

Additional file 1

Supporting analyses, full enumerations, and terminology

The choice of species representation in multi-species aquatic acute-toxicity prediction is constrained by data availability rather than by predictive accuracy

This file collects the material cited from the main text, in order of first citation: the terminology disambiguation; the water-solubility check; the complete run inventory; the decomposition of per-seed variance by processor; the naive-predictor results for the two taxonomy representations; the full gatekeeping enumeration and its six exploratory families; the ensemble sensitivity analysis; the species-axis variance-capture estimates; the species-by-chemical interaction analysis; the out-of-vocabulary re-aggregation; the concentration-fold interpretation; and the comparison of one-hot identity with the species mean offset. Every numeric value is taken from the audit and gatekeeping artefacts in the code repository (<https://github.com/Pluswick/species-representation-toxicity>). The data tables referenced below (`run_manifest.csv`, `excluded_records.csv`, and `gatekeeping_enumeration.csv`) accompany this file. Sections are numbered S1–S13; tables and figures are numbered independently of the main text; the bibliography is independent of the main text.

The audit tables use the pre-registration identifiers for the species representations, which correspond to the last column of main-text Table 2. The identifiers, partitions, splits, and codes used throughout are listed here.

Species representations (last column of main-text Table 2)

t0	no species information
t1	species mean offset
t1'	out-of-fold residual calibration
t2	one-hot identity
t3a	source-authority taxonomy
t3b	NCBI taxonomy
t4	learned embedding
t3a_g	source-authority taxonomy, genus rank only
t3a_gf	source-authority taxonomy, genus and family ranks
shuf	permuted (permute-and-relearn) control

Datasets, partitions, splits, backbones, and codes

discovery	single-condition partition (LC ₅₀ at 96 h)
replication	seven-condition partition (LC ₅₀ /EC ₅₀ , 24–96 h)
b1, runs_b1	ECOTOX-external set
group	compound-registry (CASRN) split
scaffold	Bemis–Mureko scaffold split
scaffold_generic	generic scaffold split
designed_leaky	deliberately leaky split (leakage positive control)
dmpnn, graphconv	D-MPNN, GraphConv
lightgbm	LightGBM
path3	individual-inspection exclusion path (source or literature error)

Contents

S1	Terminology disambiguation	3
S2	Water-solubility check	3
S3	Complete run inventory	3
S4	Decomposition of per-seed variance by processor	5
S5	Naive predictor with taxonomy	5
S6	Full gatekeeping enumeration	5
S7	Exploratory families	12
S8	Ensemble sensitivity analysis	14
S9	Species-axis variance capture	15
S10	Species-by-chemical interaction	15
S11	Out-of-vocabulary re-aggregation	16
S12	Concentration-fold interpretation	17
S13	One-hot identity versus the species mean offset	17

S1. Terminology disambiguation

Cited from Section 2.3 of the main text. The three terms below are used in conflicting senses across the studies we surveyed; we state the sense we adopt.

What “cross-species” denotes. The term is used in at least five different senses. Zubrod et al. use it as a claim that trait surrogates permit prediction for new species, under a random split [1]; Forastiere et al. evaluate genuine held-out species [2]; Viljanen et al. mean unobserved (species, chemical) pairs [3]; Schlender et al. mean pooling many species in one model [4]; and Yuan et al. use the term in the title while splitting by compound registry [5]. We reserve *cross-species* for evaluation on species absent from training.

“Species embedding” denotes three different objects: a latent factor learned from the toxicity matrix itself [3]; a vector computed from external data without reference to the response [2, 1]; and the output of a learned encoder whose *input* is an external descriptor [6, 5]. We distinguish them as *learned-from-target*, *derived-without-learning*, and *learned-from-descriptor*.

“Ablation” does not always mean removal of species information: the ablation condition in Yuan et al. is a one-hot species identity [5], so the comparison it anchors is between degrees of structure, not presence versus absence. Conditions that remove species information entirely do exist – Wu et al. quantify the gain of taxonomic and experimental information over chemistry alone [7], and Yang et al. report a chemistry-and-exposure baseline [6] – but they are not the same comparison.

S2. Water-solubility check

Cited from Sections 3.1 and 5.1. We did not apply a water-solubility filter; this section reports the check of the ten compounds with the widest concentration spread, including the DDT case. The pre-registered screening log (the fourteen individually judged records) follows.

Table S1: Solubility check for the ten compounds with the widest concentration spread. Aqueous solubility is web-verifiable for eight; records exceeded it for five. The largest excess is a factor of 2.97×10^5 (DDT). Source abbreviations as recorded: PubChem, HSDB, PPDB, HMDB, e-Pest Manual.

Compound	CAS	Solub. (mg/L)	Source	Exceeds?	Factor	Rows over
Cypermethrin	52315-07-8	4.00e-03	PubChem/e-Pest Manual 4e-3 mg/L @20C	yes	4.71e+04	69
Methamidophos	10265-92-6	2.00e+05	PPDB 200000 mg/L @20C	no	1.04e-03	0
Deltamethrin	52918-63-5	2.00e-03	PubChem/HSDB <0.002 mg/L @25C	yes	3.50e+04	57
lambda-Cyhalothrin	91465-08-6	5.00e-03	PPDB 0.005 mg/L @20C	yes	3.85e+03	4
Trichlorfon	52-68-6	1.20e+05	PubChem/HSDB 1.2e5 mg/L @25C	no	4.88e-03	0
Copper(II) chloride	7447-39-4	–	salt, freely soluble; no 20-25C number	–	–	–
p,p'-DDT	50-29-3	5.50e-03	PubChem/HMDB 5.5e-6 mg/mL @25C	yes	2.97e+05	202
Nickel(metal)	7440-02-0	–	elemental metal, insoluble; N/A (salt-dependent)	–	–	–
Mercuric chloride	7487-94-7	6.90e+04	PubChem/HSDB 69 g/L @20C	no	6.29e-04	0
Fenvalerate	51630-58-1	0.024	PubChem/HSDB 0.024 mg/L @22C seawater	yes	6.14e+03	29

S3. Complete run inventory

Cited from Section 3.4. The 6,286 prediction files behind the reported statistics are enumerated by model group below and listed file-by-file in `run_manifest.csv`. Each file name encodes the represen-

Table S2: Screening log: the fourteen records excluded by the pre-registered screening rule. All are from the seven-condition partition: none is LC₅₀ at 96 h (durations 24–72 h), so none belongs to the single-condition partition (Section 3.1). The lower concentration bound ($-\log_{10}(\text{mol/L}) > -1.75$) excluded no records here; all fourteen were removed by the mass rule (ten) or on individual inspection as source or literature errors (four). Mass in ng/L (two significant figures); pLC₅₀ = $-\log_{10}(\text{mol/L})$. The raw log, with full-precision values, is provided as `excluded_records.csv`; the *Reason* column expands the `exclusion_reason` codes of that file (Table S3).

Result id	CAS	Group	Endpoint	Dur. (h)	Mass (ng/L)	pLC ₅₀	Reason
77418	72-20-8	crusta	LC50	48	0.28	12.13	source / literature mismatch (path 3)
67764	298-00-0	crusta	LC50	24	0.14	12.27	source / literature mismatch (path 3)
77425	72-20-8	crusta	LC50	72	0.19	12.30	source / literature mismatch (path 3)
77419	72-20-8	crusta	LC50	72	0.13	12.47	source / literature mismatch (path 3)
77325	72-20-8	crusta	LC50	72	0.071	12.73	mass concentration below 0.1 ng/L
77320	72-20-8	crusta	LC50	24	0.057	12.82	mass concentration below 0.1 ng/L
77321	72-20-8	crusta	LC50	48	0.022	13.24	mass concentration below 0.1 ng/L
77322	72-20-8	crusta	LC50	72	0.0088	13.64	mass concentration below 0.1 ng/L
77420	72-20-8	crusta	LC50	24	0.0029	14.12	mass concentration below 0.1 ng/L
2034949	688-73-3	crusta	EC50	48	0.001	14.46	mass concentration below 0.1 ng/L
77421	72-20-8	crusta	LC50	48	0.0011	14.54	mass concentration below 0.1 ng/L
77422	72-20-8	crusta	LC50	72	0.00061	14.80	mass concentration below 0.1 ng/L
2137595	2921-88-2	crusta	LC50	24	7.3e-05	15.68	mass concentration below 0.1 ng/L
117362	121-75-5	crusta	EC50	24	2.7e-09	20.09	mass concentration below 0.1 ng/L

Table S3: Exclusion-reason codes used in `excluded_records.csv`, with the human-readable expansion shown in Table S2.

<code>exclusion_reason</code> code	Reason	<i>n</i>
<code>excluded:below_detection_limit(<0.1ng/L)</code>	mass concentration below 0.1 ng/L	10
<code>excluded:source_error_literature_mismatch(path3)</code>	source / literature mismatch (path 3)	4

tation, control arm, dataset, split, and seed.

Table S4: Complete run inventory: the 6,286 per-run prediction files behind the reported statistics, by model group. Each file name encodes representation, control, dataset, split, and seed; the full per-file listing is provided as `run_manifest.csv`. The ECOTOX-external runs live under a parallel `runs_b1/` tree.

Model group	Prediction files
gnn (D-MPNN/GraphConv, 10 seeds)	4880
gnn ensemble (90 seeds)	1260
LightGBM	74
naive baselines	64
LightGBM species-SVD	8
Total	6286

S4. Decomposition of per-seed variance by processor

Cited from Section 3.6. The seed-to-processor assignment was fixed in the pre-registration; the per-seed variance was decomposed by processor after unblinding.

Table S5: Decomposition of per-seed variance by processor. The ten seeds were run across two GPUs under the pre-registered card assignment (Pre-registration Section A-5; scheduler-arbitrary placement prohibited): ADORE seeds 0–6 on an RTX 5060 Ti and 7–9 on an RTX 4090; ECOTOX-external seeds 0–4 on the RTX 4090 and 5–9 on the RTX 5060 Ti. The between-processor share is the pooled SSB/SST of the per-seed RMSE across the 14 primary cells. The share is small, so δ measures re-run noise rather than a hardware effect. This decomposition was specified after unblinding.

Margin	Seed $\rightarrow$ processor map	Pooled share of per-seed variance	Mean	Median	δ (full)
ADORE δ	5060Ti(0-6) vs 4090(7-9)	4.5%	5.3%	3.9%	0.0198
ECOTOX-external δ	4090(0-4) vs 5060Ti(5-9)	6.2%	6.4%	2.7%	0.0356
ECOTOX-external δ'	4090 ens(1-4) vs 5060Ti ens(6-9); ens0,5 device-mixed (excluded)	2.7%	5.5%	3.5%	0.0160
ADORE δ' : every ensemble device-mixed (card rule splits each decade across both GPUs); no homogeneous subset; decomposition N/A					

S5. Naive predictor with taxonomy

Cited from Table 2, footnote c: the naive-predictor results for the two taxonomy representations, evaluated but reported only here.

S6. Full gatekeeping enumeration

Cited from Table 7. Every comparison behind the verdict counts is listed in Table S7 (ADORE 225 comparisons; ECOTOX-external 189; the external set has no confirmatory family). The six exploratory sub-families (Section S7) appear in the same table under their *Subtype* label, as do the deterministic and ensemble-sensitivity families, which are judged against different margins. The complete table is also provided as `gatekeeping_enumeration.csv`. Individual deterministic equivalence intervals are broken out in Table S8.

