

# Patterns of Competitive Exclusion in the Mammalian Fossil Record

## *Appendix*

### AUTHORS:

**Esther Galbrun** (corresponding author)

`esther.galbrun@uef.fi`

School of Computing, University of Eastern Finland

Technopolis, Microkatu 1, FI-70210 Kuopio, Finland.

**Jo Skeie Hermansen**

Centre for Ecological and Evolutionary Synthesis, University of Oslo

P.O. Box 1066, University of Oslo, NO-0316 Oslo, Norway.

**Indrė Žliobaitė**

Department of Computer Science, University of Helsinki,

P.O. Box 64, FI-00014 University of Helsinki, Finland.

## A Supplementary information about the data

This section presents supplementary information about the datasets used in our study.

Table 1 lists age boundaries used as time bins in the two continents considered in this study, Europe and North America.

Table 1: Age boundaries of the time units used as time bins for the two continents.

Europe			North America		
Time units	Bounds (Ma)	Rank	Time units	Bounds (Ma)	Rank
MN2	21.70 – 19.50	Zone	Orellan	33.90 – 32.10	Age
MN3	19.50 – 17.20	Zone	Whitneyan	32.10 – 29.75	Age
MN4	17.20 – 16.40	Zone	Arikareean-1	29.75 – 28.00	Subage
MN5	16.40 – 14.20	Zone	Arikareean-2	28.00 – 22.20	Subage
MN6	14.20 – 12.85	Zone	Arikareean-3	22.20 – 19.09	Subage
MN7-8	12.85 – 11.20	Zone	Arikareean-4	19.09 – 18.50	Subage
MN9	11.20 – 9.90	Zone	Hemingfordian-1	18.50 – 17.34	Subage
MN10	9.90 – 8.90	Zone	Hemingfordian-2	17.34 – 15.97	Subage
MN11	8.90 – 7.60	Zone	Barstovian-1	15.97 – 14.78	Subage
MN12	7.60 – 7.10	Zone	Barstovian-2	14.78 – 12.57	Subage
MN13	7.10 – 5.30	Zone	Clarendonian-1	12.57 – 12.11	Subage
MN14	5.30 – 5.00	Zone	Clarendonian-2	12.11 – 10.09	Subage
MN15	5.00 – 3.55	Zone	Clarendonian-3	10.09 – 9.07	Subage
MN16	3.55 – 2.50	Zone	Hemphillian-1	9.07 – 7.59	Subage
			Hemphillian-2	7.59 – 6.88	Subage
			Hemphillian-3	6.88 – 5.91	Subage
			Hemphillian-4	5.91 – 4.91	Subage
			Blancan-Early	4.91 – 2.63	Subage

Tables 2–4 give summary statistics of the datasets for the three different faunal groups, respectively. Visual summaries of the distribution of species occurrences between different orders and genera over time for each faunal group and continent are provided in Figs. 1–6. We are aware of the presence of outliers, possibly due to misidentifications. They are under investigation following the NOW curatorial process. To ensure the transparency of analysis, we deliberately choose not to manually amend the dataset extracted from the NOW database and we therefore do not modify or delete any occurrences outside the database.

## B Supplementary results

Recall that results are obtained in three major steps: *(i)* computing co-occurrence statistics from the original data, *(ii)* comparing the values from the original data to null models, *(iii)* examining trends in the environmental context. Such three-steps experiment is carried out separately for each faunal group over each continent. Furthermore, we consider three different co-occurrence statistics and three different null models. Hence, we performed a total of  $2 \times 3 \times 3 \times 3 = 54$  three-steps computational experiments, one for each of the combinations outlined below.

$$\begin{array}{c} \text{continent} \\ \text{Europe} \\ \text{North America} \end{array} \left\} \times \left\{ \begin{array}{c} \text{faunal group} \\ \text{large herbivores} \\ \text{small mammals} \\ \text{large carnivorans} \end{array} \right\} \times \left\{ \begin{array}{c} \text{co-occurrence statistic} \\ C\text{-score} \\ \text{mid-}P \text{ value} \\ \text{nb. of genera with co-occ. sp.} \end{array} \right\} \times \left\{ \begin{array}{c} \text{null model} \\ \text{UG} \\ \text{Curveball} \\ \text{shuffle} \end{array} \right\}$$

This section presents an exhaustive report of our experiments with different co-occurrence statistics and null models.

Table 2: Counts for large herbivores. The columns indicate: the total number of localities and the number of localities with a single occurrence, the median and the maximum number of occurrences per locality, the total number of species and the number of species with only single occurrence, the median and the maximum number of occurrences per species, the total number of occurrence, i.e. the number of non-zero species–locality entries. In each column, the pair of values correspond to the counts excluding/including undeterminate and unnamed species. The first count, marked with a  $\star$ , indicates the number of general localities, where many small fossil localities were lumped together into a larger unit.

	nb. of localities				occ. per locality		nb. of species		occ. per species		occ.
	total		s. occ.		median	max	total	s. occ.	median	max	total
Europe											
MN2	4*	47/ 52	6/ 9		3/ 3	16/16	41/ 57	14/24	3/2	18/18	186/219
MN3	10*	77/ 85	31/25		3/ 3	14/14	47/ 68	14/25	3/2	24/24	319/376
MN4	1*	60/ 66	20/19		2/ 3	19/19	66/ 94	23/35	2/2	22/22	274/348
MN5	9*	126/143	36/45		3/ 4	27/27	108/148	40/56	2/2	43/43	764/886
MN6	2*	77/ 92	31/37		2/ 2	23/23	76/106	21/38	2/2	20/20	299/368
MN7-8	5*	103/122	39/48		2/ 2	19/19	85/129	36/56	2/2	38/38	384/476
MN9	5*	104/144	32/39		3/ 3	20/21	105/164	44/67	2/2	25/25	441/626
MN10	3*	53/ 62	16/14		3/ 4	21/23	94/143	49/80	1/1	19/19	258/362
MN11	5*	41/ 51	10/ 8		4/ 6	18/23	111/168	62/89	1/1	11/16	233/372
MN12	11*	99/121	25/26		4/ 5	36/39	139/200	47/77	3/2	28/36	724/921
MN13	1*	42/ 48	6/ 6		3/ 5	21/21	104/148	56/85	1/1	13/13	248/340
MN14	2*	24/ 34	7/11		3/ 2	14/17	43/ 62	27/34	1/1	10/10	88/128
MN15	1*	23/ 28	4/ 6		4/ 4	14/14	53/ 82	38/60	1/1	9/ 9	106/146
MN16	0*	49/ 54	16/16		3/ 3	15/15	73/109	41/66	1/1	20/20	205/260
North America											
Orellan	7*	19/ 20	4/ 3		3/ 6	33/36	47/ 77	19/41	2/1	10/10	138/179
Whitneyan	7*	14/ 14	5/ 2		3/ 8	39/41	51/ 68	23/32	2/2	5/ 6	97/132
Arikareean-1	7*	17/ 17	3/ 3		6/ 7	18/21	46/ 62	19/30	2/2	10/10	113/134
Arikareean-2	7*	18/ 18	5/ 4		3/ 5	12/14	33/ 53	19/36	1/1	10/10	78/103
Arikareean-3	8*	19/ 23	4/ 5		5/ 5	18/26	67/ 99	40/60	1/1	7/ 7	115/169
Arikareean-4	6*	13/ 13	2/ 2		4/ 4	25/28	48/ 68	25/43	1/1	4/ 4	84/106
Hemingfordian-1	13*	32/ 32	9/ 7		3/ 4	17/28	67/102	37/56	1/1	11/11	134/200
Hemingfordian-2	9*	28/ 30	6/ 5		3/ 6	25/33	65/105	33/53	1/1	8/ 8	138/232
Barstovian-1	21*	36/ 39	4/ 6		6/10	31/39	110/165	50/73	2/2	13/13	296/436
Barstovian-2	22*	42/ 50	15/12		3/ 5	32/40	116/170	53/76	2/2	15/15	300/447
Clarendonian-1	2*	5/ 6	1/ 1		3/ 3	29/35	32/ 42	27/37	1/1	3/ 3	39/ 49
Clarendonian-2	12*	29/ 34	5/ 8		5/ 6	26/32	85/122	40/58	2/2	13/13	228/318
Clarendonian-3	8*	18/ 18	3/ 0		3/ 6	17/20	42/ 70	30/45	1/1	7/ 7	75/133
Hemphillian-1	10*	18/ 20	3/ 3		4/ 8	18/20	53/ 88	26/43	2/2	9/12	107/186
Hemphillian-2	8*	21/ 22	3/ 3		7/ 9	16/21	63/ 96	34/48	1/1	9/ 9	143/217
Hemphillian-3	8*	14/ 15	5/ 4		3/ 5	14/18	27/ 48	11/22	2/2	6/ 6	58/ 92
Hemphillian-4	9*	20/ 23	3/ 3		4/ 7	25/27	47/ 84	26/44	1/1	9/ 9	98/167
Blancan-Early	22*	48/ 56	12/ 7		2/ 5	8/12	36/ 72	16/27	2/2	33/33	131/275

Table 3: Counts for small mammals. The columns indicate: the total number of localities and the number of localities with a single occurrence, the median and the maximum number of occurrences per locality, the total number of species and the number of species with only single occurrence, the median and the maximum number of occurrences per species, the total number of occurrence, i.e. the number of non-zero species–locality entries. In each column, the pair of values correspond to the counts excluding/including undeterminate and unnamed species. The first count, marked with a  $\star$ , indicates the number of general localities, where many small fossil localities were lumped together into a larger unit.

