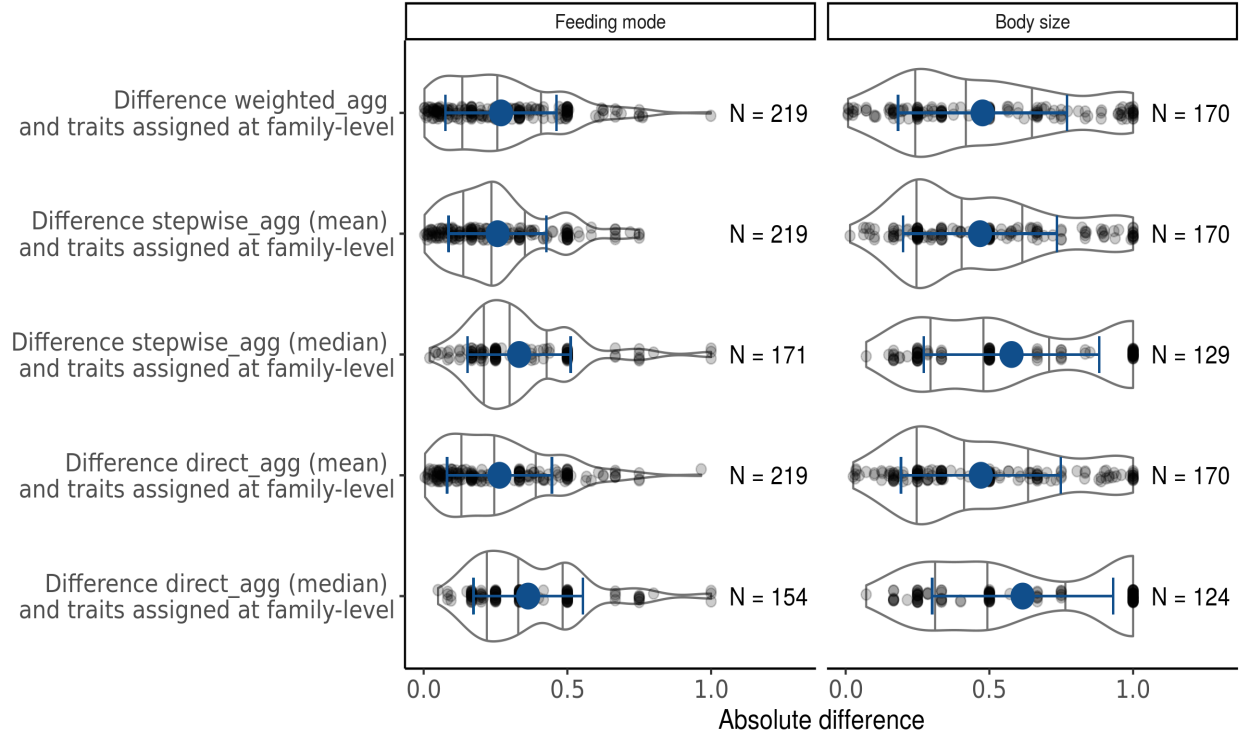
Figure 1: Display of the cases where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the Australian dataset for the grouping features feeding mode and body size. Absolute differences in trait affinities of the cases, which resulted in differences, are depicted as gray dots. N denotes the number of cases per comparison where differences occurred. The blue dot indicates the mean of absolute differences and the error bars the standard deviation. The gray vertical lines show the 25th, 50th and 75th quantile of the density estimate.
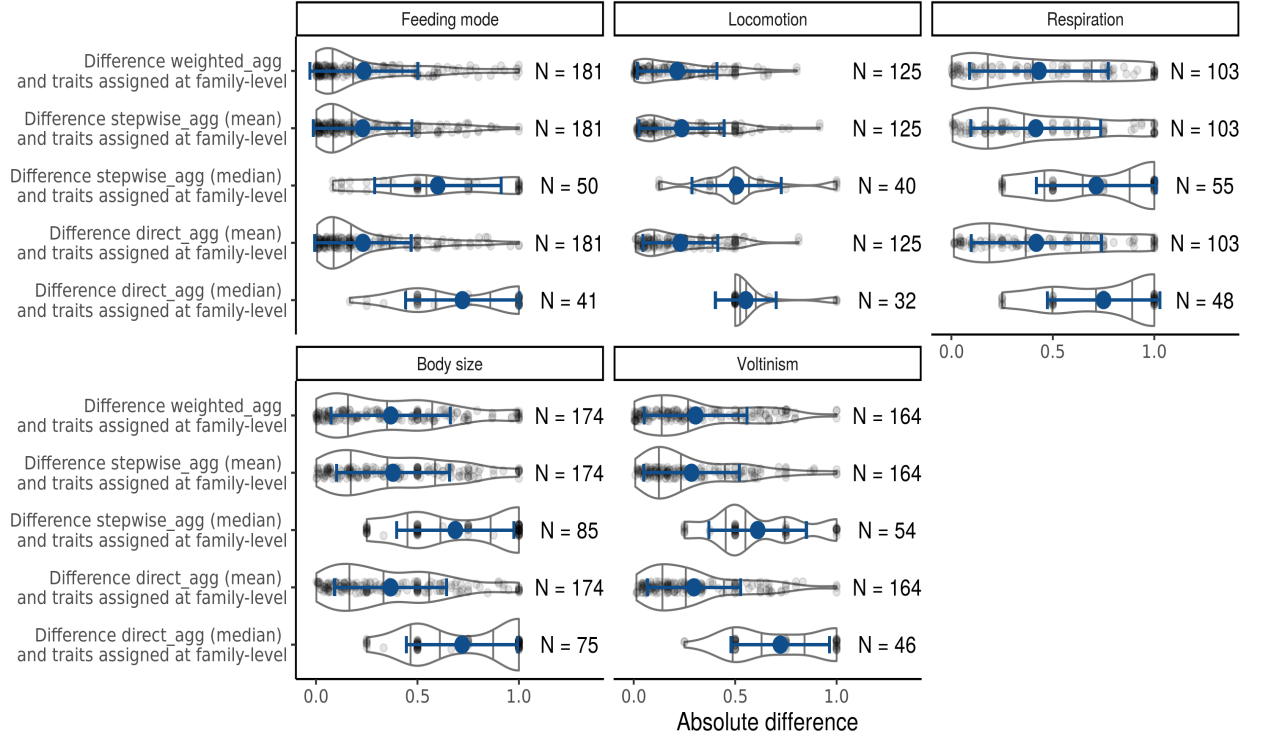
Figure 2: Display of the cases where differences occurred between aggregated traits and traits assigned at family-level. Violin plots - a mirrored density plot - show the density of the absolute trait affinity differences for the North America dataset for the grouping features feeding mode, locomotion, respiration, body size and voltinism. Absolute differences in trait affinities of the cases, which resulted in differences, are depicted as gray dots. N denotes the number of cases per comparison where differences occurred. The black dot indicates the mean of absolute differences and the error bars the standard deviation. The gray vertical lines show the 25th, 50th and 75th quantile of the density estimate.