Table S6: Naive predictor with taxonomic species representations (Table 2, footnote c). The naive baseline predicts a hierarchical taxon-group mean with back-off over four ranks; RMSE on the $\log_{10}$ scale, whole test set, one seed. Only the untouched (non-permuted) arms are shown.

Dataset	Split	Representation	RMSE	n
ADORE	single-condition, compound-registry split	NCBI taxonomy	1.7488	2029
ADORE	single-condition, compound-registry split	source-authority taxonomy	1.7561	2029
ADORE	single-condition, Bemis–Murcko scaffold split	NCBI taxonomy	1.6545	2029
ADORE	single-condition, Bemis–Murcko scaffold split	source-authority taxonomy	1.6585	2029
ADORE	single-condition, generic scaffold split	NCBI taxonomy	1.6874	2029
ADORE	single-condition, generic scaffold split	source-authority taxonomy	1.6908	2029
ADORE	single-condition, leaky split	NCBI taxonomy	1.6720	2029
ADORE	single-condition, leaky split	source-authority taxonomy	1.6698	2029
ADORE	seven-condition, compound-registry split	NCBI taxonomy	1.6014	4351
ADORE	seven-condition, compound-registry split	source-authority taxonomy	1.6066	4351
ADORE	seven-condition, Bemis–Murcko scaffold split	NCBI taxonomy	2.0303	4351
ADORE	seven-condition, Bemis–Murcko scaffold split	source-authority taxonomy	2.0355	4351
ADORE	seven-condition, generic scaffold split	NCBI taxonomy	1.8248	4351
ADORE	seven-condition, generic scaffold split	source-authority taxonomy	1.8287	4351
ADORE	seven-condition, leaky split	NCBI taxonomy	1.6295	4310
ADORE	seven-condition, leaky split	source-authority taxonomy	1.6339	4310
ECOTOX-external	ECOTOX-external, compound-registry split	NCBI taxonomy	1.5748	4561
ECOTOX-external	ECOTOX-external, compound-registry split	source-authority taxonomy	1.5787	4561
ECOTOX-external	ECOTOX-external, Bemis–Murcko scaffold split	NCBI taxonomy	1.7360	4561
ECOTOX-external	ECOTOX-external, Bemis–Murcko scaffold split	source-authority taxonomy	1.7126	4561
ECOTOX-external	ECOTOX-external, generic scaffold split	NCBI taxonomy	1.6935	4561
ECOTOX-external	ECOTOX-external, generic scaffold split	source-authority taxonomy	1.6687	4561
ECOTOX-external	ECOTOX-external, leaky split	NCBI taxonomy	1.6540	4579
ECOTOX-external	ECOTOX-external, leaky split	source-authority taxonomy	1.6584	4579

Table S7: Full gatekeeping enumeration: every pre-registered and exploratory comparison behind the verdict counts of Table 7. Effect size dd (difference-in-differences in RMSE on the $\log_{10}$ scale), 90% confidence interval, Benjamini–Hochberg q within family, and verdict (*significant* for superiority tests, *different* for two one-sided tests, both meaning the interval excludes zero). The *Split* column distinguishes the compound-registry (group), Bemis–Murcko and generic scaffold, and leaky schemes; the discovery/replication partition is carried by the primary/confirmatory family label. Tier codes: t0 no species, t1 mean offset, t1' OOF calibration, t2 one-hot, t3a source-authority taxonomy, t3b NCBI taxonomy, t4 learned embedding; t3a_g / t3a_gf are genus- and genus-family rank truncations. Deterministic rows are judged against δ_{det} and ensemble rows against δ' ; all others against δ .

Family	Subtype	Dataset	Split	Comparison	dd	90% CI	q	Verdict
primary	–	ADORE	group	dmpnn/t1 > shuf	-0.077	[-0.113, -0.042]	0.000	significant
primary	–	ADORE	group	dmpnn/t1 > t0	-0.068	[-0.103, -0.035]	0.000	significant
primary	–	ADORE	group	dmpnn/t1' > shuf	-0.273	[-0.350, -0.210]	0.000	significant
primary	–	ADORE	group	dmpnn/t1' > t0	-0.052	[-0.085, -0.018]	0.010	significant
primary	–	ADORE	group	dmpnn/t2 > shuf	-0.128	[-0.177, -0.074]	0.000	significant
primary	–	ADORE	group	dmpnn/t2 > t0	-0.076	[-0.124, -0.025]	0.012	significant
primary	–	ADORE	group	dmpnn/t3a > shuf	-0.146	[-0.204, -0.083]	0.000	significant
primary	–	ADORE	group	dmpnn/t3a > t0	-0.094	[-0.157, -0.032]	0.011	significant
primary	–	ADORE	group	dmpnn/t3a~t2	-0.018	[-0.045, +0.009]	0.317	indeterminate
primary	–	ADORE	group	dmpnn/t3b > shuf	-0.134	[-0.198, -0.069]	0.002	significant
primary	–	ADORE	group	dmpnn/t3b > t0	-0.093	[-0.158, -0.028]	0.014	significant
primary	–	ADORE	group	dmpnn/t3b~t2	-0.017	[-0.046, +0.014]	0.369	indeterminate
primary	–	ADORE	group	dmpnn/t4 > shuf	-0.104	[-0.156, -0.049]	0.002	significant
primary	–	ADORE	group	dmpnn/t4 > t0	-0.065	[-0.116, -0.010]	0.050	significant
primary	–	ADORE	group	dmpnn/t4~t2	0.012	[-0.013, +0.034]	0.451	indeterminate
primary	–	ADORE	group	graphconv/t1 > shuf	-0.048	[-0.075, -0.018]	0.008	significant
primary	–	ADORE	group	graphconv/t1 > t0	-0.050	[-0.077, -0.020]	0.009	significant
primary	–	ADORE	group	graphconv/t1' > shuf	-0.288	[-0.364, -0.223]	0.000	significant
primary	–	ADORE	group	graphconv/t1' > t0	-0.052	[-0.087, -0.016]	0.013	significant
primary	–	ADORE	group	graphconv/t2 > shuf	-0.106	[-0.156, -0.049]	0.002	significant
primary	–	ADORE	group	graphconv/t2 > t0	-0.067	[-0.114, -0.018]	0.039	significant
primary	–	ADORE	group	graphconv/t3a > shuf	-0.118	[-0.175, -0.061]	0.000	significant
primary	–	ADORE	group	graphconv/t3a > t0	-0.101	[-0.159, -0.044]	0.010	significant
primary	–	ADORE	group	graphconv/t3a~t2	-0.033	[-0.069, +0.002]	0.133	indeterminate

Family	Subtype	Dataset	Split	Comparison	dd	90% CI	q	Verdict
primary	-	ADORE	group	graphconv/t3b > shuf	-0.127	[-0.181, -0.071]	0.000	significant
primary	-	ADORE	group	graphconv/t3b > t0	-0.110	[-0.169, -0.052]	0.002	significant
primary	-	ADORE	group	graphconv/t3b~t2	-0.043	[-0.079, -0.009]	0.060	different
primary	-	ADORE	group	graphconv/t4 > shuf	-0.113	[-0.155, -0.066]	0.000	significant
primary	-	ADORE	group	graphconv/t4 > t0	-0.082	[-0.126, -0.037]	0.004	significant
primary	-	ADORE	group	graphconv/t4~t2	-0.015	[-0.048, +0.016]	0.397	indeterminate
primary	-	ext	group	dmpnn/t1 > shuf	-0.056	[-0.110, +0.001]	0.127	indeterminate
primary	-	ext	group	dmpnn/t1 > t0	-0.088	[-0.124, -0.052]	0.000	significant
primary	-	ext	group	dmpnn/t1' > shuf	-0.402	[-0.471, -0.333]	0.000	significant
primary	-	ext	group	dmpnn/t1' > t0	-0.109	[-0.143, -0.075]	0.000	significant
primary	-	ext	group	dmpnn/t2 > shuf	-0.134	[-0.205, -0.060]	0.007	significant
primary	-	ext	group	dmpnn/t2 > t0	-0.150	[-0.211, -0.092]	0.000	significant
primary	-	ext	group	dmpnn/t3a > shuf	-0.118	[-0.174, -0.065]	0.000	significant
primary	-	ext	group	dmpnn/t3a > t0	-0.158	[-0.225, -0.090]	0.000	significant
primary	-	ext	group	dmpnn/t3a~t2	-0.008	[-0.067, +0.047]	0.924	indeterminate
primary	-	ext	group	dmpnn/t3b > shuf	-0.110	[-0.163, -0.057]	0.002	significant
primary	-	ext	group	dmpnn/t3b > t0	-0.148	[-0.221, -0.081]	0.000	significant
primary	-	ext	group	dmpnn/t3b~t2	0.002	[-0.055, +0.053]	0.924	indeterminate
primary	-	ext	group	dmpnn/t4 > shuf	-0.104	[-0.153, -0.057]	0.000	significant
primary	-	ext	group	dmpnn/t4 > t0	-0.151	[-0.216, -0.089]	0.000	significant
primary	-	ext	group	dmpnn/t4~t2	-0.000	[-0.048, +0.040]	0.926	indeterminate
primary	-	ext	group	graphconv/t1 > shuf	-0.080	[-0.125, -0.033]	0.020	significant
primary	-	ext	group	graphconv/t1 > t0	-0.119	[-0.150, -0.088]	0.000	significant
primary	-	ext	group	graphconv/t1' > shuf	-0.403	[-0.472, -0.334]	0.000	significant
primary	-	ext	group	graphconv/t1' > t0	-0.104	[-0.137, -0.073]	0.000	significant
primary	-	ext	group	graphconv/t2 > shuf	-0.158	[-0.231, -0.084]	0.002	significant
primary	-	ext	group	graphconv/t2 > t0	-0.195	[-0.262, -0.128]	0.000	significant
primary	-	ext	group	graphconv/t3a > shuf	-0.116	[-0.176, -0.061]	0.002	significant
primary	-	ext	group	graphconv/t3a > t0	-0.186	[-0.267, -0.114]	0.000	significant
primary	-	ext	group	graphconv/t3a~t2	0.009	[-0.043, +0.054]	0.811	indeterminate
primary	-	ext	group	graphconv/t3b > shuf	-0.102	[-0.166, -0.045]	0.015	significant
primary	-	ext	group	graphconv/t3b > t0	-0.162	[-0.224, -0.101]	0.000	significant
primary	-	ext	group	graphconv/t3b~t2	0.032	[-0.017, +0.078]	0.302	indeterminate
primary	-	ext	group	graphconv/t4 > shuf	-0.093	[-0.139, -0.049]	0.003	significant
primary	-	ext	group	graphconv/t4 > t0	-0.151	[-0.205, -0.097]	0.000	significant
primary	-	ext	group	graphconv/t4~t2	0.044	[+0.009, +0.080]	0.045	different
confirmatory	-	ADORE	group	dmpnn/t1 > shuf	-0.056	[-0.080, -0.029]	0.004	significant
confirmatory	-	ADORE	group	dmpnn/t1 > t0	-0.064	[-0.097, -0.031]	0.005	significant
confirmatory	-	ADORE	group	dmpnn/t1' > shuf	-0.223	[-0.290, -0.162]	0.000	significant
confirmatory	-	ADORE	group	dmpnn/t1' > t0	-0.024	[-0.053, +0.003]	0.188	indeterminate
confirmatory	-	ADORE	group	dmpnn/t2 > shuf	-0.088	[-0.174, +0.006]	0.154	indeterminate
confirmatory	-	ADORE	group	dmpnn/t2 > t0	-0.151	[-0.200, -0.101]	0.000	significant
confirmatory	-	ADORE	group	dmpnn/t3a > shuf	-0.122	[-0.174, -0.072]	0.000	significant
confirmatory	-	ADORE	group	dmpnn/t3a > t0	-0.158	[-0.206, -0.105]	0.000	significant
confirmatory	-	ADORE	group	dmpnn/t3a~t2	-0.007	[-0.035, +0.022]	0.724	indeterminate
confirmatory	-	ADORE	group	dmpnn/t3b > shuf	-0.122	[-0.174, -0.072]	0.000	significant
confirmatory	-	ADORE	group	dmpnn/t3b > t0	-0.161	[-0.215, -0.103]	0.000	significant
confirmatory	-	ADORE	group	dmpnn/t3b~t2	-0.011	[-0.038, +0.020]	0.587	indeterminate
confirmatory	-	ADORE	group	dmpnn/t4 > shuf	-0.088	[-0.139, -0.037]	0.016	significant
confirmatory	-	ADORE	group	dmpnn/t4 > t0	-0.141	[-0.193, -0.085]	0.000	significant
confirmatory	-	ADORE	group	dmpnn/t4~t2	0.009	[-0.023, +0.037]	0.525	indeterminate
confirmatory	-	ADORE	group	graphconv/t1 > shuf	-0.052	[-0.075, -0.027]	0.008	significant
confirmatory	-	ADORE	group	graphconv/t1 > t0	-0.054	[-0.080, -0.028]	0.002	significant
confirmatory	-	ADORE	group	graphconv/t1' > shuf	-0.207	[-0.272, -0.148]	0.000	significant
confirmatory	-	ADORE	group	graphconv/t1' > t0	-0.018	[-0.049, +0.011]	0.308	indeterminate
confirmatory	-	ADORE	group	graphconv/t2 > shuf	-0.051	[-0.129, +0.027]	0.295	indeterminate
confirmatory	-	ADORE	group	graphconv/t2 > t0	-0.134	[-0.196, -0.072]	0.000	significant
confirmatory	-	ADORE	group	graphconv/t3a > shuf	-0.112	[-0.164, -0.059]	0.000	significant
confirmatory	-	ADORE	group	graphconv/t3a > t0	-0.173	[-0.238, -0.106]	0.000	significant
confirmatory	-	ADORE	group	graphconv/t3a~t2	-0.040	[-0.075, -0.004]	0.090	different
confirmatory	-	ADORE	group	graphconv/t3b > shuf	-0.130	[-0.177, -0.081]	0.000	significant
confirmatory	-	ADORE	group	graphconv/t3b > t0	-0.187	[-0.256, -0.111]	0.000	significant
confirmatory	-	ADORE	group	graphconv/t3b~t2	-0.053	[-0.091, -0.014]	0.029	different
confirmatory	-	ADORE	group	graphconv/t4 > shuf	-0.085	[-0.133, -0.029]	0.028	significant
confirmatory	-	ADORE	group	graphconv/t4 > t0	-0.160	[-0.224, -0.098]	0.000	significant
confirmatory	-	ADORE	group	graphconv/t4~t2	-0.026	[-0.058, +0.003]	0.188	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t3a~t2/bin1-5	-0.006	[-0.096, +0.070]	0.926	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t3a~t2/ bin100+	-0.003	[-0.022, +0.016]	0.858	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t3a~t2/ bin21-100	-0.045	[-0.077, -0.011]	0.067	different
exploratory	support bin	ADORE	group	dmpnn/t3a~t2/ bin6-20	-0.042	[-0.102, +0.027]	0.469	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t3b~t2/bin1-5	-0.005	[-0.090, +0.071]	0.926	indeterminate