	nb. of localities		occ. per locality		nb. of species		occ. per species		occ.
	total	s. occ.	median	max	total	s. occ.	median	max	total
<b>Europe</b>									
MN2	1 $\star$ 90/ 94	18/18	4/ 5	31/33	145/200	50/ 79	2/2	40/40	555/ 685
MN3	8 $\star$ 128/133	21/22	5/ 7	39/48	164/242	52/ 83	2/2	54/54	952/1171
MN4	4 $\star$ 116/127	18/14	8/11	44/48	187/283	61/ 99	3/3	52/52	1247/1580
MN5	6 $\star$ 223/231	37/33	8/10	34/42	178/282	53/ 94	4/3	84/84	1800/2247
MN6	2 $\star$ 101/102	16/13	7/ 9	35/38	139/215	54/ 89	2/2	44/44	802/1084
MN7-8	3 $\star$ 104/108	9/ 8	9/12	48/49	195/291	71/113	3/2	56/56	1149/1452
MN9	3 $\star$ 99/104	26/20	4/ 5	26/35	163/254	75/110	2/2	31/31	610/ 824
MN10	2 $\star$ 105/111	14/15	4/ 7	34/42	118/183	42/ 78	3/2	56/56	678/ 893
MN11	0 $\star$ 41/ 45	7/ 8	7/ 9	30/35	89/138	35/ 73	2/1	23/23	363/ 467
MN12	5 $\star$ 66/ 71	16/14	5/ 6	16/21	82/123	41/ 63	1/1	37/37	342/ 509
MN13	2 $\star$ 98/101	17/13	6/ 8	20/26	141/216	77/116	1/1	41/41	593/ 852
MN14	2 $\star$ 71/ 81	14/15	7/ 8	43/57	159/244	67/105	2/2	26/26	489/ 703
MN15	1 $\star$ 56/ 61	6/10	7/ 9	47/64	204/282	104/150	1/1	29/29	577/ 726
MN16	0 $\star$ 44/ 52	6/10	5/ 6	37/37	153/235	80/139	1/1	11/11	335/ 456
<b>North America</b>									
Orellan	7 $\star$ 15/ 16	2/ 3	7/10	44/47	85/116	45/ 63	1/1	10/10	193/ 243
Whitneyan	7 $\star$ 10/ 10	0/ 0	3/ 5	24/35	41/ 72	28/ 50	1/1	7/ 7	68/ 109
Arikareean-1	7 $\star$ 17/ 18	5/ 2	3/ 5	34/37	79/119	55/ 84	1/1	3/ 3	115/ 169
Arikareean-2	7 $\star$ 12/ 14	3/ 1	3/ 5	25/38	77/131	59/ 97	1/1	4/ 4	99/ 174
Arikareean-3	7 $\star$ 14/ 17	2/ 4	5/ 6	21/38	76/116	62/ 95	1/1	3/ 3	93/ 140
Arikareean-4	4 $\star$ 7/ 7	1/ 0	3/ 4	19/24	35/ 43	33/ 40	1/1	3/ 3	39/ 48
Hemingfordian-1	11 $\star$ 19/ 22	6/ 5	2/ 3	21/25	62/103	46/ 74	1/1	3/ 3	84/ 141
Hemingfordian-2	6 $\star$ 15/ 17	5/ 3	3/ 5	12/22	50/ 94	39/ 72	1/1	4/ 4	65/ 122
Barstovian-1	16 $\star$ 22/ 27	5/ 4	4/ 5	15/21	85/121	62/ 84	1/1	5/ 9	122/ 188
Barstovian-2	18 $\star$ 25/ 29	6/ 4	3/ 5	32/43	105/164	66/100	1/1	6/ 9	186/ 306
Clarendonian-1	2 $\star$ 4/ 4	2/ 1	4/ 7	9/17	18/ 31	18/ 29	1/1	1/ 2	18/ 33
Clarendonian-2	8 $\star$ 13/ 18	3/ 6	5/ 4	16/18	66/ 93	48/ 67	1/1	4/ 6	87/ 129
Clarendonian-3	7 $\star$ 11/ 12	1/ 2	2/ 3	11/22	33/ 54	29/ 48	1/1	4/ 4	39/ 63
Hemphillian-1	8 $\star$ 12/ 14	1/ 1	3/ 4	11/16	39/ 64	31/ 49	1/1	5/ 5	51/ 89
Hemphillian-2	4 $\star$ 12/ 12	3/ 2	4/ 5	15/23	41/ 63	30/ 47	1/1	4/ 4	57/ 84
Hemphillian-3	8 $\star$ 11/ 12	4/ 3	3/ 4	22/26	49/ 73	39/ 59	1/1	4/ 4	61/ 90
Hemphillian-4	7 $\star$ 12/ 13	5/ 2	2/ 3	13/29	41/ 66	35/ 54	1/1	2/ 3	47/ 80
Blancan-Early	24 $\star$ 52/ 57	10/ 8	4/ 6	29/32	154/206	69/ 95	2/2	13/13	364/ 486

Table 4: Counts for large carnivorans. The columns indicate: the total number of localities and the number of localities with a single occurrence, the median and the maximum number of occurrences per locality, the total number of species and the number of species with only single occurrence, the median and the maximum number of occurrences per species, the total number of occurrence, i.e. the number of non-zero species–locality entries. In each column, the pair of values correspond to the counts excluding/including undeterminate and unnamed species. The first count, marked with a  $\star$ , indicates the number of general localities, where many small fossil localities were lumped together into a larger unit.

	nb. of localities		occ. per locality		nb. of species		occ. per species		occ.
	total	s. occ.	median	max	total	s. occ.	median	max	total
<b>Europe</b>									
MN2	0 $\star$ 26/29	15/15	1/1	21/21	40/51	16/24	2/2	7/ 7	88/104
MN3	5 $\star$ 44/49	14/17	3/3	19/20	45/56	14/24	3/2	18/18	223/235
MN4	2 $\star$ 29/35	14/17	2/2	18/18	47/57	23/31	2/1	8/ 8	117/131
MN5	7 $\star$ 58/63	17/16	3/3	22/22	61/78	23/33	2/2	16/16	274/306
MN6	2 $\star$ 23/36	13/22	1/1	19/22	37/58	22/36	1/1	9/ 9	84/116
MN7-8	2 $\star$ 27/34	8/13	4/4	39/41	62/84	20/37	2/2	9/ 9	193/221
MN9	3 $\star$ 46/64	20/30	2/2	11/16	59/87	31/46	1/1	10/10	134/191
MN10	2 $\star$ 27/37	9/13	2/2	23/26	58/85	33/53	1/1	9/ 9	111/154
MN11	2 $\star$ 18/27	8/12	2/2	17/22	40/64	27/46	1/1	8/ 8	68/101
MN12	10 $\star$ 55/67	18/27	3/2	19/20	53/75	19/32	2/2	29/29	221/262
MN13	1 $\star$ 21/27	5/10	3/3	13/19	45/67	29/46	1/1	6/ 6	84/112
MN14	1 $\star$ 12/19	6/11	1/1	7/ 9	17/27	11/17	1/1	4/ 4	26/ 42
MN15	0 $\star$ 14/16	3/ 2	2/4	10/12	37/53	29/39	1/1	6/ 6	58/ 82
MN16	0 $\star$ 24/26	8/ 7	3/3	13/14	40/59	23/38	1/1	10/10	93/118
<b>North America</b>									
Orellan	7 $\star$ 13/16	3/ 5	2/2	12/15	22/31	11/18	1/1	9/ 9	56/ 68
Whitneyan	7 $\star$ 8/11	4/ 4	2/2	17/18	26/33	17/22	1/1	4/ 4	39/ 48
Arikareean-1	7 $\star$ 14/14	5/ 4	2/3	17/18	42/50	26/33	1/1	7/ 7	73/ 82
Arikareean-2	7 $\star$ 9/13	1/ 4	4/4	10/12	31/39	23/26	1/1	5/ 5	43/ 56
Arikareean-3	6 $\star$ 8/10	0/ 2	4/3	10/12	30/38	23/29	1/1	3/ 3	40/ 50
Arikareean-4	5 $\star$ 8/ 8	1/ 0	3/4	13/15	25/31	15/19	1/1	5/ 5	40/ 50
Hemingfordian-1	9 $\star$ 19/22	7/ 9	2/2	18/28	30/48	17/31	1/1	6/ 6	58/ 80
Hemingfordian-2	7 $\star$ 16/19	7/ 5	2/2	12/14	21/30	10/15	2/1	8/ 8	48/ 62
Barstovian-1	17 $\star$ 25/30	8/ 9	2/3	26/27	57/72	30/40	1/1	11/11	132/156
Barstovian-2	18 $\star$ 29/32	12/10	3/2	18/22	52/69	26/35	1/1	10/10	133/161
Clarendonian-1	3 $\star$ 4/ 4	2/ 2	1/2	13/16	15/19	13/17	1/1	2/ 2	17/ 21
Clarendonian-2	11 $\star$ 20/25	6/ 7	2/2	16/18	30/48	14/24	2/1	11/11	83/114
Clarendonian-3	6 $\star$ 13/15	5/ 3	2/5	8/12	26/43	18/28	1/1	5/ 5	42/ 70
Hemphillian-1	9 $\star$ 15/15	2/ 0	3/4	8/11	28/39	19/25	1/1	11/11	55/ 77
Hemphillian-2	9 $\star$ 17/17	5/ 3	3/6	9/10	35/49	23/31	1/1	7/ 7	66/ 89
Hemphillian-3	8 $\star$ 12/13	4/ 4	3/4	9/10	21/29	13/19	1/1	7/ 7	42/ 56
Hemphillian-4	7 $\star$ 16/17	5/ 3	2/4	15/20	36/53	17/26	2/2	7/ 7	76/106
Blancan-Early	19 $\star$ 31/42	9/11	2/2	17/19	43/65	19/32	2/2	12/12	122/167

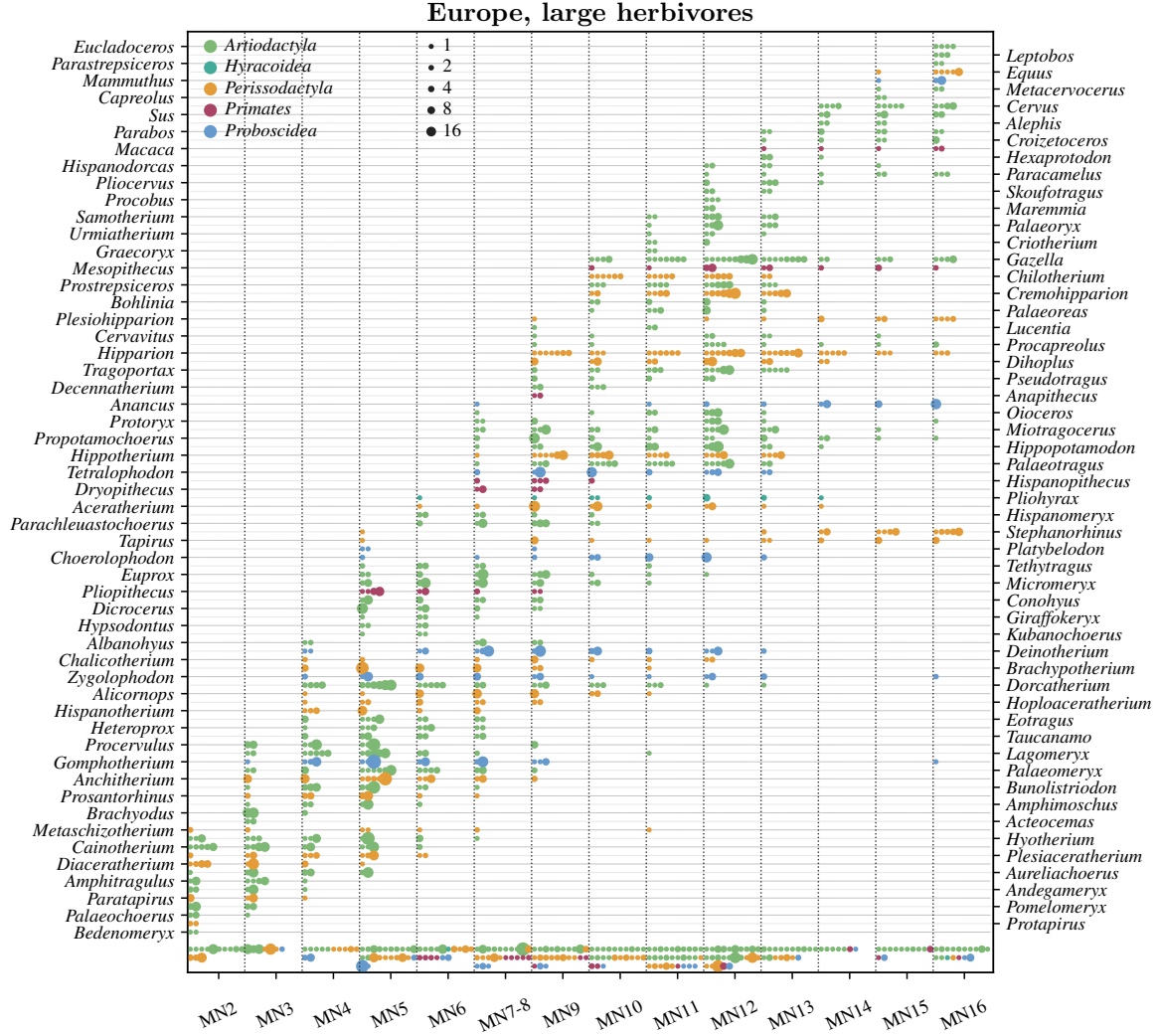


Figure 1: Distribution of species occurrences between different orders and genera over time for Europe, large herbivores. Each horizontal line represents a genus. For each time bin, each circle on the line represents a distinct species of that genus, with size proportional to its number of occurrences within the time bin, and colored according to the order it belongs to. Note that multiple circles correspond to the same species, if it occurs in multiple time bins. Genera are ordered in such a way that earliest occurring genera are placed at the bottom and recent ones at the top. Occurrences for genera such that at most one species occurs within any time bin are gathered at the bottom of the plot. Undetermined and unnamed species are excluded. (A scalable vector graphics version with species labels is available at [https://zliobaite.github.io/patterns\\_complex/bbl-gen0\\_EU-L.svg](https://zliobaite.github.io/patterns_complex/bbl-gen0_EU-L.svg))