Simulation of varying taxonomic hierarchies

For the *sim_base* scenario the trait aggregation methods based on the mean result in the same ranges of aggregated trait affinities within each level of trait variation. Aggregation methods using the median yielded greater ranges of aggregated trait affinities than methods using the mean. Thereby, the (stepwise_agg $_{median}$) yielded the greatest range of aggregated trait affinities (Figure S**??**). By contrast, the ranges of trait affinities differed for all aggregation methods for the simulation scenarios *sim_extreme* and *sim_variation*. *Weighted_agg* and *stepwise_agg$_{mean}$* produced a wider range of values than the *direct_agg$_{mean}$* in the *sim_extreme* scenario. For most levels of trait variability the *stepwise_agg $_{median}$* resulted in the largest range of trait affinities in both scenarios

Analysing the individual simulation runs showed that differences between the results of the

aggregation methods increased with increasing trait variability. Most differences occurred for the *sim_extreme* scenario. Differences greater than an absolute trait affinity of 0.1 occurred between methods using the median and the mean. Thereby, the most frequent differences above 0.1 were found between the aggregation methods *direct_agg$_{mean}$* and *stepwise_agg$_{median}$*, *direct_agg$_{median}$* and *stepwise_agg$_{median}$*, and *stepwise_agg$_{median}$* and *weighted_agg* (**??**).