Family	Subtype	Dataset	Split	Comparison	dd	90% CI	q	Verdict
exploratory	support bin	ADORE	group	dmpnn/t3b~t2/ bin100+	0.002	[-0.017, +0.021]	0.923	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t3b~t2/ bin21-100	-0.053	[-0.090, -0.015]	0.052	different
exploratory	support bin	ADORE	group	dmpnn/t3b~t2/ bin6-20	-0.038	[-0.106, +0.040]	0.577	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t4~t2/bin1-5	0.029	[-0.039, +0.084]	0.611	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t4~t2/bin100+	0.009	[-0.010, +0.026]	0.600	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t4~t2/ bin21-100	-0.015	[-0.051, +0.021]	0.634	indeterminate
exploratory	support bin	ADORE	group	dmpnn/t4~t2/bin6-20	0.005	[-0.045, +0.053]	0.933	indeterminate
exploratory	support bin	ADORE	group	graphconv/t3a~t2/ bin1-5	-0.007	[-0.106, +0.081]	0.926	indeterminate
exploratory	support bin	ADORE	group	graphconv/t3a~t2/ bin100+	-0.027	[-0.051, -0.004]	0.122	different
exploratory	support bin	ADORE	group	graphconv/t3a~t2/ bin21-100	-0.066	[-0.118, -0.018]	0.073	different
exploratory	support bin	ADORE	group	graphconv/t3a~t2/ bin6-20	-0.030	[-0.101, +0.042]	0.655	indeterminate
exploratory	support bin	ADORE	group	graphconv/t3b~t2/ bin1-5	-0.011	[-0.107, +0.076]	0.902	indeterminate
exploratory	support bin	ADORE	group	graphconv/t3b~t2/ bin100+	-0.034	[-0.061, -0.007]	0.077	different
exploratory	support bin	ADORE	group	graphconv/t3b~t2/ bin21-100	-0.073	[-0.122, -0.027]	0.022	different
exploratory	support bin	ADORE	group	graphconv/t3b~t2/ bin6-20	-0.060	[-0.128, +0.012]	0.257	indeterminate
exploratory	support bin	ADORE	group	graphconv/t4~t2/ bin1-5	-0.024	[-0.088, +0.032]	0.622	indeterminate
exploratory	support bin	ADORE	group	graphconv/t4~t2/ bin100+	-0.022	[-0.054, +0.008]	0.390	indeterminate
exploratory	support bin	ADORE	group	graphconv/t4~t2/ bin21-100	-0.035	[-0.080, +0.011]	0.332	indeterminate
exploratory	support bin	ADORE	group	graphconv/t4~t2/ bin6-20	0.012	[-0.041, +0.063]	0.827	indeterminate
exploratory	support bin	ext	group	dmpnn/t3a~t2/bin1-5	-0.044	[-0.133, +0.044]	0.571	indeterminate
exploratory	support bin	ext	group	dmpnn/t3a~t2/ bin100+	-0.008	[-0.047, +0.029]	0.805	indeterminate
exploratory	support bin	ext	group	dmpnn/t3a~t2/ bin21-100	-0.020	[-0.075, +0.034]	0.683	indeterminate
exploratory	support bin	ext	group	dmpnn/t3a~t2/ bin6-20	-0.023	[-0.123, +0.080]	0.891	indeterminate
exploratory	support bin	ext	group	dmpnn/t3b~t2/bin1-5	-0.042	[-0.115, +0.035]	0.560	indeterminate
exploratory	support bin	ext	group	dmpnn/t3b~t2/ bin100+	-0.004	[-0.038, +0.029]	0.929	indeterminate
exploratory	support bin	ext	group	dmpnn/t3b~t2/ bin21-100	-0.019	[-0.080, +0.037]	0.709	indeterminate
exploratory	support bin	ext	group	dmpnn/t3b~t2/ bin6-20	0.011	[-0.086, +0.106]	0.842	indeterminate
exploratory	support bin	ext	group	dmpnn/t4~t2/bin1-5	0.019	[-0.038, +0.078]	0.683	indeterminate
exploratory	support bin	ext	group	dmpnn/t4~t2/bin100+	-0.018	[-0.047, +0.010]	0.466	indeterminate
exploratory	support bin	ext	group	dmpnn/t4~t2/ bin21-100	-0.021	[-0.066, +0.024]	0.603	indeterminate
exploratory	support bin	ext	group	dmpnn/t4~t2/bin6-20	-0.028	[-0.106, +0.043]	0.738	indeterminate
exploratory	support bin	ext	group	graphconv/t3a~t2/ bin1-5	-0.007	[-0.087, +0.074]	0.932	indeterminate
exploratory	support bin	ext	group	graphconv/t3a~t2/ bin100+	-0.023	[-0.064, +0.015]	0.491	indeterminate
exploratory	support bin	ext	group	graphconv/t3a~t2/ bin21-100	-0.025	[-0.079, +0.022]	0.578	indeterminate
exploratory	support bin	ext	group	graphconv/t3a~t2/ bin6-20	0.027	[-0.048, +0.100]	0.603	indeterminate
exploratory	support bin	ext	group	graphconv/t3b~t2/ bin1-5	-0.005	[-0.079, +0.074]	0.932	indeterminate
exploratory	support bin	ext	group	graphconv/t3b~t2/ bin100+	-0.014	[-0.053, +0.022]	0.637	indeterminate
exploratory	support bin	ext	group	graphconv/t3b~t2/ bin21-100	-0.019	[-0.059, +0.021]	0.586	indeterminate
exploratory	support bin	ext	group	graphconv/t3b~t2/ bin6-20	0.075	[+0.004, +0.139]	0.150	different
exploratory	support bin	ext	group	graphconv/t4~t2/ bin1-5	0.068	[+0.011, +0.133]	0.067	different
exploratory	support bin	ext	group	graphconv/t4~t2/ bin100+	-0.010	[-0.041, +0.019]	0.715	indeterminate