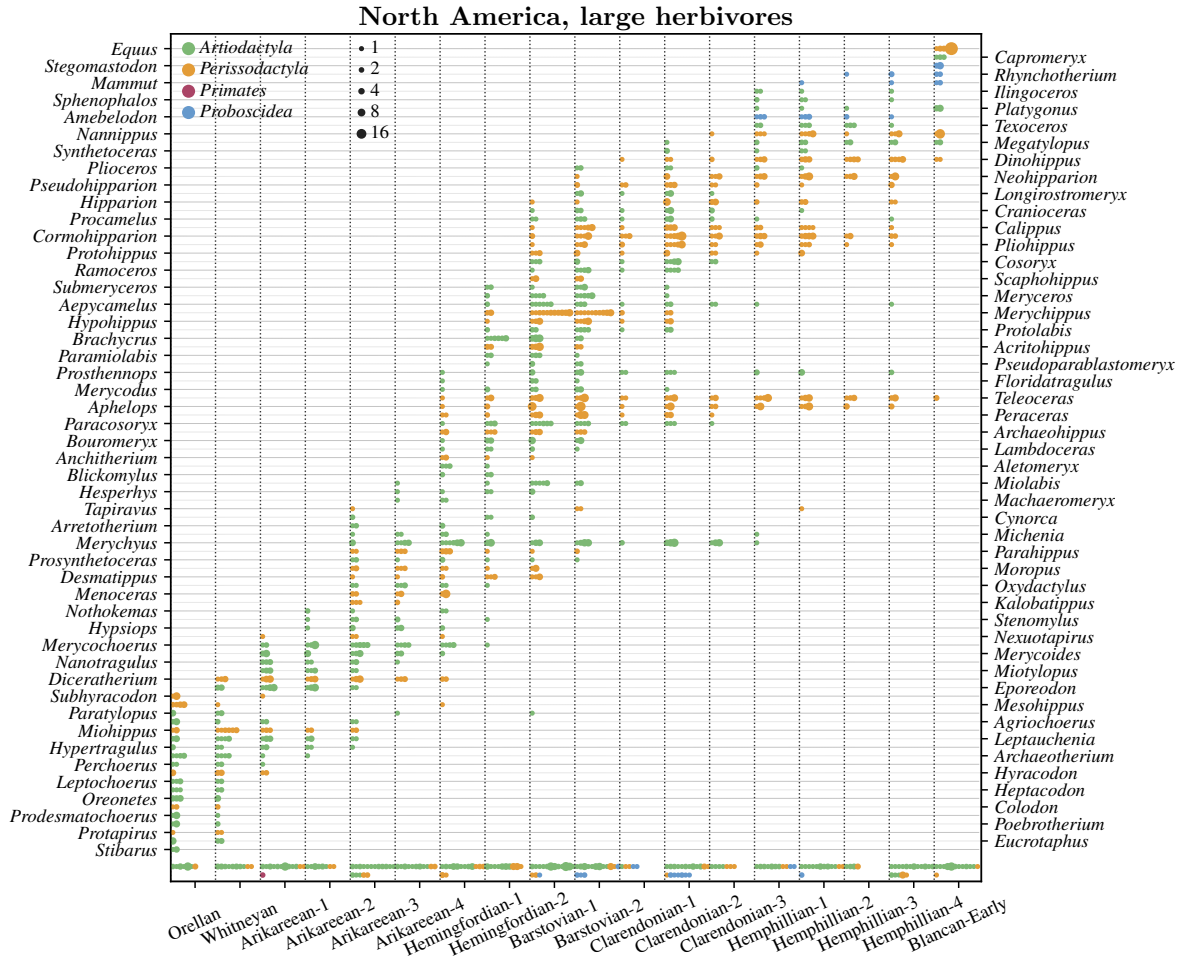


Figure 2: Distribution of species occurrences between different orders and genera over time for North America, large herbivores. See Fig. 1 for a detailed explanation. (A scalable vector graphics version with species labels is available at [https://zliobaite.github.io/patterns\\_compex/bbl-gen0\\_NA-L.svg](https://zliobaite.github.io/patterns_compex/bbl-gen0_NA-L.svg))

### Europe, small mammals

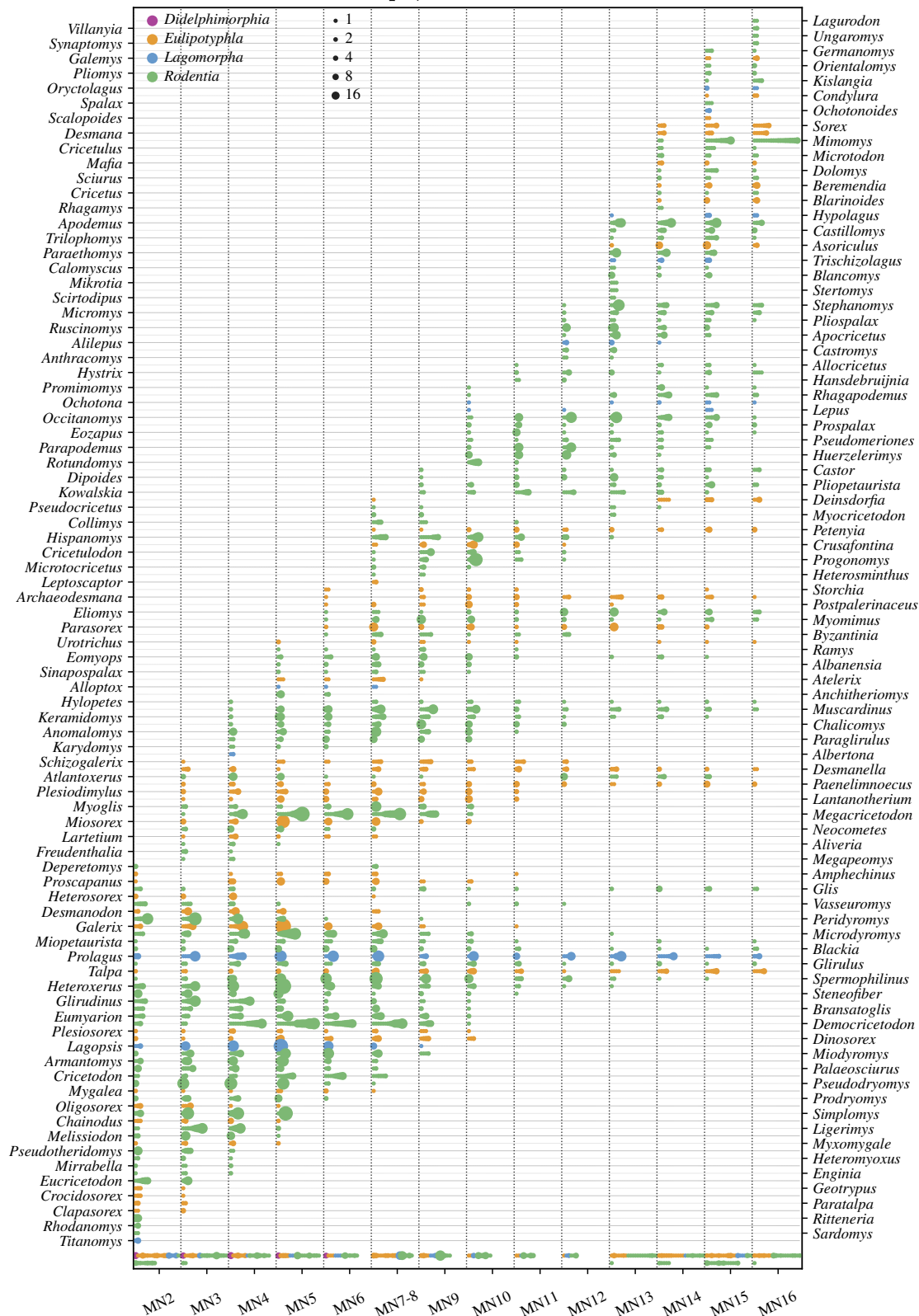


Figure 3: Distribution of species occurrences between different orders and genera over time for Europe, small mammals. See Fig. 1 for a detailed explanation. (A scalable vector graphics version with species labels is available at [https://zliobaite.github.io/patterns\\_compex/bb1-gen0\\_EU-S.svg](https://zliobaite.github.io/patterns_compex/bb1-gen0_EU-S.svg))