Family	Subtype	Dataset	Split	Comparison	dd	90% CI	q	Verdict
exploratory	support bin	ext	group	graphconv/t4~t2/ bin21-100	-0.004	[-0.040, +0.031]	0.918	indeterminate
exploratory	support bin	ext	group	graphconv/t4~t2/ bin6-20	0.055	[+0.011, +0.105]	0.093	different
exploratory	taxon group	ADORE	group	dmpnn/t3a~t2/algae	0.041	[-0.195, +0.332]	0.923	indeterminate
exploratory	taxon group	ADORE	group	dmpnn/t3a~t2/crusta	0.014	[-0.043, +0.072]	0.858	indeterminate
exploratory	taxon group	ADORE	group	dmpnn/t3a~t2/fish	-0.029	[-0.053, -0.006]	0.076	different
exploratory	taxon group	ADORE	group	dmpnn/t3b~t2/algae	-0.003	[-0.241, +0.243]	0.972	indeterminate
exploratory	taxon group	ADORE	group	dmpnn/t3b~t2/crusta	0.019	[-0.035, +0.074]	0.751	indeterminate
exploratory	taxon group	ADORE	group	dmpnn/t3b~t2/fish	-0.030	[-0.057, -0.002]	0.144	different
exploratory	taxon group	ADORE	group	dmpnn/t4~t2/algae	0.165	[-0.147, +0.433]	0.544	indeterminate
exploratory	taxon group	ADORE	group	dmpnn/t4~t2/crusta	0.007	[-0.033, +0.042]	0.906	indeterminate
exploratory	taxon group	ADORE	group	dmpnn/t4~t2/fish	0.012	[-0.013, +0.037]	0.544	indeterminate
exploratory	taxon group	ADORE	group	graphconv/t3a~t2/ algae	0.093	[-0.131, +0.322]	0.634	indeterminate
exploratory	taxon group	ADORE	group	graphconv/t3a~t2/ crusta	0.001	[-0.078, +0.078]	0.961	indeterminate
exploratory	taxon group	ADORE	group	graphconv/t3a~t2/fish	-0.046	[-0.076, -0.015]	0.034	different
exploratory	taxon group	ADORE	group	graphconv/t3b~t2/ algae	0.046	[-0.228, +0.306]	0.884	indeterminate
exploratory	taxon group	ADORE	group	graphconv/t3b~t2/ crusta	-0.014	[-0.083, +0.055]	0.807	indeterminate
exploratory	taxon group	ADORE	group	graphconv/t3b~t2/fish	-0.053	[-0.086, -0.021]	0.018	different
exploratory	taxon group	ADORE	group	graphconv/t4~t2/algae	0.031	[-0.259, +0.279]	0.923	indeterminate
exploratory	taxon group	ADORE	group	graphconv/t4~t2/ crusta	-0.024	[-0.073, +0.021]	0.539	indeterminate
exploratory	taxon group	ADORE	group	graphconv/t4~t2/fish	-0.012	[-0.045, +0.022]	0.622	indeterminate
exploratory	taxon group	ext	group	dmpnn/t3a~t2/algae	0.020	[-0.037, +0.091]	0.657	indeterminate
exploratory	taxon group	ext	group	dmpnn/t3a~t2/crusta	-0.032	[-0.101, +0.031]	0.591	indeterminate
exploratory	taxon group	ext	group	dmpnn/t3a~t2/fish	0.002	[-0.039, +0.042]	0.932	indeterminate
exploratory	taxon group	ext	group	dmpnn/t3b~t2/algae	-0.004	[-0.058, +0.057]	0.932	indeterminate
exploratory	taxon group	ext	group	dmpnn/t3b~t2/crusta	-0.005	[-0.074, +0.056]	0.959	indeterminate
exploratory	taxon group	ext	group	dmpnn/t3b~t2/fish	0.021	[-0.028, +0.065]	0.603	indeterminate
exploratory	taxon group	ext	group	dmpnn/t4~t2/algae	-0.004	[-0.053, +0.043]	0.932	indeterminate
exploratory	taxon group	ext	group	dmpnn/t4~t2/crusta	0.006	[-0.058, +0.064]	0.896	indeterminate
exploratory	taxon group	ext	group	dmpnn/t4~t2/fish	0.021	[-0.018, +0.060]	0.571	indeterminate
exploratory	taxon group	ext	group	graphconv/t3a~t2/ algae	0.014	[-0.062, +0.092]	0.815	indeterminate
exploratory	taxon group	ext	group	graphconv/t3a~t2/ crusta	-0.028	[-0.090, +0.026]	0.550	indeterminate
exploratory	taxon group	ext	group	graphconv/t3a~t2/fish	-0.008	[-0.070, +0.055]	0.824	indeterminate
exploratory	taxon group	ext	group	graphconv/t3b~t2/ algae	0.036	[-0.026, +0.100]	0.482	indeterminate
exploratory	taxon group	ext	group	graphconv/t3b~t2/ crusta	0.029	[-0.042, +0.112]	0.709	indeterminate
exploratory	taxon group	ext	group	graphconv/t3b~t2/fish	-0.005	[-0.045, +0.036]	0.891	indeterminate
exploratory	taxon group	ext	group	graphconv/t4~t2/algae	0.032	[-0.029, +0.092]	0.560	indeterminate
exploratory	taxon group	ext	group	graphconv/t4~t2/ crusta	0.066	[+0.007, +0.137]	0.104	different
exploratory	taxon group	ext	group	graphconv/t4~t2/fish	0.018	[-0.039, +0.072]	0.738	indeterminate
exploratory	scaffold	ADORE	Bemis	dmpnn/t1 > shuf	-0.098	[-0.130, -0.065]	0.000	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t1 > t0	-0.094	[-0.128, -0.058]	0.000	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t1' > shuf	-0.262	[-0.317, -0.210]	0.000	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t1' > t0	-0.047	[-0.066, -0.026]	0.000	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t2 > shuf	-0.104	[-0.147, -0.058]	0.000	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t2 > t0	-0.064	[-0.114, -0.014]	0.077	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t3a > shuf	-0.091	[-0.136, -0.041]	0.011	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t3a > t0	-0.070	[-0.122, -0.012]	0.103	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t3a~t2	-0.006	[-0.051, +0.049]	0.884	indeterminate
exploratory	scaffold	ADORE	Bemis	dmpnn/t3b > shuf	-0.111	[-0.164, -0.054]	0.005	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t3b > t0	-0.078	[-0.134, -0.018]	0.077	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t3b~t2	-0.014	[-0.061, +0.042]	0.751	indeterminate
exploratory	scaffold	ADORE	Bemis	dmpnn/t4 > shuf	-0.097	[-0.141, -0.051]	0.000	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t4 > t0	-0.077	[-0.125, -0.022]	0.043	significant
exploratory	scaffold	ADORE	Bemis	dmpnn/t4~t2	-0.013	[-0.039, +0.012]	0.551	indeterminate
exploratory	scaffold	ADORE	Bemis	graphconv/t1 > shuf	-0.092	[-0.127, -0.055]	0.000	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t1 > t0	-0.089	[-0.120, -0.056]	0.000	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t1' > shuf	-0.262	[-0.312, -0.213]	0.000	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t1' > t0	-0.054	[-0.075, -0.032]	0.000	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t2 > shuf	-0.128	[-0.180, -0.067]	0.005	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t2 > t0	-0.057	[-0.106, -0.008]	0.120	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t3a > shuf	-0.063	[-0.099, -0.023]	0.018	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t3a > t0	-0.077	[-0.127, -0.021]	0.046	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t3a~t2	-0.019	[-0.055, +0.017]	0.549	indeterminate

Family	Subtype	Dataset	Split	Comparison	dd	90% CI	q	Verdict
exploratory	scaffold	ADORE	Bemis	graphconv/t3b > shuf	-0.065	[-0.101, -0.028]	0.008	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t3b > t0	-0.073	[-0.123, -0.020]	0.045	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t3b~t2	-0.015	[-0.056, +0.027]	0.655	indeterminate
exploratory	scaffold	ADORE	Bemis	graphconv/t4 > shuf	-0.059	[-0.097, -0.016]	0.084	significant
exploratory	scaffold	ADORE	Bemis	graphconv/t4 > t0	-0.041	[-0.098, +0.019]	0.416	indeterminate
exploratory	scaffold	ADORE	Bemis	graphconv/t4~t2	0.016	[-0.024, +0.054]	0.664	indeterminate
exploratory	scaffold	ADORE	generic	dmpnn/t1 > shuf	-0.061	[-0.082, -0.039]	0.000	significant
exploratory	scaffold	ADORE	generic	dmpnn/t1 > t0	-0.054	[-0.077, -0.030]	0.000	significant
exploratory	scaffold	ADORE	generic	dmpnn/t1' > shuf	-0.256	[-0.300, -0.214]	0.000	significant
exploratory	scaffold	ADORE	generic	dmpnn/t1' > t0	-0.043	[-0.063, -0.022]	0.003	significant
exploratory	scaffold	ADORE	generic	dmpnn/t2 > shuf	-0.088	[-0.120, -0.053]	0.000	significant
exploratory	scaffold	ADORE	generic	dmpnn/t2 > t0	-0.031	[-0.062, +0.002]	0.199	indeterminate
exploratory	scaffold	ADORE	generic	dmpnn/t3a > shuf	-0.075	[-0.107, -0.044]	0.000	significant
exploratory	scaffold	ADORE	generic	dmpnn/t3a > t0	-0.024	[-0.062, +0.014]	0.438	indeterminate
exploratory	scaffold	ADORE	generic	dmpnn/t3a~t2	0.007	[-0.014, +0.029]	0.689	indeterminate
exploratory	scaffold	ADORE	generic	dmpnn/t3b > shuf	-0.084	[-0.114, -0.052]	0.000	significant
exploratory	scaffold	ADORE	generic	dmpnn/t3b > t0	-0.030	[-0.066, +0.004]	0.257	indeterminate
exploratory	scaffold	ADORE	generic	dmpnn/t3b~t2	0.001	[-0.018, +0.022]	0.926	indeterminate
exploratory	scaffold	ADORE	generic	dmpnn/t4 > shuf	-0.089	[-0.118, -0.059]	0.000	significant
exploratory	scaffold	ADORE	generic	dmpnn/t4 > t0	-0.039	[-0.073, -0.003]	0.150	significant
exploratory	scaffold	ADORE	generic	dmpnn/t4~t2	-0.008	[-0.027, +0.010]	0.634	indeterminate
exploratory	scaffold	ADORE	generic	graphconv/t1 > shuf	-0.074	[-0.095, -0.051]	0.000	significant
exploratory	scaffold	ADORE	generic	graphconv/t1 > t0	-0.066	[-0.087, -0.043]	0.000	significant
exploratory	scaffold	ADORE	generic	graphconv/t1' > shuf	-0.265	[-0.307, -0.224]	0.000	significant
exploratory	scaffold	ADORE	generic	graphconv/t1' > t0	-0.052	[-0.072, -0.031]	0.000	significant
exploratory	scaffold	ADORE	generic	graphconv/t2 > shuf	-0.092	[-0.126, -0.056]	0.000	significant
exploratory	scaffold	ADORE	generic	graphconv/t2 > t0	-0.045	[-0.088, -0.003]	0.150	significant
exploratory	scaffold	ADORE	generic	graphconv/t3a > shuf	-0.066	[-0.101, -0.033]	0.003	significant
exploratory	scaffold	ADORE	generic	graphconv/t3a > t0	-0.044	[-0.088, -0.002]	0.163	significant
exploratory	scaffold	ADORE	generic	graphconv/t3a~t2	0.001	[-0.026, +0.027]	0.961	indeterminate
exploratory	scaffold	ADORE	generic	graphconv/t3b > shuf	-0.073	[-0.105, -0.043]	0.000	significant
exploratory	scaffold	ADORE	generic	graphconv/t3b > t0	-0.047	[-0.094, +0.002]	0.196	indeterminate
exploratory	scaffold	ADORE	generic	graphconv/t3b~t2	-0.002	[-0.033, +0.028]	0.961	indeterminate
exploratory	scaffold	ADORE	generic	graphconv/t4 > shuf	-0.073	[-0.101, -0.048]	0.000	significant
exploratory	scaffold	ADORE	generic	graphconv/t4 > t0	-0.045	[-0.090, +0.001]	0.193	indeterminate
exploratory	scaffold	ADORE	generic	graphconv/t4~t2	0.001	[-0.024, +0.023]	0.961	indeterminate
exploratory	scaffold	ext	Bemis	dmpnn/t1 > shuf	-0.107	[-0.140, -0.071]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t1 > t0	-0.085	[-0.115, -0.052]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t1' > shuf	-0.380	[-0.436, -0.322]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t1' > t0	-0.105	[-0.144, -0.057]	0.002	significant
exploratory	scaffold	ext	Bemis	dmpnn/t2 > shuf	-0.201	[-0.269, -0.124]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t2 > t0	-0.129	[-0.184, -0.065]	0.002	significant
exploratory	scaffold	ext	Bemis	dmpnn/t3a > shuf	-0.137	[-0.191, -0.078]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t3a > t0	-0.102	[-0.154, -0.047]	0.004	significant
exploratory	scaffold	ext	Bemis	dmpnn/t3a~t2	0.027	[-0.015, +0.068]	0.460	indeterminate
exploratory	scaffold	ext	Bemis	dmpnn/t3b > shuf	-0.139	[-0.190, -0.084]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t3b > t0	-0.117	[-0.166, -0.066]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t3b~t2	0.012	[-0.032, +0.055]	0.738	indeterminate
exploratory	scaffold	ext	Bemis	dmpnn/t4 > shuf	-0.126	[-0.176, -0.072]	0.000	significant
exploratory	scaffold	ext	Bemis	dmpnn/t4 > t0	-0.107	[-0.152, -0.058]	0.002	significant
exploratory	scaffold	ext	Bemis	dmpnn/t4~t2	0.022	[-0.012, +0.055]	0.482	indeterminate
exploratory	scaffold	ext	Bemis	graphconv/t1 > shuf	-0.100	[-0.133, -0.063]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t1 > t0	-0.084	[-0.117, -0.046]	0.002	significant
exploratory	scaffold	ext	Bemis	graphconv/t1' > shuf	-0.380	[-0.432, -0.324]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t1' > t0	-0.106	[-0.145, -0.056]	0.002	significant
exploratory	scaffold	ext	Bemis	graphconv/t2 > shuf	-0.206	[-0.277, -0.132]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t2 > t0	-0.158	[-0.216, -0.091]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t3a > shuf	-0.151	[-0.206, -0.089]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t3a > t0	-0.147	[-0.207, -0.089]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t3a~t2	0.011	[-0.029, +0.048]	0.762	indeterminate
exploratory	scaffold	ext	Bemis	graphconv/t3b > shuf	-0.183	[-0.249, -0.113]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t3b > t0	-0.164	[-0.222, -0.104]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t3b~t2	-0.007	[-0.041, +0.026]	0.815	indeterminate
exploratory	scaffold	ext	Bemis	graphconv/t4 > shuf	-0.130	[-0.184, -0.072]	0.000	significant
exploratory	scaffold	ext	Bemis	graphconv/t4 > t0	-0.115	[-0.165, -0.060]	0.002	significant
exploratory	scaffold	ext	Bemis	graphconv/t4~t2	0.043	[+0.013, +0.071]	0.034	different
exploratory	scaffold	ext	generic	dmpnn/t1 > shuf	-0.120	[-0.161, -0.078]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t1 > t0	-0.125	[-0.163, -0.084]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t1' > shuf	-0.379	[-0.436, -0.318]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t1' > t0	-0.121	[-0.165, -0.073]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t2 > shuf	-0.158	[-0.222, -0.087]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t2 > t0	-0.150	[-0.211, -0.086]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t3a > shuf	-0.117	[-0.173, -0.059]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t3a > t0	-0.144	[-0.201, -0.086]	0.000	significant

Family	Subtype	Dataset	Split	Comparison	dd	90% CI	q	Verdict
exploratory	scaffold	ext	generic	dmpnn/t3a~t2	0.006	[-0.024, +0.036]	0.817	indeterminate
exploratory	scaffold	ext	generic	dmpnn/t3b > shuf	-0.120	[-0.175, -0.064]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t3b > t0	-0.148	[-0.209, -0.087]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t3b~t2	0.002	[-0.033, +0.035]	0.932	equivalent
exploratory	scaffold	ext	generic	dmpnn/t4 > shuf	-0.107	[-0.160, -0.049]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t4 > t0	-0.130	[-0.184, -0.076]	0.000	significant
exploratory	scaffold	ext	generic	dmpnn/t4~t2	0.020	[-0.005, +0.042]	0.297	indeterminate
exploratory	scaffold	ext	generic	graphconv/t1 > shuf	-0.091	[-0.131, -0.049]	0.000	significant
exploratory	scaffold	ext	generic	graphconv/t1 > t0	-0.093	[-0.130, -0.053]	0.000	significant
exploratory	scaffold	ext	generic	graphconv/t1' > shuf	-0.378	[-0.437, -0.315]	0.000	significant
exploratory	scaffold	ext	generic	graphconv/t1' > t0	-0.119	[-0.162, -0.070]	0.000	significant
exploratory	scaffold	ext	generic	graphconv/t2 > shuf	-0.112	[-0.180, -0.045]	0.013	significant
exploratory	scaffold	ext	generic	graphconv/t2 > t0	-0.099	[-0.160, -0.037]	0.019	significant
exploratory	scaffold	ext	generic	graphconv/t3a > shuf	-0.094	[-0.154, -0.033]	0.027	significant
exploratory	scaffold	ext	generic	graphconv/t3a > t0	-0.150	[-0.209, -0.091]	0.000	significant
exploratory	scaffold	ext	generic	graphconv/t3a~t2	-0.052	[-0.089, -0.015]	0.044	different
exploratory	scaffold	ext	generic	graphconv/t3b > shuf	-0.090	[-0.146, -0.032]	0.025	significant
exploratory	scaffold	ext	generic	graphconv/t3b > t0	-0.142	[-0.203, -0.081]	0.000	significant
exploratory	scaffold	ext	generic	graphconv/t3b~t2	-0.044	[-0.076, -0.009]	0.067	different
exploratory	scaffold	ext	generic	graphconv/t4 > shuf	-0.073	[-0.123, -0.020]	0.042	significant
exploratory	scaffold	ext	generic	graphconv/t4 > t0	-0.123	[-0.172, -0.072]	0.000	significant
exploratory	scaffold	ext	generic	graphconv/t4~t2	-0.025	[-0.059, +0.005]	0.325	indeterminate
exploratory	rank	ADORE	group	dmpnn/t3a_gf~t2	-0.006	[-0.034, +0.018]	0.827	indeterminate
exploratory	rank	ADORE	group	dmpnn/t3a_gf~t3a	0.011	[-0.006, +0.028]	0.416	indeterminate
exploratory	rank	ADORE	group	dmpnn/t3a_g~t2	-0.006	[-0.033, +0.019]	0.811	indeterminate
exploratory	rank	ADORE	group	dmpnn/t3a_g~t3a	0.012	[-0.007, +0.029]	0.449	indeterminate
exploratory	rank	ADORE	group	graphconv/t3a_gf~t2	-0.021	[-0.054, +0.011]	0.416	indeterminate
exploratory	rank	ADORE	group	graphconv/t3a_gf~t3a	0.012	[-0.006, +0.029]	0.416	indeterminate
exploratory	rank	ADORE	group	graphconv/t3a_g~t2	-0.012	[-0.046, +0.020]	0.640	indeterminate
exploratory	rank	ADORE	group	graphconv/t3a_g~t3a	0.022	[+0.001, +0.040]	0.150	different
exploratory	rank	ADORE	group	lightgbm/t3a_gf~t2	-0.014	[-0.034, +0.007]	0.416	indeterminate
exploratory	rank	ADORE	group	lightgbm/t3a_gf~t3a	0.013	[-0.002, +0.028]	0.288	indeterminate
exploratory	rank	ADORE	group	lightgbm/t3a_g~t2	-0.011	[-0.030, +0.007]	0.460	indeterminate
exploratory	rank	ADORE	group	lightgbm/t3a_g~t3a	0.016	[-0.010, +0.040]	0.479	indeterminate
exploratory	rank	ext	group	dmpnn/t3a_gf~t2	0.004	[-0.050, +0.052]	0.891	indeterminate
exploratory	rank	ext	group	dmpnn/t3a_gf~t3a	0.012	[-0.011, +0.034]	0.550	equivalent
exploratory	rank	ext	group	dmpnn/t3a_g~t2	0.009	[-0.042, +0.054]	0.797	indeterminate
exploratory	rank	ext	group	dmpnn/t3a_g~t3a	0.017	[-0.005, +0.039]	0.365	indeterminate
exploratory	rank	ext	group	graphconv/t3a_gf~t2	0.043	[+0.001, +0.086]	0.162	different
exploratory	rank	ext	group	graphconv/t3a_gf~t3a	0.034	[+0.000, +0.080]	0.160	different
exploratory	rank	ext	group	graphconv/t3a_g~t2	0.055	[+0.009, +0.104]	0.104	different
exploratory	rank	ext	group	graphconv/t3a_g~t3a	0.046	[+0.010, +0.094]	0.041	different
exploratory	rank	ext	group	lightgbm/t3a_gf~t2	-0.013	[-0.039, +0.014]	0.591	indeterminate
exploratory	rank	ext	group	lightgbm/t3a_gf~t3a	-0.012	[-0.029, +0.006]	0.470	equivalent
exploratory	rank	ext	group	lightgbm/t3a_g~t2	0.006	[-0.024, +0.038]	0.824	indeterminate
exploratory	rank	ext	group	lightgbm/t3a_g~t3a	0.007	[-0.034, +0.046]	0.815	indeterminate
exploratory	cross-backbone	ADORE	group	dmpnn/t3a ~ lgbm/t2	-0.031	[-0.073, +0.012]	0.386	indeterminate
exploratory	cross-backbone	ADORE	group	dmpnn/t3b ~ lgbm/t2	-0.031	[-0.071, +0.012]	0.416	indeterminate
exploratory	cross-backbone	ADORE	group	dmpnn/t4 ~ lgbm/t2	-0.002	[-0.040, +0.035]	0.961	indeterminate
exploratory	cross-backbone	ADORE	group	graphconv/t3a ~ lgbm/t2	-0.038	[-0.082, +0.005]	0.257	indeterminate
exploratory	cross-backbone	ADORE	group	graphconv/t3b ~ lgbm/t2	-0.048	[-0.091, -0.004]	0.150	different
exploratory	cross-backbone	ADORE	group	graphconv/t4 ~ lgbm/t2	-0.020	[-0.059, +0.017]	0.495	indeterminate
exploratory	cross-backbone	ext	group	dmpnn/t3a ~ lgbm/t2	-0.051	[-0.141, +0.034]	0.521	indeterminate
exploratory	cross-backbone	ext	group	dmpnn/t3b ~ lgbm/t2	-0.041	[-0.131, +0.045]	0.591	indeterminate
exploratory	cross-backbone	ext	group	dmpnn/t4 ~ lgbm/t2	-0.043	[-0.130, +0.037]	0.554	indeterminate
exploratory	cross-backbone	ext	group	graphconv/t3a ~ lgbm/t2	-0.078	[-0.172, +0.011]	0.259	indeterminate
exploratory	cross-backbone	ext	group	graphconv/t3b ~ lgbm/t2	-0.055	[-0.135, +0.022]	0.387	indeterminate
exploratory	cross-backbone	ext	group	graphconv/t4 ~ lgbm/t2	-0.043	[-0.114, +0.028]	0.482	indeterminate
exploratory	leaky	ADORE	leaky	dmpnn/t1 > shuf	-0.092	[-0.109, -0.074]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t1 > t0	-0.088	[-0.105, -0.070]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t1' > shuf	-0.284	[-0.319, -0.248]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t1' > t0	-0.052	[-0.071, -0.031]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t2 > shuf	-0.143	[-0.170, -0.115]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t2 > t0	-0.064	[-0.089, -0.036]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t3a > shuf	-0.160	[-0.186, -0.135]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t3a > t0	-0.104	[-0.131, -0.077]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t3a~t2	-0.040	[-0.061, -0.019]	0.003	different
exploratory	leaky	ADORE	leaky	dmpnn/t3b > shuf	-0.155	[-0.181, -0.128]	0.000	significant