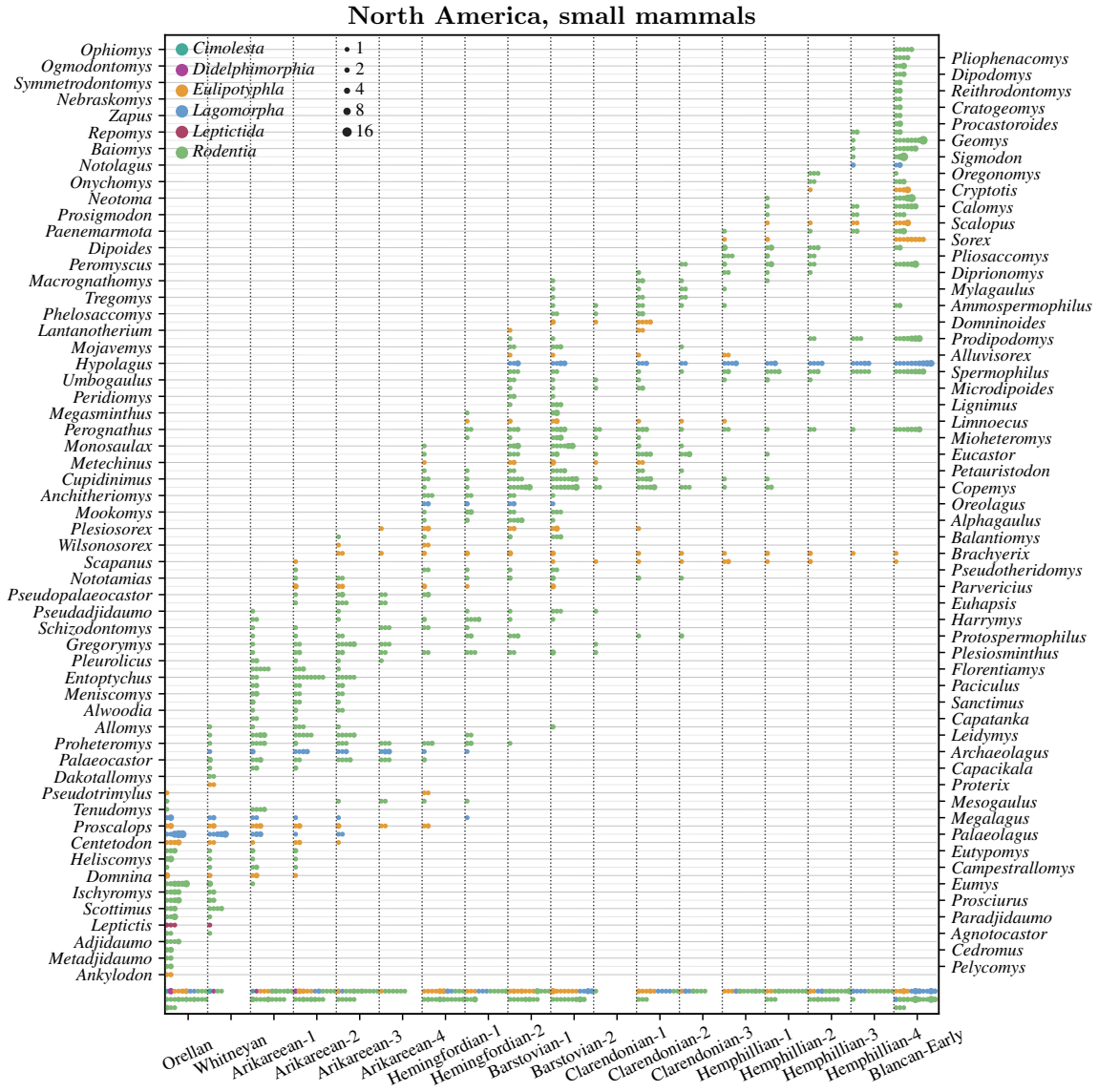


Figure 4: Distribution of species occurrences between different orders and genera over time for North America, small mammals. See Fig. 1 for a detailed explanation. (A scalable vector graphics version with species labels is available at [https://zliobaite.github.io/patterns\\_compex/bbl-gen0\\_NA-S.svg](https://zliobaite.github.io/patterns_compex/bbl-gen0_NA-S.svg))

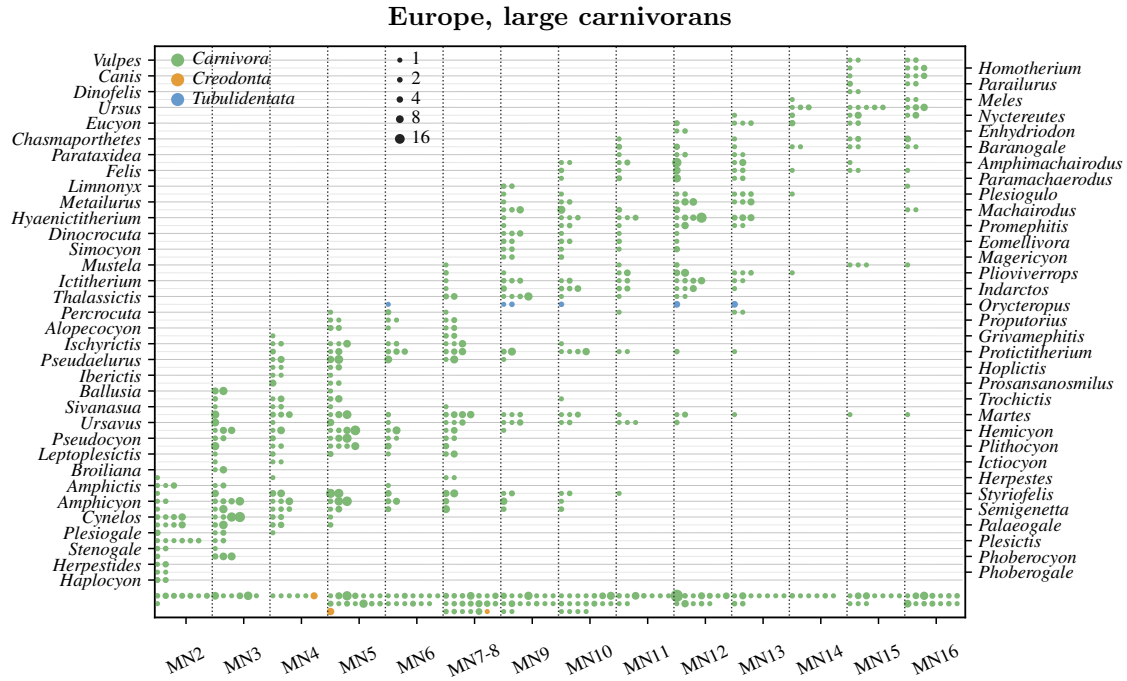


Figure 5: Distribution of species occurrences between different orders and genera over time for Europe, large carnivorans. See Fig. 1 for a detailed explanation. (A scalable vector graphics version with species labels is available at [https://zliobaite.github.io/patterns\\_compex/bbl-gen0\\_EU-C.svg](https://zliobaite.github.io/patterns_compex/bbl-gen0_EU-C.svg))

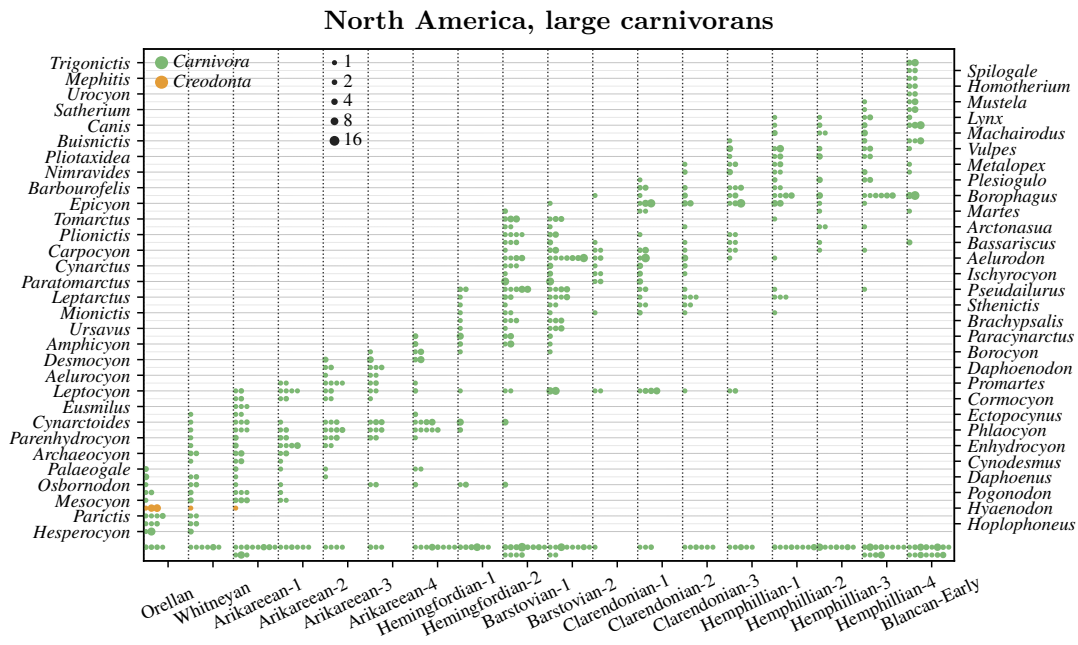


Figure 6: Distribution of species occurrences between different orders and genera over time for North America, large carnivorans. See Fig. 1 for a detailed explanation. (A scalable vector graphics version with species labels is available at [https://zliobaite.github.io/patterns\\_compex/bbl-gen0\\_NA-C.svg](https://zliobaite.github.io/patterns_compex/bbl-gen0_NA-C.svg))

In the main text, we focused on one co-occurrence statistic, namely the number of genera with co-occurring species, and on one null model, namely the fixed-fixed model generated by the **Curveball** randomization algorithm. Furthermore, to illustrate the first and second steps of the experiments, we focused exclusively on large herbivores.

Here, we first present the complete set of results for these experiments, including the results for all three faunal groups for the first step (see Fig. 7) and the second step (see Fig. 8). Results for the third step, already presented in the main text, are repeated for completeness (see Fig. 9).

Next we present the corresponding results with the other two null models. The first step does not depend on the chosen null model (see Fig. 7). Results from the second and third steps for the proportional-proportional model generated by the **UG** randomization algorithm can be seen in Figs. 10 and 11, respectively. Results from the second and third steps with the taxonomic shuffling method using the **shuffle** randomization algorithm can be seen in Figs. 12 and 13, respectively.

Then, we move on to co-occurrence statistics that look at species pairs, namely the  $C$ -score and mid- $P$  value.

Results for the  $C$ -score from the first step, i.e. distribution of values in the original datasets, for all three faunal groups and both continents can be seen in Fig. 14. Results from the second and third steps, i.e. respectively comparing these values to their distributions in null models and plotting the relative difference between observed and expected values against environmental context variable, are then presented for the three null models in turn: for the **Curveball** randomization algorithm in Figs. 15 and 16, for the **UG** randomization algorithm in Figs. 17 and 18, and for the **shuffle** randomization algorithm in Figs. 19 and 20.

Similarly, results with the mid- $P$  value from the first step can be seen in Fig. 21 while results from the second and third step can be seen in Figs. 22 and 23 for the **Curveball** randomization algorithm, in Figs. 24 and 25 for the **UG** randomization algorithm, and in Figs. 26 and 27 for the **shuffle** randomization algorithm.

Below, we describe the results in more details, focusing primarily on large herbivores as an example, since the same visual representations are used across the different groups.

## B.1 Computing co-occurrence statistics from the original data

Panels A and D in Fig. 21 show histograms of mid- $P$  value for pairs of species of different genera (orange) and of the same genus (yellow). In addition to the histograms showing the overall distribution of values, pairs with exceptional values (i.e. falling in bins that contain very few pairs) are represented by individual dots.

Most values are concentrated around 0.5, as well as, to a lower extent and only for some of the time units of North America, in the top of the unit interval (0.8–1). To an extent, the marked difference between the two continents might be related to the fact that Europe covers more localities, but with fewer occurrences per site and more localities which contain a single specimen (of the considered taxonomic group), cf. Table 2.

Overall, there are many species with very few, or even only one occurrence. For a pair of two such species (e.g.  $s_i = s_j = 1$ ), the size of the intersection can take only a few different values (at most the smallest of  $s_i$  and  $s_j$ ), and is expected to be zero. If the species do not co-occur ( $s_{ij} = 0$ ), then the mid- $P$  value is 0.5, if they do co-occur ( $s_{ij} = 1$ ) then the mid- $P$  value is close to 1. The latter might be considered as a signal of aggregation, but the former does not really indicate segregation, the absence of overlap is expected. Only pairs of fairly pervasive species, i.e. present in a large number of localities, can potentially result in mid- $P$  values close to 0 which might be interpreted as indicating segregation, but they are very rare. The distribution for pairs of different genera (orange) and of the same genus (yellow) look overall quite similar to each other. The latter involves much fewer pairs than the former, which makes the comparison more difficult.

Panels A and D in Fig. 14 show the distributions of  $C$ -scores for pairs of species of different genera (orange) and of the same genus (yellow). Note that the scale of the values differs for each time unit, scaling factors are indicated above the histograms, where relevant. Indeed, one way in which  $C$ -scores

differ from mid- $P$  values is that the maximum attainable value depends on the number of localities, and hence the values cannot be easily normalized and compared across time bins.

However, it is difficult to tell whether the observed values are specific to a considered time bin and related to some underlying process, or a straight-forward consequence of more general properties of the data. In addition, because different time bins contain different numbers of localities and of species, it is difficult to meaningfully compare values across time bins and across continents. For instance,  $C$ -scores are very low for Clarendonian-1 (Panel D in Fig. 14) due to the very limited number of localities.