Family	Subtype	Dataset	Split	Comparison	dd	90% CI	q	Verdict
exploratory	leaky	ADORE	leaky	dmpnn/t3b > t0	-0.097	[-0.125, -0.070]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t3b~t2	-0.034	[-0.053, -0.014]	0.016	different
exploratory	leaky	ADORE	leaky	dmpnn/t4 > shuf	-0.111	[-0.134, -0.086]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t4 > t0	-0.056	[-0.079, -0.032]	0.000	significant
exploratory	leaky	ADORE	leaky	dmpnn/t4~t2	0.008	[-0.008, +0.025]	0.514	indeterminate
exploratory	leaky	ADORE	leaky	graphconv/t1 > shuf	-0.089	[-0.105, -0.071]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t1 > t0	-0.083	[-0.099, -0.064]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t1' > shuf	-0.283	[-0.316, -0.248]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t1' > t0	-0.051	[-0.070, -0.030]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t2 > shuf	-0.151	[-0.179, -0.121]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t2 > t0	-0.069	[-0.093, -0.042]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t3a > shuf	-0.155	[-0.181, -0.130]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t3a > t0	-0.113	[-0.139, -0.087]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t3a~t2	-0.045	[-0.067, -0.022]	0.003	different
exploratory	leaky	ADORE	leaky	graphconv/t3b > shuf	-0.159	[-0.185, -0.134]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t3b > t0	-0.113	[-0.138, -0.086]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t3b~t2	-0.044	[-0.066, -0.021]	0.000	different
exploratory	leaky	ADORE	leaky	graphconv/t4 > shuf	-0.118	[-0.141, -0.094]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t4 > t0	-0.072	[-0.094, -0.049]	0.000	significant
exploratory	leaky	ADORE	leaky	graphconv/t4~t2	-0.003	[-0.019, +0.013]	0.884	equivalent
exploratory	leaky	ext	leaky	dmpnn/t1 > shuf	-0.101	[-0.121, -0.081]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t1 > t0	-0.085	[-0.105, -0.065]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t1' > shuf	-0.410	[-0.455, -0.364]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t1' > t0	-0.095	[-0.127, -0.065]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t2 > shuf	-0.195	[-0.232, -0.160]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t2 > t0	-0.111	[-0.147, -0.076]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t3a > shuf	-0.176	[-0.207, -0.145]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t3a > t0	-0.132	[-0.162, -0.101]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t3a~t2	-0.021	[-0.051, +0.006]	0.365	indeterminate
exploratory	leaky	ext	leaky	dmpnn/t3b > shuf	-0.166	[-0.195, -0.136]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t3b > t0	-0.122	[-0.153, -0.090]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t3b~t2	-0.011	[-0.039, +0.015]	0.621	indeterminate
exploratory	leaky	ext	leaky	dmpnn/t4 > shuf	-0.095	[-0.128, -0.062]	0.000	significant
exploratory	leaky	ext	leaky	dmpnn/t4 > t0	-0.051	[-0.085, -0.017]	0.023	significant
exploratory	leaky	ext	leaky	dmpnn/t4~t2	0.060	[+0.043, +0.077]	0.000	different
exploratory	leaky	ext	leaky	graphconv/t1 > shuf	-0.104	[-0.124, -0.083]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t1 > t0	-0.089	[-0.109, -0.068]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t1' > shuf	-0.409	[-0.456, -0.362]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t1' > t0	-0.097	[-0.129, -0.067]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t2 > shuf	-0.174	[-0.214, -0.137]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t2 > t0	-0.091	[-0.129, -0.055]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t3a > shuf	-0.176	[-0.209, -0.145]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t3a > t0	-0.136	[-0.168, -0.104]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t3a~t2	-0.045	[-0.075, -0.017]	0.013	different
exploratory	leaky	ext	leaky	graphconv/t3b > shuf	-0.169	[-0.199, -0.140]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t3b > t0	-0.126	[-0.157, -0.095]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t3b~t2	-0.035	[-0.061, -0.010]	0.039	different
exploratory	leaky	ext	leaky	graphconv/t4 > shuf	-0.093	[-0.128, -0.059]	0.000	significant
exploratory	leaky	ext	leaky	graphconv/t4 > t0	-0.051	[-0.087, -0.017]	0.036	significant
exploratory	leaky	ext	leaky	graphconv/t4~t2	0.039	[+0.024, +0.055]	0.000	different

Table S8: Deterministic (LightGBM) equivalence tests against δ_{det} (ADORE 0.087, ext 0.074), which is 2–4 $\times$ wider than the network margin δ and can therefore contain the intervals. This is the only family that returned *equivalent* verdicts on a species-representation contrast.

Dataset	Backbone	Comparison	dd	90% CI	q	Verdict
ADORE	LightGBM	lgbm/t3a~t2	-0.027	[-0.052, +0.001]	0.154	equivalent
ADORE	LightGBM	lgbm/t3b~t2	-0.022	[-0.045, +0.003]	0.154	equivalent
ADORE	LightGBM	lgbm/t4~t2	0.036	[+0.008, +0.064]	0.105	equivalent
ext	LightGBM	lgbm/t3a~t2	-0.000	[-0.029, +0.028]	0.980	equivalent
ext	LightGBM	lgbm/t3b~t2	-0.015	[-0.045, +0.018]	0.661	equivalent
ext	LightGBM	lgbm/t4~t2	0.082	[+0.051, +0.112]	0.000	different

S7. Exploratory families

Cited from Sections 4.2–4.3 and Table 7. Six exploratory sub-families decompose, extend, or control the primary comparisons. *Rank truncation* repeats the source-authority taxonomy contrast using only

the genus rank, and the genus and family ranks (Table S9). *Support bin* splits each one-hot-to-richer TOST pair by per-species training-record count (1–5, 6–20, 21–100, more than 100). *Taxon group* splits the same pairs by fish / crustacean / algae. *Scaffold* repeats the primary set under the Bemis–Murcko and generic scaffold splits. *Leakage control* is the deliberately leaky (SMILES, species)-pair split, a pre-registered positive control for the leakage tripwire (Table S10). *Cross-backbone* contrasts a graph-network representation against the gradient-boosted one-hot as a true four-arm difference-in-differences, each arm anchored to its own no-species baseline. All rows appear in Table S7 (Subtype column); the rank-truncation and leakage-control sub-families are additionally tabulated in Tables S9 and S10.

Table S9: Rank-truncation family (exploratory). t3a_g truncates source-authority taxonomy to genus, t3a_gf to genus and family; each is compared against full source-authority taxonomy (t3a) and against one-hot identity (t2). The two *equivalent* verdicts on the ext set (t3a_gf vs t3a) are the equivalence cases noted in Section 4.3.

Dataset	Backbone	Comparison	dd	90% CI	q	Verdict
ADORE	D-MPNN	dmpnn/t3a_gf~t2	-0.006	[-0.034, +0.018]	0.827	indeterminate
ADORE	D-MPNN	dmpnn/t3a_gf~t3a	0.011	[-0.006, +0.028]	0.416	indeterminate
ADORE	D-MPNN	dmpnn/t3a_g~t2	-0.006	[-0.033, +0.019]	0.811	indeterminate
ADORE	D-MPNN	dmpnn/t3a_g~t3a	0.012	[-0.007, +0.029]	0.449	indeterminate
ADORE	GraphConv	graphconv/t3a_gf~t2	-0.021	[-0.054, +0.011]	0.416	indeterminate
ADORE	GraphConv	graphconv/t3a_gf~t3a	0.012	[-0.006, +0.029]	0.416	indeterminate
ADORE	GraphConv	graphconv/t3a_g~t2	-0.012	[-0.046, +0.020]	0.640	indeterminate
ADORE	GraphConv	graphconv/t3a_g~t3a	0.022	[+0.001, +0.040]	0.150	different
ADORE	LightGBM	lightgbm/t3a_gf~t2	-0.014	[-0.034, +0.007]	0.416	indeterminate
ADORE	LightGBM	lightgbm/t3a_gf~t3a	0.013	[-0.002, +0.028]	0.288	indeterminate
ADORE	LightGBM	lightgbm/t3a_g~t2	-0.011	[-0.030, +0.007]	0.460	indeterminate
ADORE	LightGBM	lightgbm/t3a_g~t3a	0.016	[-0.010, +0.040]	0.479	indeterminate
ext	D-MPNN	dmpnn/t3a_gf~t2	0.004	[-0.050, +0.052]	0.891	indeterminate
ext	D-MPNN	dmpnn/t3a_gf~t3a	0.012	[-0.011, +0.034]	0.550	equivalent
ext	D-MPNN	dmpnn/t3a_g~t2	0.009	[-0.042, +0.054]	0.797	indeterminate
ext	D-MPNN	dmpnn/t3a_g~t3a	0.017	[-0.005, +0.039]	0.365	indeterminate
ext	GraphConv	graphconv/t3a_gf~t2	0.043	[+0.001, +0.086]	0.162	different
ext	GraphConv	graphconv/t3a_gf~t3a	0.034	[+0.000, +0.080]	0.160	different
ext	GraphConv	graphconv/t3a_g~t2	0.055	[+0.009, +0.104]	0.104	different
ext	GraphConv	graphconv/t3a_g~t3a	0.046	[+0.010, +0.094]	0.041	different
ext	LightGBM	lightgbm/t3a_gf~t2	-0.013	[-0.039, +0.014]	0.591	indeterminate
ext	LightGBM	lightgbm/t3a_gf~t3a	-0.012	[-0.029, +0.006]	0.470	equivalent
ext	LightGBM	lightgbm/t3a_g~t2	0.006	[-0.024, +0.038]	0.824	indeterminate
ext	LightGBM	lightgbm/t3a_g~t3a	0.007	[-0.034, +0.046]	0.815	indeterminate

Table S10: Leakage-control family: the deliberately leaky (SMILES, species)-pair split, retained as a pre-registered positive control for the leakage tripwire. The tripwire (a stopping threshold on implausibly low error) fired only here. Verdict tally on ADORE = 1 equivalent / 28 different / 1 indeterminate.