## B.2 Comparing the values from the original data to null models

In order to alleviate this issue, we resort to a comparison with null models. Values observed across a number of randomized variants of the data provide a background distribution against which we can contrast the value observed in the original data. However, we cannot apply this process while considering the collection of scores for all pairs of species with each dataset. Instead, each dataset must be summarized by a single value, i.e. the scores must be aggregated. In addition to aggregating over all pairs, we can also consider specific subsets of interest within each dataset, such as the subset of pairs of species of the same genus and the subset of pairs of species of different genera, for instance.

First, we consider  $C$ -scores, which we aggregate by taking the average. Figs. 15–19 shows average  $C$ -scores for subsets of pairs in the original dataset (black crosses) compared to the corresponding distribution across one thousand null models (colored boxplots). The averages over whole datasets and over species pairs of different genera behave very similarly to each other, which is obviously expected, since this subset constitutes the overwhelming majority of pairs in each dataset, hence we only depict the latter. Also, due to taking the average over a more limited number of pairs, there is much more variance in values computed within the subsets consisting from pairs of the same genus as compared to the subsets consisting from pairs of different genera, or the whole collection of pairs.

Note that with taxonomic shuffling (`shuffle`) the collection of pairs for a given time unit is always the same, what varies is only whether a pair consists of species from the same or different genus, i.e. in which of the subsets it falls. Hence it does not make sense to compare averages across the whole dataset, since it is constant by definition of the randomization procedure.

There is greater variance in the values computed from the `UG` null models than from `Curveball` null model. This is expected, from the stricter requirements on the conservation of the margins (fixed rather than proportional) imposed by the latter.

In Europe, for both types of null models, we observe larger average  $C$ -scores and greater variance for the earlier time units (MN2–MN9, as well as MN12). The average  $C$ -score in the original data appears to be consistently below the null model values for the subset of pairs of different genera (with the exception of MN6, MN13, MN16 with `UG`), indicating that pairs of species tend to congregate more than expected at random. There does not seem to be a strong and consistent trend for pairs of the same genus, when comparing the original values to null model distribution, and to the other pairs.

In North America, Blacan-Early shows a very distinct pattern, with large average  $C$ -scores and very high variance for the subset of pairs of the same genus. This is probably linked to the fact that there are few distinct species per genus, and occurrences are dominated by a single taxon (see Fig. 2). There does not seem to be a consistent trend for subsets of pairs across other time bins.

Second, we consider mid- $P$  values, which we aggregate by computing the 95<sup>th</sup> percentile. Figs. 22–26 show the 95<sup>th</sup> percentile mid- $P$  value for subsets of pairs in the original dataset (black crosses) compared to the corresponding distribution across one thousand null models (colored boxplots). In both continents and for both types of null models, with few exceptions, the 95<sup>th</sup> percentiles mid- $P$  values in the original data are higher than in the null models, indicating the presence pairs of species with more extreme large mid- $P$  values, i.e. pairs with surprisingly large overlap. As mentioned earlier, because species tend to occur in very few localities, detecting segregated pairs is rather challenging. In particular, looking at the 5 percentile of mid- $P$  values was not conclusive.

By averaging  $C$ -scores, extreme values get diluted into the mass, and that statistic might reveal broad trends rather than specific behavior among a few individual pairs. Looking at 95<sup>th</sup> percentiles mid- $P$  values instead sets the focus on extreme cases. The statistic has a higher sensitivity and, especially for the

subset of pairs of the same genus, pairs with exceptional values can have a big impact on the aggregated statistic.

### B.3 Examining trends in the environmental context

Finally, we examine segregation trends in the environmental context, as captured by mean ordinated hypsodonty and number of species, in particular. To do so, we look at the direction and magnitude of the deviation between the value observed in the original data and the null models. Formally, denoting the value from the original data as  $x$  and the mean of the null model distribution as  $\mu$ , we compute the *relative difference* as  $\delta = (x - \mu)/\mu$ . Intuitively, this difference captures how far crosses are from the associated boxplots in Figs. 8–12, 15–19 and 22–26. A cross far below the boxplot, i.e. a case where the value observed in the original data is much lower than the expected value in the null model, corresponds to a large negative difference.

The relative difference between observed and expected average  $C$ -score versus environmental context variables is shown in Fig. 16. Each dot represents a time bin, with more recent times depicted in darker shades. As in the previous step, shown in Fig. 15, we separately consider two subsets of pairs. The values computed while considering only species pairs of different genera are depicted as smaller grey dots, whereas the values computed while considering only species pairs of the same genus as yellow–brown dots. Similarly, Fig. 23 shows the relative difference between observed and expected average  $C$ -score versus environmental context variables, again considering separately the pairs of species of different genera (in grey) and of the same genus (in yellow–brown).

Most yellow–brown dots in the panels of the top row of Fig. 16 are above the dotted line, whereas most grey dots lie below it. This suggests that, among large herbivores, pairs of species of the same genus are on average slightly more segregated than expected at random, whereas pairs of species of different genera are on average slightly less segregated than expected at random, as measured by the  $C$ -score. No strong trend is apparent over time, only a slight increase in segregation among congeneric species pairs of large herbivores in North America. On the other hand, mean ordinated hypsodonty clearly increases over time, whereas there is no clearly discernable trend in the number of species. Note that the North America data subset spans a longer time, starting from older time-units than the Europe data subset.

Looking at the top row of Fig. 23, instead, we see that most dots lie along the dotted line, with a few exceptions of yellow–brown points lying further away above or below the line. The 95<sup>th</sup> percentile is sensitive to the presence of extreme high values, whereas averaging tends to dilute the impact of extreme values, especially if there are extremes on both ends of the range that counter-balance each other. Hence, the 95<sup>th</sup> percentile mid- $P$  value and average  $C$ -score should be more suited for the detection of the presence of pairs of species with exceptional co-occurrence behavior and for capturing general trends, respectively.

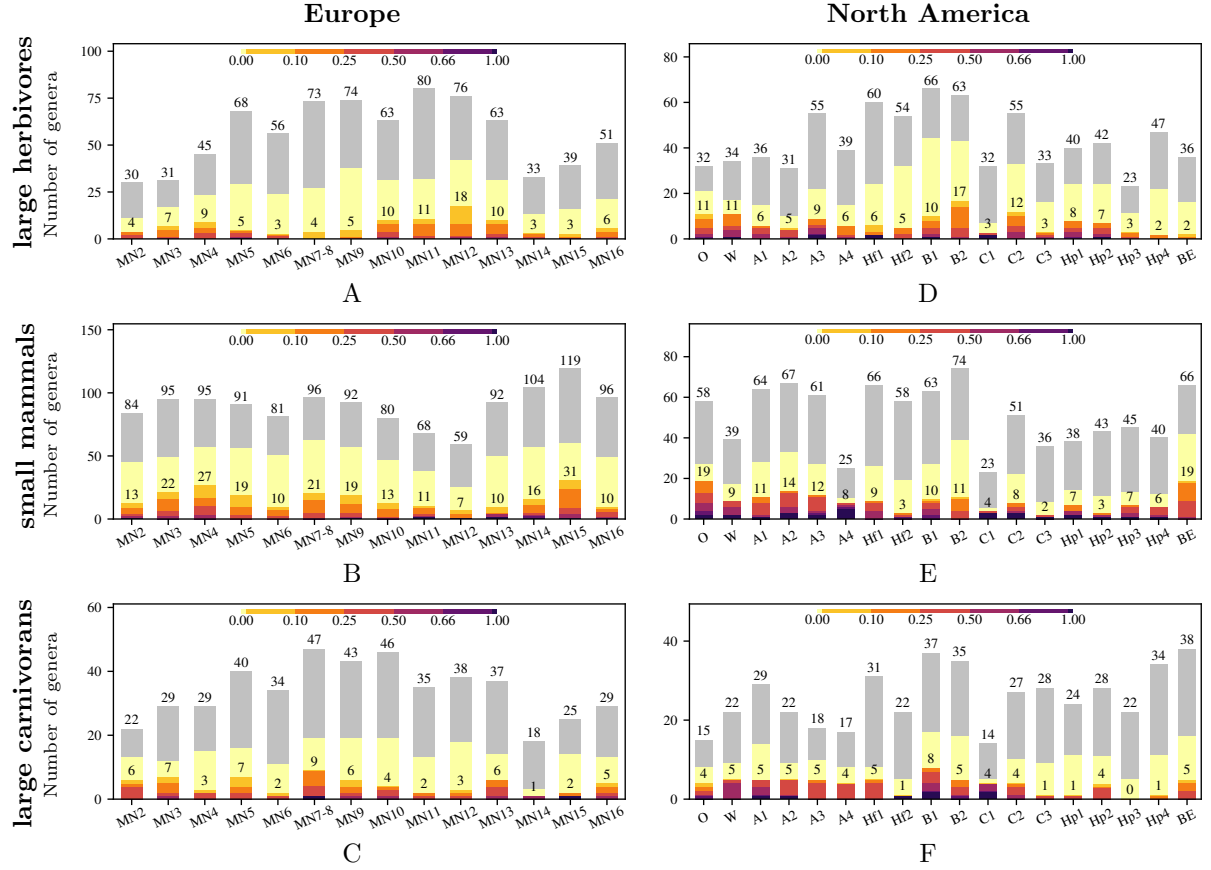


Figure 7: Counts of genera by fraction of multi-species occurrence localities for each time bin. The total number of genera in a time bin is indicated above each bar and the number of these genera having co-occurring species is indicated above the portion of the bar that represents them. Singleton genera are drawn in grey.

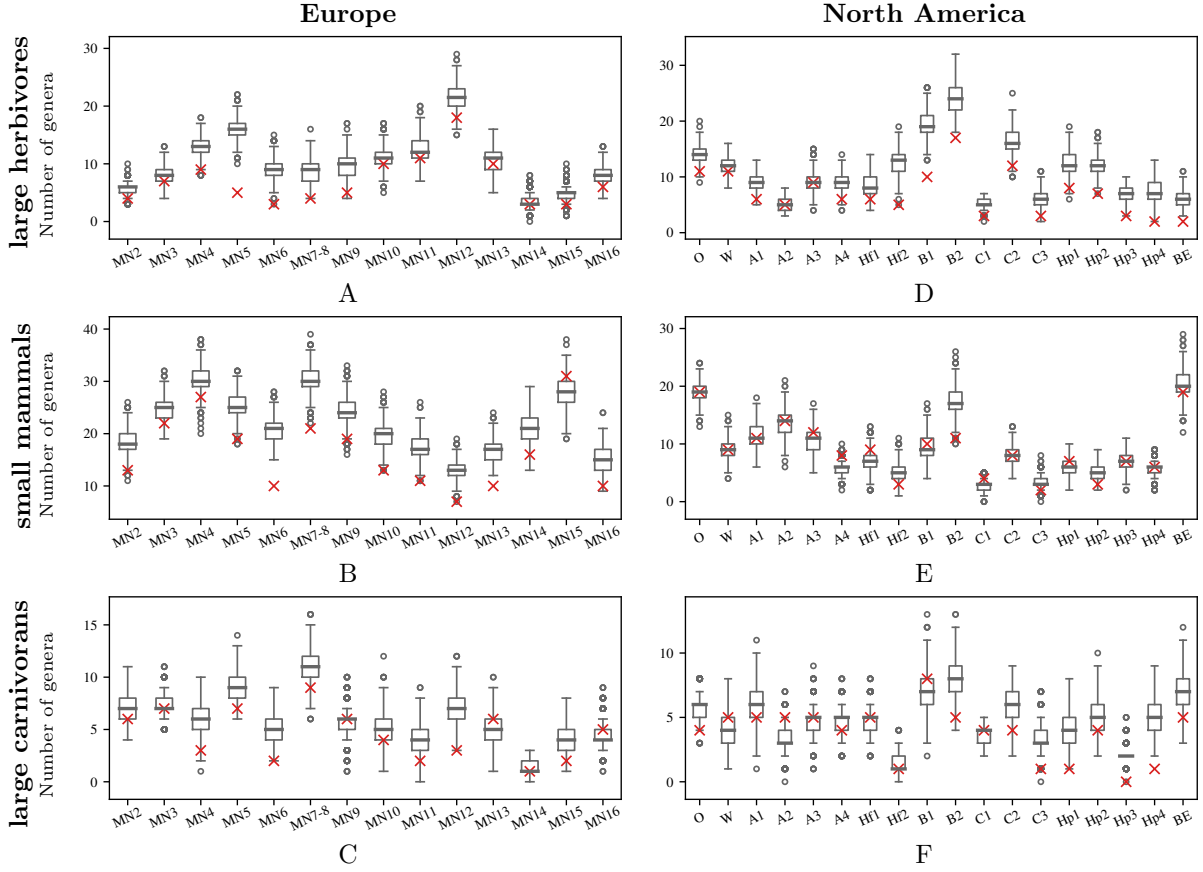


Figure 8: Number of genera with co-occurring species for each time bin. Comparing the value computed for the original data (red crosses) to its distribution across one thousand **Curveball** null models (grey boxplots).

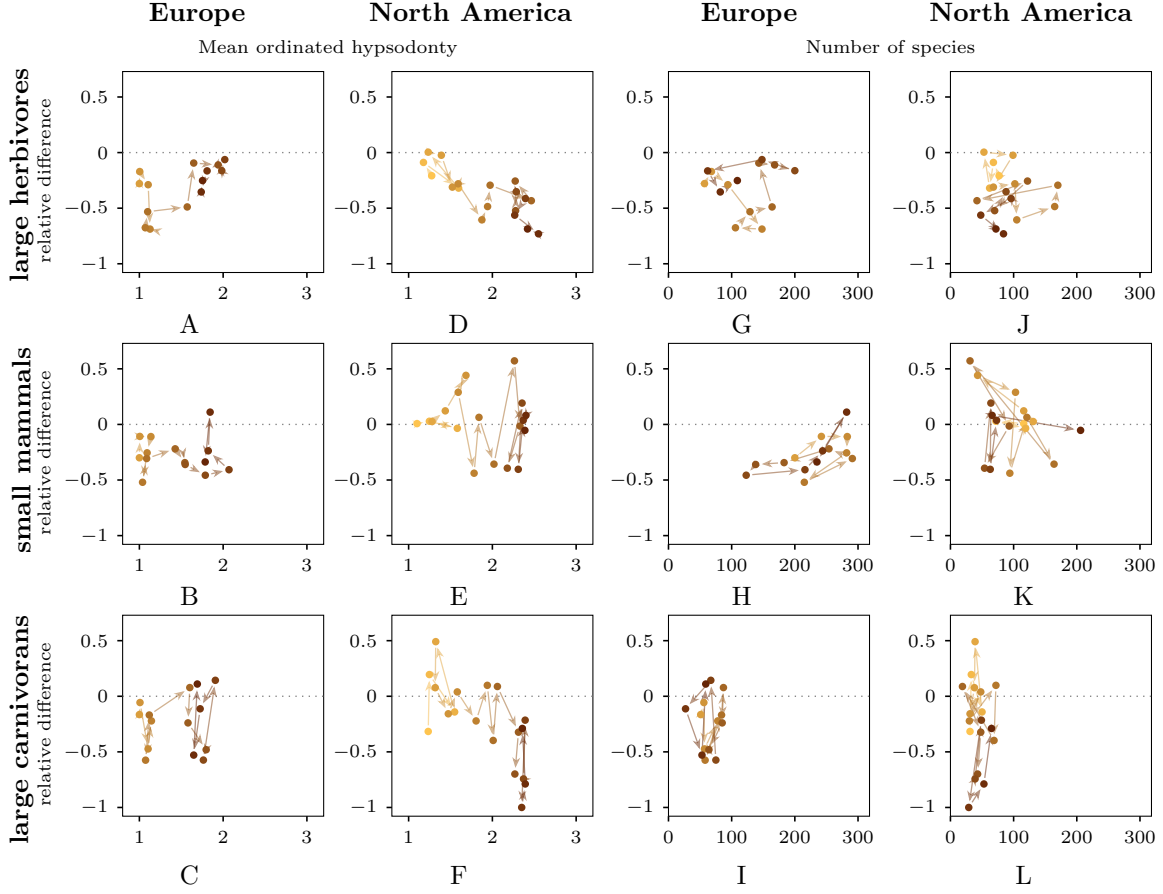


Figure 9: Relative difference between observed and expected number of genera with co-occurring species versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of brown and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the number of genera with co-occurring species computed from the original data and the *Curveball* null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger negative values indicate stronger segregation of congeneric species. Time bins span from 21.7 to 2.5 MA for Europe and from 33.9 to 2.6 Ma for North America.



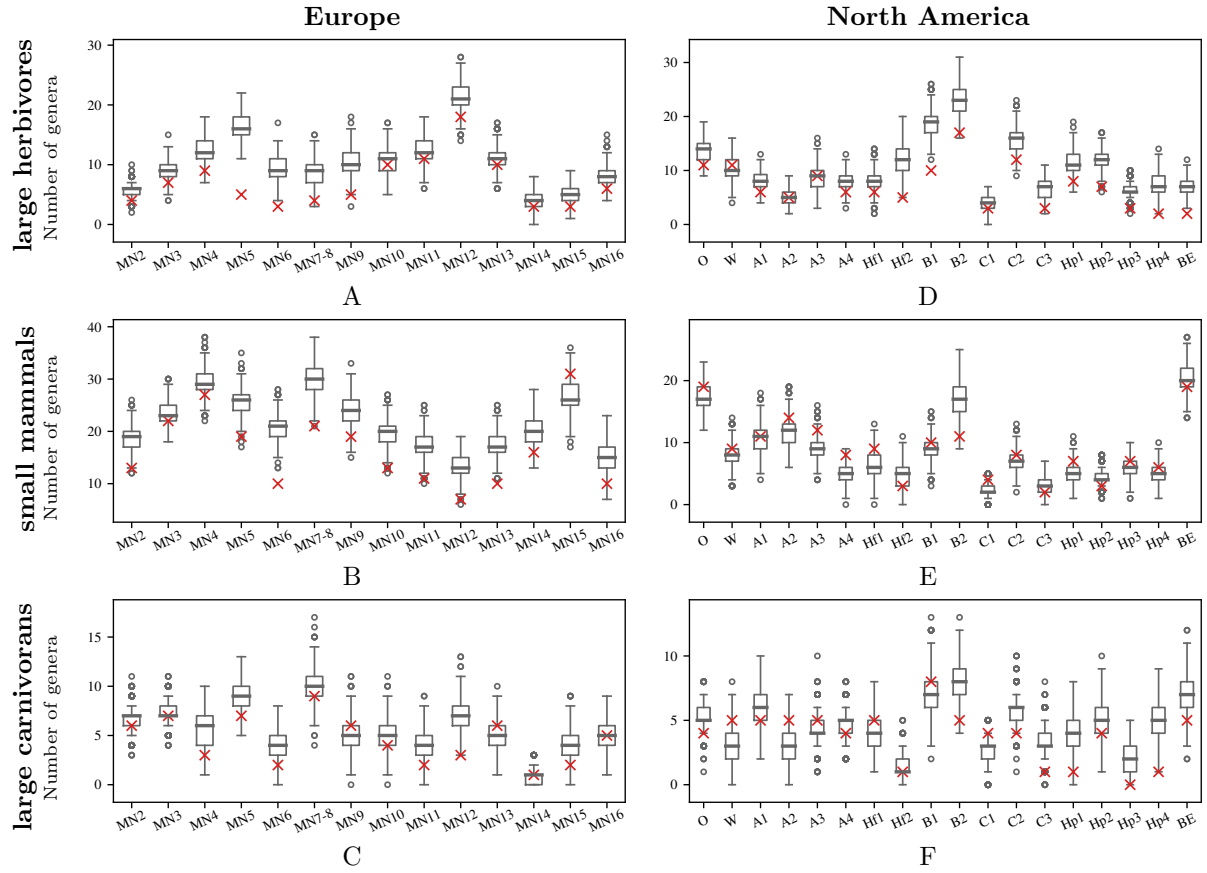


Figure 10: Number of genera with co-occurring species for each time bin. Comparing the value computed for the original data (red crosses) to its distribution across one thousand **UG** null models (grey boxplots).

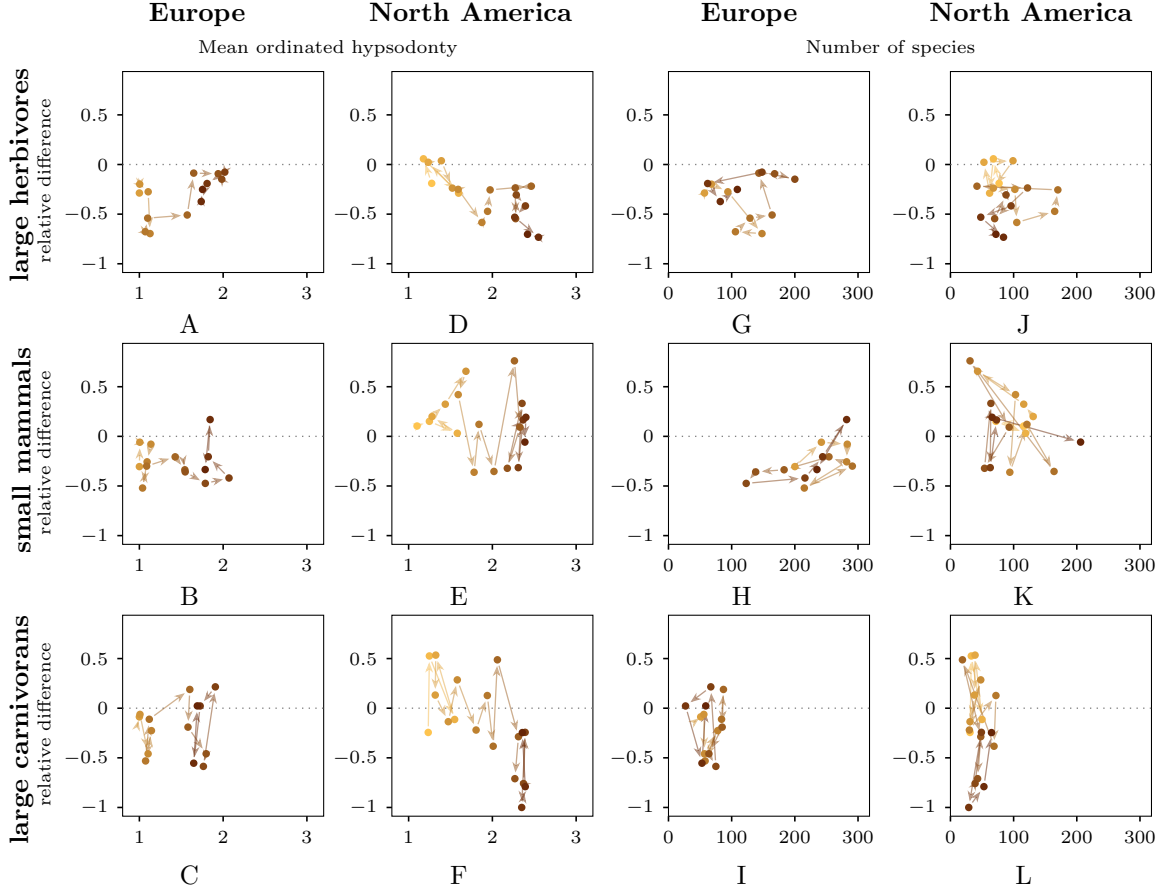


Figure 11: Relative difference between observed and expected number of genera with co-occurring species versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of brown and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the number of genera with co-occurring species computed from the original data and the UG null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger negative values indicate stronger segregation of congeneric species. Time bins span from 21.7 to 2.5 MA for Europe and from 33.9 to 2.6 Ma for North America.