Dataset	Backbone	Comparison	dd	90% CI	q	Verdict
ADORE	D-MPNN	dmpnn/t1 > shuf	-0.092	[-0.109, -0.074]	0.000	significant
ADORE	D-MPNN	dmpnn/t1 > t0	-0.088	[-0.105, -0.070]	0.000	significant
ADORE	D-MPNN	dmpnn/t1' > shuf	-0.284	[-0.319, -0.248]	0.000	significant
ADORE	D-MPNN	dmpnn/t1' > t0	-0.052	[-0.071, -0.031]	0.000	significant
ADORE	D-MPNN	dmpnn/t2 > shuf	-0.143	[-0.170, -0.115]	0.000	significant
ADORE	D-MPNN	dmpnn/t2 > t0	-0.064	[-0.089, -0.036]	0.000	significant
ADORE	D-MPNN	dmpnn/t3a > shuf	-0.160	[-0.186, -0.135]	0.000	significant
ADORE	D-MPNN	dmpnn/t3a > t0	-0.104	[-0.131, -0.077]	0.000	significant
ADORE	D-MPNN	dmpnn/t3a~t2	-0.040	[-0.061, -0.019]	0.003	different
ADORE	D-MPNN	dmpnn/t3b > shuf	-0.155	[-0.181, -0.128]	0.000	significant
ADORE	D-MPNN	dmpnn/t3b > t0	-0.097	[-0.125, -0.070]	0.000	significant
ADORE	D-MPNN	dmpnn/t3b~t2	-0.034	[-0.053, -0.014]	0.016	different

Dataset	Backbone	Comparison	dd	90% CI	q	Verdict
ADORE	D-MPNN	dmpnn/t4 > shuf	-0.111	[-0.134, -0.086]	0.000	significant
ADORE	D-MPNN	dmpnn/t4 > t0	-0.056	[-0.079, -0.032]	0.000	significant
ADORE	D-MPNN	dmpnn/t4~t2	0.008	[-0.008, +0.025]	0.514	indeterminate
ADORE	GraphConv	graphconv/t1 > shuf	-0.089	[-0.105, -0.071]	0.000	significant
ADORE	GraphConv	graphconv/t1 > t0	-0.083	[-0.099, -0.064]	0.000	significant
ADORE	GraphConv	graphconv/t1' > shuf	-0.283	[-0.316, -0.248]	0.000	significant
ADORE	GraphConv	graphconv/t1' > t0	-0.051	[-0.070, -0.030]	0.000	significant
ADORE	GraphConv	graphconv/t2 > shuf	-0.151	[-0.179, -0.121]	0.000	significant
ADORE	GraphConv	graphconv/t2 > t0	-0.069	[-0.093, -0.042]	0.000	significant
ADORE	GraphConv	graphconv/t3a > shuf	-0.155	[-0.181, -0.130]	0.000	significant
ADORE	GraphConv	graphconv/t3a > t0	-0.113	[-0.139, -0.087]	0.000	significant
ADORE	GraphConv	graphconv/t3a~t2	-0.045	[-0.067, -0.022]	0.003	different
ADORE	GraphConv	graphconv/t3b > shuf	-0.159	[-0.185, -0.134]	0.000	significant
ADORE	GraphConv	graphconv/t3b > t0	-0.113	[-0.138, -0.086]	0.000	significant
ADORE	GraphConv	graphconv/t3b~t2	-0.044	[-0.066, -0.021]	0.000	different
ADORE	GraphConv	graphconv/t4 > shuf	-0.118	[-0.141, -0.094]	0.000	significant
ADORE	GraphConv	graphconv/t4 > t0	-0.072	[-0.094, -0.049]	0.000	significant
ADORE	GraphConv	graphconv/t4~t2	-0.003	[-0.019, +0.013]	0.884	equivalent
ext	D-MPNN	dmpnn/t1 > shuf	-0.101	[-0.121, -0.081]	0.000	significant
ext	D-MPNN	dmpnn/t1 > t0	-0.085	[-0.105, -0.065]	0.000	significant
ext	D-MPNN	dmpnn/t1' > shuf	-0.410	[-0.455, -0.364]	0.000	significant
ext	D-MPNN	dmpnn/t1' > t0	-0.095	[-0.127, -0.065]	0.000	significant
ext	D-MPNN	dmpnn/t2 > shuf	-0.195	[-0.232, -0.160]	0.000	significant
ext	D-MPNN	dmpnn/t2 > t0	-0.111	[-0.147, -0.076]	0.000	significant
ext	D-MPNN	dmpnn/t3a > shuf	-0.176	[-0.207, -0.145]	0.000	significant
ext	D-MPNN	dmpnn/t3a > t0	-0.132	[-0.162, -0.101]	0.000	significant
ext	D-MPNN	dmpnn/t3a~t2	-0.021	[-0.051, +0.006]	0.365	indeterminate
ext	D-MPNN	dmpnn/t3b > shuf	-0.166	[-0.195, -0.136]	0.000	significant
ext	D-MPNN	dmpnn/t3b > t0	-0.122	[-0.153, -0.090]	0.000	significant
ext	D-MPNN	dmpnn/t3b~t2	-0.011	[-0.039, +0.015]	0.621	indeterminate
ext	D-MPNN	dmpnn/t4 > shuf	-0.095	[-0.128, -0.062]	0.000	significant
ext	D-MPNN	dmpnn/t4 > t0	-0.051	[-0.085, -0.017]	0.023	significant
ext	D-MPNN	dmpnn/t4~t2	0.060	[+0.043, +0.077]	0.000	different
ext	GraphConv	graphconv/t1 > shuf	-0.104	[-0.124, -0.083]	0.000	significant
ext	GraphConv	graphconv/t1 > t0	-0.089	[-0.109, -0.068]	0.000	significant
ext	GraphConv	graphconv/t1' > shuf	-0.409	[-0.456, -0.362]	0.000	significant
ext	GraphConv	graphconv/t1' > t0	-0.097	[-0.129, -0.067]	0.000	significant
ext	GraphConv	graphconv/t2 > shuf	-0.174	[-0.214, -0.137]	0.000	significant
ext	GraphConv	graphconv/t2 > t0	-0.091	[-0.129, -0.055]	0.000	significant
ext	GraphConv	graphconv/t3a > shuf	-0.176	[-0.209, -0.145]	0.000	significant
ext	GraphConv	graphconv/t3a > t0	-0.136	[-0.168, -0.104]	0.000	significant
ext	GraphConv	graphconv/t3a~t2	-0.045	[-0.075, -0.017]	0.013	different
ext	GraphConv	graphconv/t3b > shuf	-0.169	[-0.199, -0.140]	0.000	significant
ext	GraphConv	graphconv/t3b > t0	-0.126	[-0.157, -0.095]	0.000	significant
ext	GraphConv	graphconv/t3b~t2	-0.035	[-0.061, -0.010]	0.039	different
ext	GraphConv	graphconv/t4 > shuf	-0.093	[-0.128, -0.059]	0.000	significant
ext	GraphConv	graphconv/t4 > t0	-0.051	[-0.087, -0.017]	0.036	significant
ext	GraphConv	graphconv/t4~t2	0.039	[+0.024, +0.055]	0.000	different

S8. Ensemble sensitivity analysis

Cited from Section 3.4 and Table 7. The ensemble margin δ' pools the within-condition standard deviation of per-seed RMSE over $k = 10$ disjoint ten-seed ensembles drawn from a 90-seed pool (seeds 10–99), by the same procedure as δ ; it equals 0.0057 on ADORE and 0.0160 on the ECOTOX-external set. The primary and confirmatory TOST pairs re-decided against δ' are listed below (not gated).

Table S11: Ensemble sensitivity family: the primary and confirmatory TOST pairs re-decided against the ensemble margin δ' (ADORE 0.0057, ext 0.0160), computed on $k = 10$ disjoint 10-seed ensembles drawn from a 90-seed pool. The *Partition* column separates the discovery and replication partitions (both compound-registry split). Not gated.

Partition	Dataset	Backbone	Comparison	dd	90% CI	q	Verdict
discovery	ADORE	D-MPNN	dmpnn/t3a~t2/ens	-0.024	[-0.051, +0.004]	0.281	indeterminate
discovery	ADORE	D-MPNN	dmpnn/t3b~t2/ens	-0.023	[-0.051, +0.008]	0.342	indeterminate
discovery	ADORE	D-MPNN	dmpnn/t4~t2/ens	-0.007	[-0.030, +0.021]	0.730	indeterminate
discovery	ADORE	GraphConv	graphconv/t3a~t2/ens	-0.042	[-0.077, -0.003]	0.204	different
discovery	ADORE	GraphConv	graphconv/t3b~t2/ens	-0.049	[-0.083, -0.013]	0.081	different
discovery	ADORE	GraphConv	graphconv/t4~t2/ens	-0.032	[-0.060, +0.002]	0.252	indeterminate
replication	ADORE	D-MPNN	dmpnn/t3a~t2/ens	-0.013	[-0.042, +0.017]	0.542	indeterminate
replication	ADORE	D-MPNN	dmpnn/t3b~t2/ens	-0.020	[-0.049, +0.008]	0.364	indeterminate
replication	ADORE	D-MPNN	dmpnn/t4~t2/ens	-0.013	[-0.043, +0.016]	0.627	indeterminate
replication	ADORE	GraphConv	graphconv/t3a~t2/ens	-0.049	[-0.085, -0.011]	0.081	different
replication	ADORE	GraphConv	graphconv/t3b~t2/ens	-0.064	[-0.103, -0.024]	0.066	different
replication	ADORE	GraphConv	graphconv/t4~t2/ens	-0.051	[-0.081, -0.017]	0.066	different
–	ext	D-MPNN	dmpnn/t3a~t2/ens	-0.047	[-0.094, +0.008]	0.507	indeterminate
–	ext	D-MPNN	dmpnn/t3b~t2/ens	-0.028	[-0.076, +0.023]	0.685	indeterminate
–	ext	D-MPNN	dmpnn/t4~t2/ens	-0.058	[-0.094, -0.012]	0.168	different
–	ext	GraphConv	graphconv/t3a~t2/ens	-0.028	[-0.069, +0.019]	0.685	indeterminate
–	ext	GraphConv	graphconv/t3b~t2/ens	-0.001	[-0.042, +0.046]	0.939	indeterminate
–	ext	GraphConv	graphconv/t4~t2/ens	-0.009	[-0.036, +0.034]	0.939	indeterminate

S9. Species-axis variance capture

Cited from Section 4.6 and Fig. 4. The method and the full values are given in Table S12.

Table S12: Species-axis variance capture by representation. $\text{capture} = (\text{RMSE}_{t_0}^2 - \text{RMSE}_{\text{tier}}^2) / s_{\text{species}}^2$, with $s_{\text{species}} = 0.960$ on ADORE and 1.079 on the ECOTOX-external set (pairwise SD of $|\Delta_{\text{target}}|$ over fix-compound, vary-species). A rough decomposition: the denominator includes replicate noise, so the ratio is not bounded above by 1.

Representation	ADORE		ECOTOX-external	
	D-MPNN	GraphConv	D-MPNN	GraphConv
no species information	0.000	0.000	0.000	0.000
species mean offset	0.173	0.126	0.216	0.295
out-of-fold residual calibration	0.132	0.131	0.265	0.260
one-hot identity	0.192	0.169	0.359	0.471
source-authority taxonomy	0.234	0.249	0.377	0.450
NCBI taxonomy	0.232	0.271	0.355	0.397
learned embedding	0.164	0.206	0.360	0.370

S10. Species-by-chemical interaction

Cited from Section 4.6. The interaction is quantified by the cross-class rank correlation of per-species sensitivity (Table S13) and, for the classes where a within-class reliability can be estimated, by split-half reliability with a Spearman–Brown correction (Table S14).