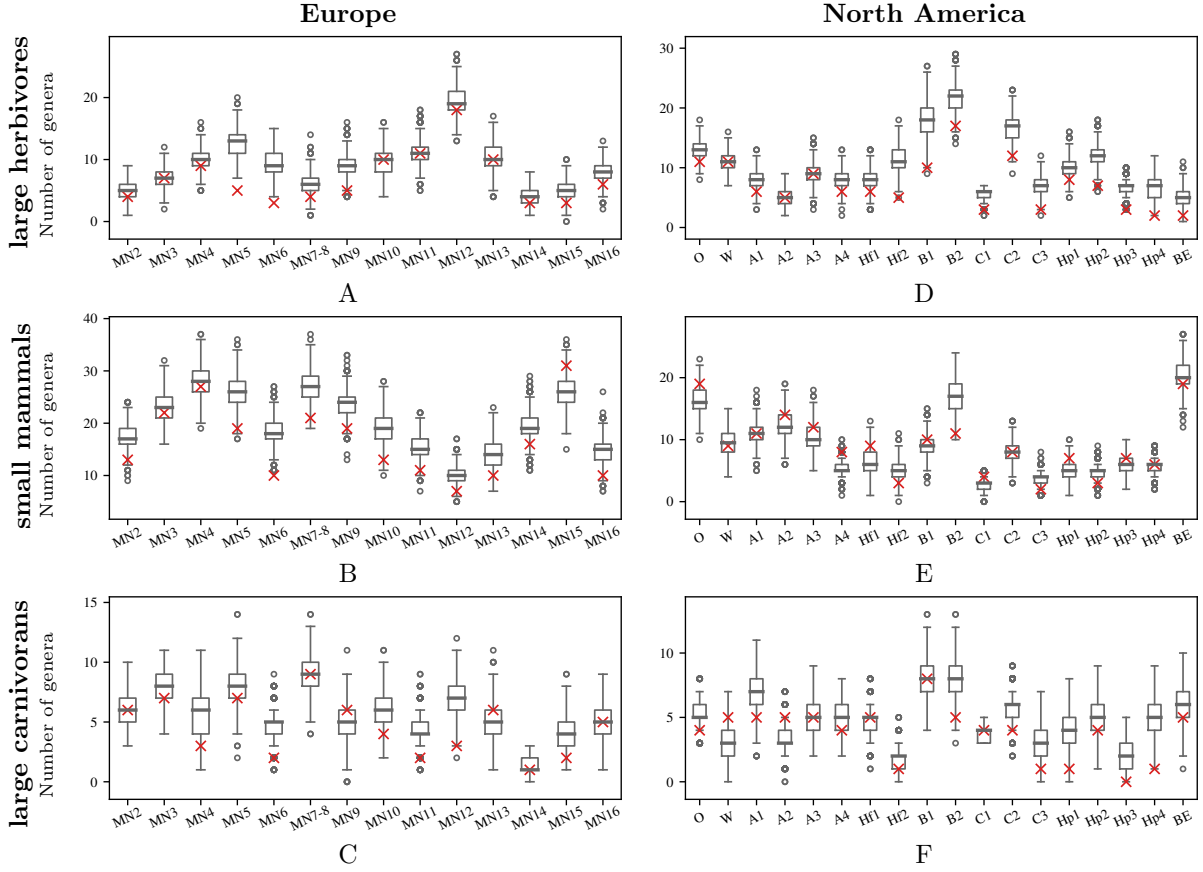


Figure 12: Number of genera with co-occurring species for each time bin. Comparing the value computed for the original data (red crosses) to its distribution across one thousand **shuffle** null models (grey boxplots).

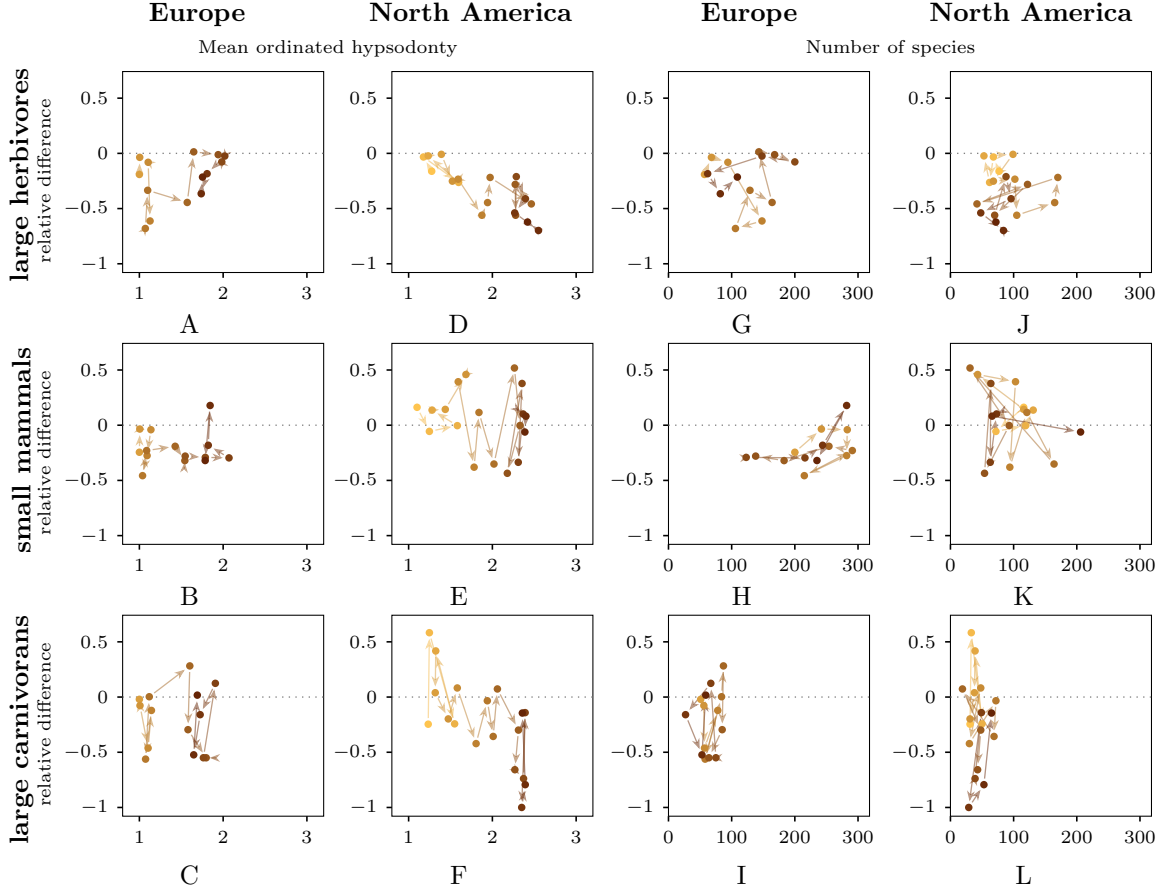


Figure 13: Relative difference between observed and expected number of genera with co-occurring species versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of brown and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the number of genera with co-occurring species computed from the original data and the `shuffle` null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger negative values indicate stronger segregation of congeneric species. Time bins span from 21.7 to 2.5 MA for Europe and from 33.9 to 2.6 Ma for North America.

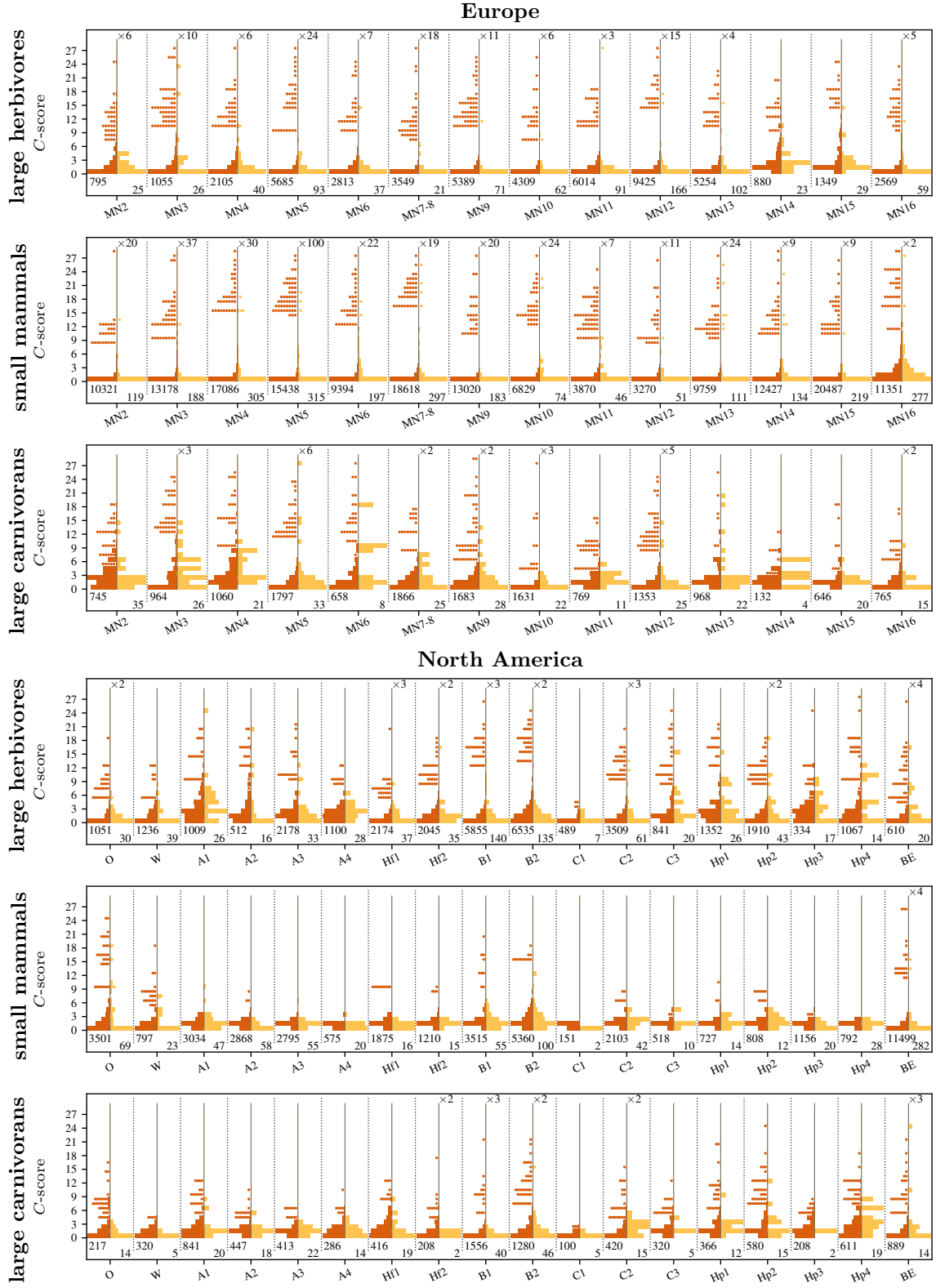


Figure 14: Distribution of  $C$ -score for pairs of species of different genera (orange) and of the same genus (yellow).