Over all 28 class pairs the median cross-class ρ is 0.225 (Table S13). Restricted to the six classes for which a within-class split-half reliability could be estimated, the median cross-class ρ over the fifteen pairs among them is 0.31. The median Spearman–Brown-corrected within-class reliability of those six classes is 0.63 (Table S14); this is the ceiling on any between-class rank correlation given the measurement noise within each class. The observed value is therefore about half the reliability ceiling, so the cross-class rearrangement of species sensitivity is not attributable to measurement noise alone. Medians are reported throughout, for consistency with the 28-pair summary and because the low reliability of the preservative class (0.10) makes the mean unrepresentative.

Table S13: Cross-class species-sensitivity rank correlation. For each pair of OECD-function chemical classes with at least 5 compounds and at least 10 shared species, the Spearman correlation of per-species mean sensitivity (28 pairs over 14 classes). Summary: median $\rho = 0.225$, IQR [0.065, 0.408], range $[-0.301, 0.891]$.

Class A	Class B	n shared species	Spearman ρ	Kendall τ
binder	biocide	18	0.414	0.268
binder	colorant	10	0.891	0.778
binder	fragrance	10	-0.176	-0.156
binder	solvent - surfactant	10	0.515	0.378
biocide	chelating agent	14	-0.301	-0.191
biocide	colorant	52	0.359	0.225
biocide	flame retardant	11	0.627	0.345
biocide	fragrance	24	0.217	0.123
biocide	pH regulating agent	10	0.042	-0.022
biocide	preservative	27	0.147	0.111
biocide	solvent	34	0.164	0.091
biocide	solvent - surfactant	42	0.334	0.228
biocide	surfactant	14	0.068	0.055
chelating agent	solvent	10	0.232	0.205
colorant	fragrance	13	0.038	0.051
colorant	preservative	13	0.407	0.359
colorant	solvent	15	0.618	0.429
colorant	solvent - surfactant	18	0.108	0.059
fragrance	preservative	13	0.308	0.231
fragrance	solvent	15	-0.018	-0.010
fragrance	solvent - surfactant	17	0.091	0.059
fragrance	surfactant	10	0.200	0.200
preservative	solvent	15	0.314	0.238
preservative	solvent - surfactant	18	0.428	0.359
preservative	surfactant	10	0.055	0.022
solvent	solvent - surfactant	23	0.673	0.486
solvent	surfactant	12	0.287	0.212
solvent - surfactant	surfactant	13	-0.060	-0.077

Table S14: Within-class split-half reliability and Spearman–Brown correction. Method: $\rho_{\text{cross}} / \sqrt{\text{rel_a} * \text{rel_b}}$; rel=within-class split-half Spearman (also Spearman-Brown full variant) ($n_{\text{reps}} = 50$ random splits of (compound $\times$ species) cells; min 5 compounds and 10 shared species per class). Reliability could be computed for 6 of the classes; the split-half value is half-data, and Spearman–Brown gives the full-data reliability, which may exceed 1.

Chemical class	n cells	Split-half ρ	Spearman–Brown ρ
biocide	3238	0.418	0.590
colorant	110	0.484	0.653
fragrance	118	0.427	0.599
preservative	111	0.050	0.095
solvent	154	0.603	0.752
solvent - surfactant	103	0.520	0.685

S11. Out-of-vocabulary re-aggregation

Cited from Section 5.1: the sensitivity of the comparison to removing the 341 warm test rows whose species are absent from training, and the 61-of-396 species analysis.

Table S15: Out-of-vocabulary (OOV) re-aggregation. The one-hot-to-richer comparisons re-computed after removing warm test rows whose species never appear in training (ADORE: 61 of 396 test species, 69 of 2,029 rows; ECOTOX-external: 187 of 934 species, 341 of 4,561 rows, 7.5%). Reference only; not used for judgment. The per-comparison row count was recorded in the artifact only for the ECOTOX-external panel; the ADORE re-aggregation removed the 69 out-of-vocabulary rows from the 2,029 test rows.

Backbone	Contrast	dd	90% CI	Verdict	<i>n</i> rows
<i>ADORE (single-condition partition, compound-registry split)</i>					
D-MPNN	source-authority taxonomy vs one-hot identity	-0.021	[-0.050, +0.005]	indeterminate	–
D-MPNN	NCBI taxonomy vs one-hot identity	-0.021	[-0.050, +0.010]	indeterminate	–
D-MPNN	learned embedding vs one-hot identity	0.004	[-0.021, +0.027]	indeterminate	–
GraphConv	source-authority taxonomy vs one-hot identity	-0.036	[-0.073, -0.001]	different	–
GraphConv	NCBI taxonomy vs one-hot identity	-0.046	[-0.083, -0.013]	different	–
GraphConv	learned embedding vs one-hot identity	-0.021	[-0.053, +0.011]	indeterminate	–
<i>ECOTOX-external (compound-registry split)</i>					
D-MPNN	source-authority taxonomy vs one-hot identity	-0.020	[-0.072, +0.030]	indeterminate	4220
D-MPNN	NCBI taxonomy vs one-hot identity	-0.010	[-0.060, +0.034]	indeterminate	4220
D-MPNN	learned embedding vs one-hot identity	-0.016	[-0.055, +0.019]	indeterminate	4220
GraphConv	source-authority taxonomy vs one-hot identity	-0.010	[-0.051, +0.030]	indeterminate	4220
GraphConv	NCBI taxonomy vs one-hot identity	0.006	[-0.035, +0.046]	indeterminate	4220
GraphConv	learned embedding vs one-hot identity	0.018	[-0.013, +0.048]	indeterminate	4220

S12. Concentration-fold interpretation

Cited from Sections 3.6 and 5.1. Section 3.6 reports effect sizes on both the $\log_{10}$ and concentration-fold scales; the concentration-fold values for the sixteen headline comparisons are given in Table S16.

Table S16: Concentration-fold interpretation of the sixteen headline effect sizes. A difference-in-differences of $|\text{dd}| \log_{10}$ units corresponds to a $10^{|\text{dd}|}$ -fold change in the predicted effect concentration. Verdict follows Section 3.7 (*significant* for the superiority tests, *different* for the two one-sided tests, both meaning the interval excludes zero); a negative dd favours the candidate.

Dataset	Backbone	Comparison	dd ($\log_{10}$)	Fold ($\times$)	Verdict
ADORE	D-MPNN	no species $\rightarrow$ one-hot	-0.076	1.19	significant
ADORE	D-MPNN	one-hot $\rightarrow$ source-authority tax.	-0.018	1.04	indeterminate
ADORE	D-MPNN	one-hot $\rightarrow$ NCBI tax.	-0.017	1.04	indeterminate
ADORE	D-MPNN	one-hot $\rightarrow$ learned embedding	+0.012	1.03	indeterminate
ADORE	GraphConv	no species $\rightarrow$ one-hot	-0.067	1.17	significant
ADORE	GraphConv	one-hot $\rightarrow$ source-authority tax.	-0.033	1.08	indeterminate
ADORE	GraphConv	one-hot $\rightarrow$ NCBI tax.	-0.043	1.10	different
ADORE	GraphConv	one-hot $\rightarrow$ learned embedding	-0.015	1.04	indeterminate
ECOTOX-external	D-MPNN	no species $\rightarrow$ one-hot	-0.150	1.41	significant
ECOTOX-external	D-MPNN	one-hot $\rightarrow$ source-authority tax.	-0.008	1.02	indeterminate
ECOTOX-external	D-MPNN	one-hot $\rightarrow$ NCBI tax.	+0.002	1.00	indeterminate
ECOTOX-external	D-MPNN	one-hot $\rightarrow$ learned embedding	-0.000	1.00	indeterminate
ECOTOX-external	GraphConv	no species $\rightarrow$ one-hot	-0.195	1.57	significant
ECOTOX-external	GraphConv	one-hot $\rightarrow$ source-authority tax.	+0.009	1.02	indeterminate
ECOTOX-external	GraphConv	one-hot $\rightarrow$ NCBI tax.	+0.032	1.08	indeterminate
ECOTOX-external	GraphConv	one-hot $\rightarrow$ learned embedding	+0.044	1.11	different

S13. One-hot identity versus the species mean offset

Cited from Section 5.1. This post-hoc comparison isolates the value of a full one-hot identity (t_2) over a single per-species offset (t_1); results are in Table S17.

Table S17: One-hot identity (t2) versus the species mean offset (t1) on ADORE, within backbone, post-hoc. Per-seed paired difference-in-differences, block bootstrap (SMILES), 90% CI, against $\delta = 0.0198$; a negative dd favours one-hot identity. q is Benjamini–Hochberg adjusted within these four comparisons only.

Partition	Backbone	dd	90% CI	q	Verdict
single-condition (discovery)	D-MPNN	-0.008	[-0.036, +0.022]	0.700	indeterminate
single-condition (discovery)	GraphConv	-0.018	[-0.050, +0.017]	0.539	indeterminate
seven-condition (confirmatory)	D-MPNN	-0.087	[-0.120, -0.049]	0.008	significant
seven-condition (confirmatory)	GraphConv	-0.080	[-0.137, -0.020]	0.042	significant

References

- [1] Jochen P. Zubrod, Nika Galic, Maxime Vaugeois, and David A. Dreier. Physiological variables in machine learning QSARs allow for both cross-chemical and cross-species predictions. *Ecotoxicology and Environmental Safety*, 263:115250, 2023.
- [2] Marco Forastiere, Martina G. Vijver, Leo Posthuma, and Sara Visser. Bridging chemical structure and genetic similarity: a Bio-QSAR model for cross-species toxicity predictions. *Ecotoxicology and Environmental Safety*, 314:120057, 2026.
- [3] Markus Viljanen, Jelle Minnema, Pim N. H. Wassenaar, Emiel Rorije, and Willie Peijnenburg. What is the ecotoxicity of a given chemical for a given aquatic species? predicting interactions between species and chemicals using recommender system techniques. *SAR and QSAR in Environmental Research*, 34(10):765, 2023.
- [4] Till Schlender, Markus Viljanen, Jan N. van Rijn, Felix Mohr, Willie Peijnenburg, Holger H. Hoos, Emiel Rorije, and Albert Wong. The bigger fish: a comparison of meta-learning QSAR models on low-resourced aquatic toxicity regression tasks. *Environmental Science & Technology*, 57(46):17818, 2023.
- [5] Y. Yuan, S. Shaw, X. Tang, and Y. Ye. Cross-chemical and cross-species toxicity prediction: benchmarking and a novel 3D-structure-based deep learning model. *bioRxiv*, 2025. 2025.11.24.690199.
- [6] Yang Yang, Ying Yang, Xin Pan, Qiao Xue, Yong Liu, Jianjie Fu, and Aiqian Zhang. Multimodal integration of chemical and biological descriptors for cross-species prediction of fish acute toxicity. *Environmental Science & Technology*, 60(3):2556, 2026.
- [7] Jin Wu, Simone D’Ambrosi, Lucas Ammann, Julita Stadnicka-Michalak, Kristin Schirmer, and Marco Baity-Jesi. Predicting chemical hazard across taxa through machine learning. *Environment International*, 163:107184, 2022.