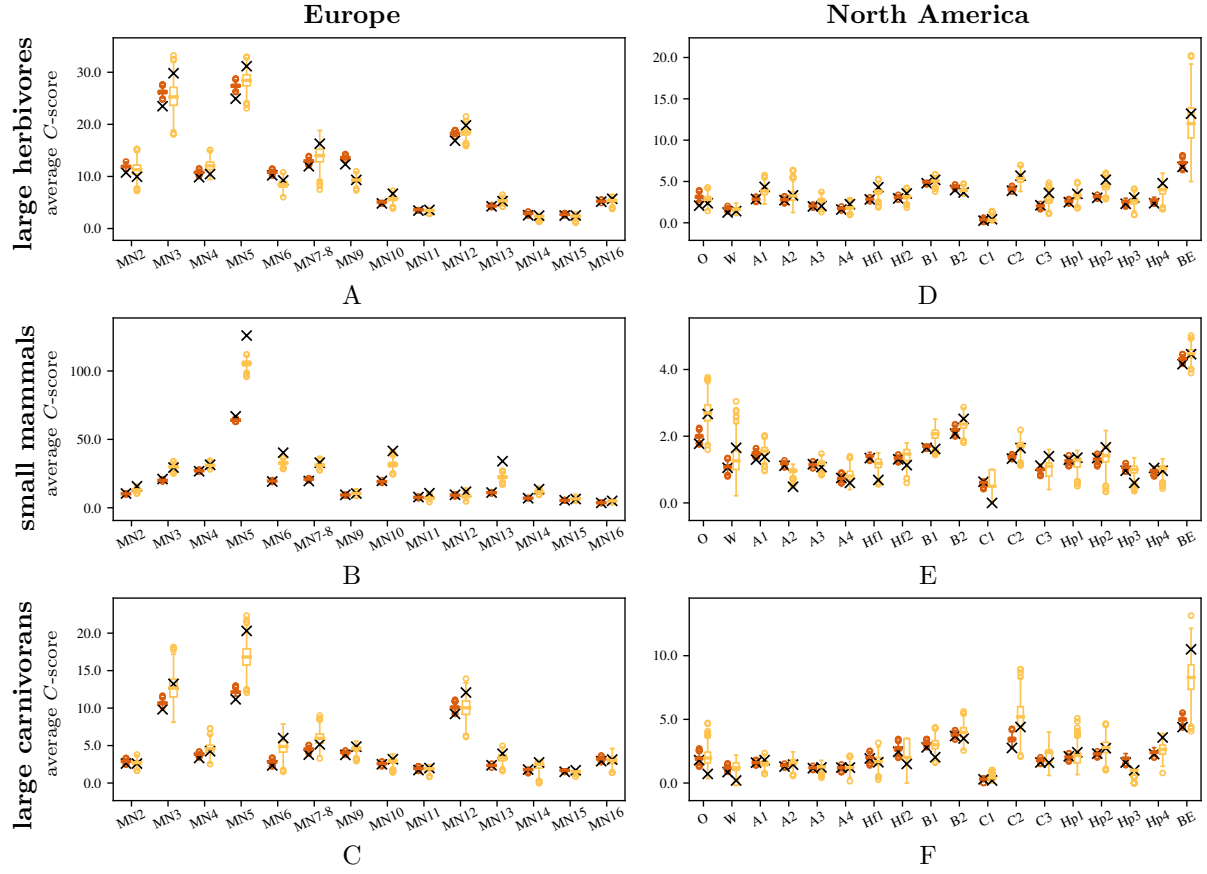


Figure 15: Comparing the average  $C$ -score in the original data (black crosses) to its distribution across one thousand *Curveball* null models, pairs of species of different genera (orange) and pairs of species of the same genus (yellow).

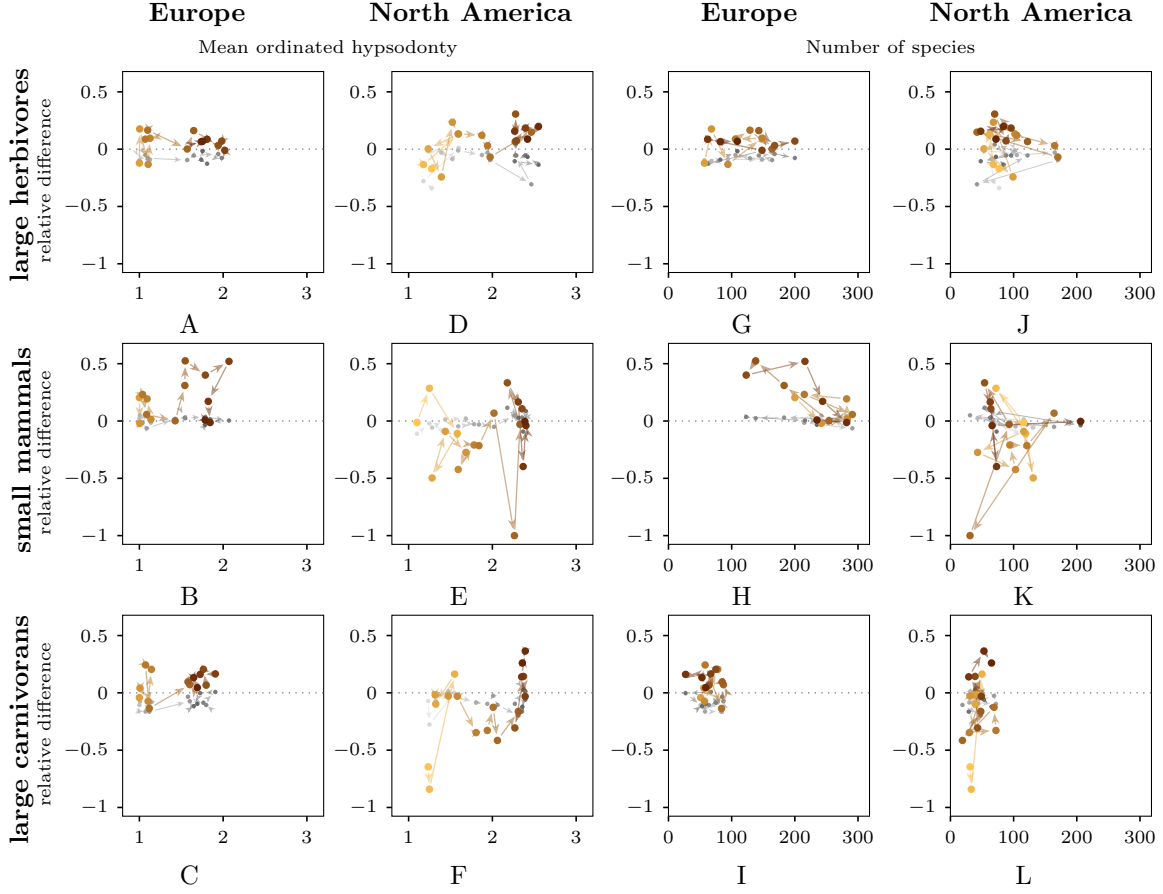


Figure 16: Relative difference between observed and expected average  $C$ -score versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of grey (pairs of species of different genera) or brown (pairs of species of the same genus), and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the average  $C$ -score computed from the original data and the **Curveball** null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger positive values indicate more segregated species pairs overall. Time bins span from 21.7 to 2.5 Ma for Europe and from 33.9 to 2.6 Ma for North America.

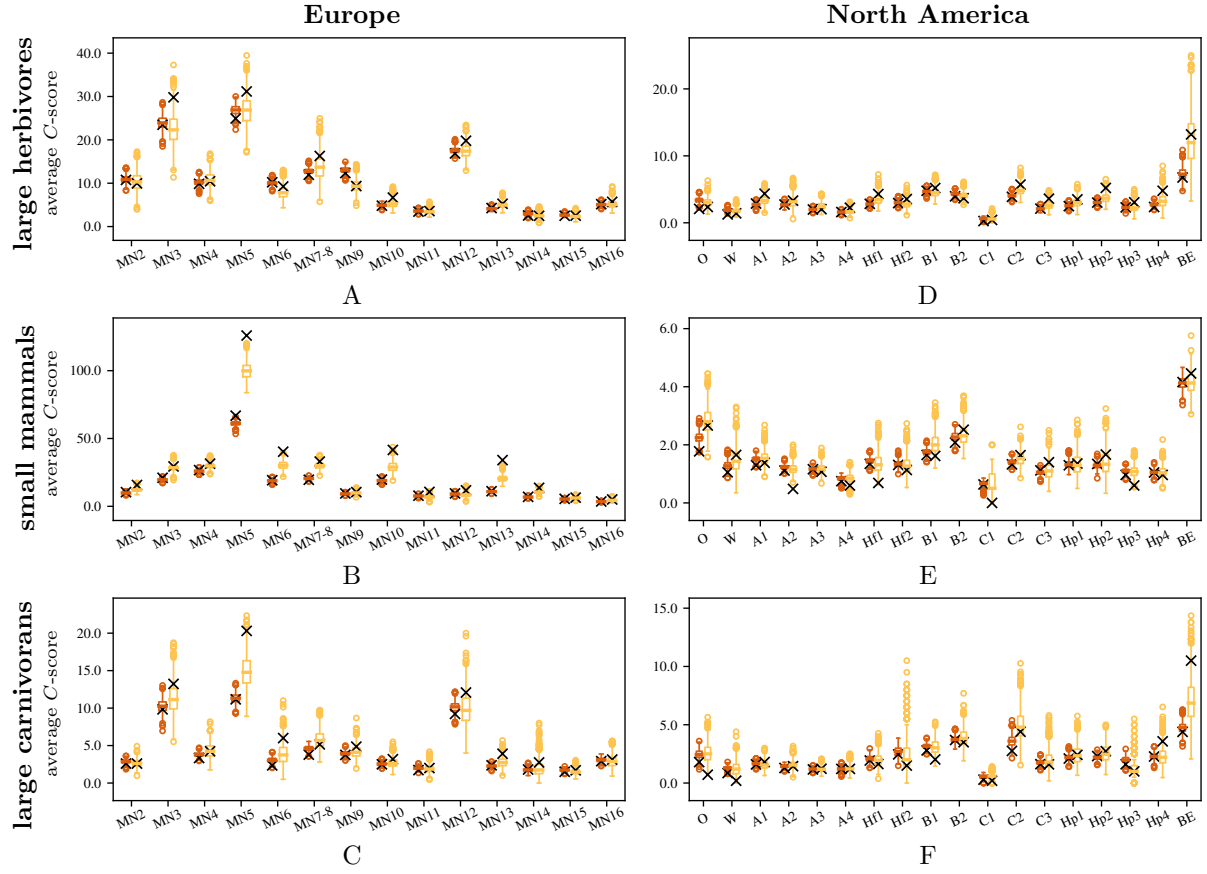


Figure 17: Comparing the average  $C$ -score in the original data (black crosses) to its distribution across one thousand  $UG$  null models, pairs of species of different genera (orange) and pairs of species of the same genus (yellow).



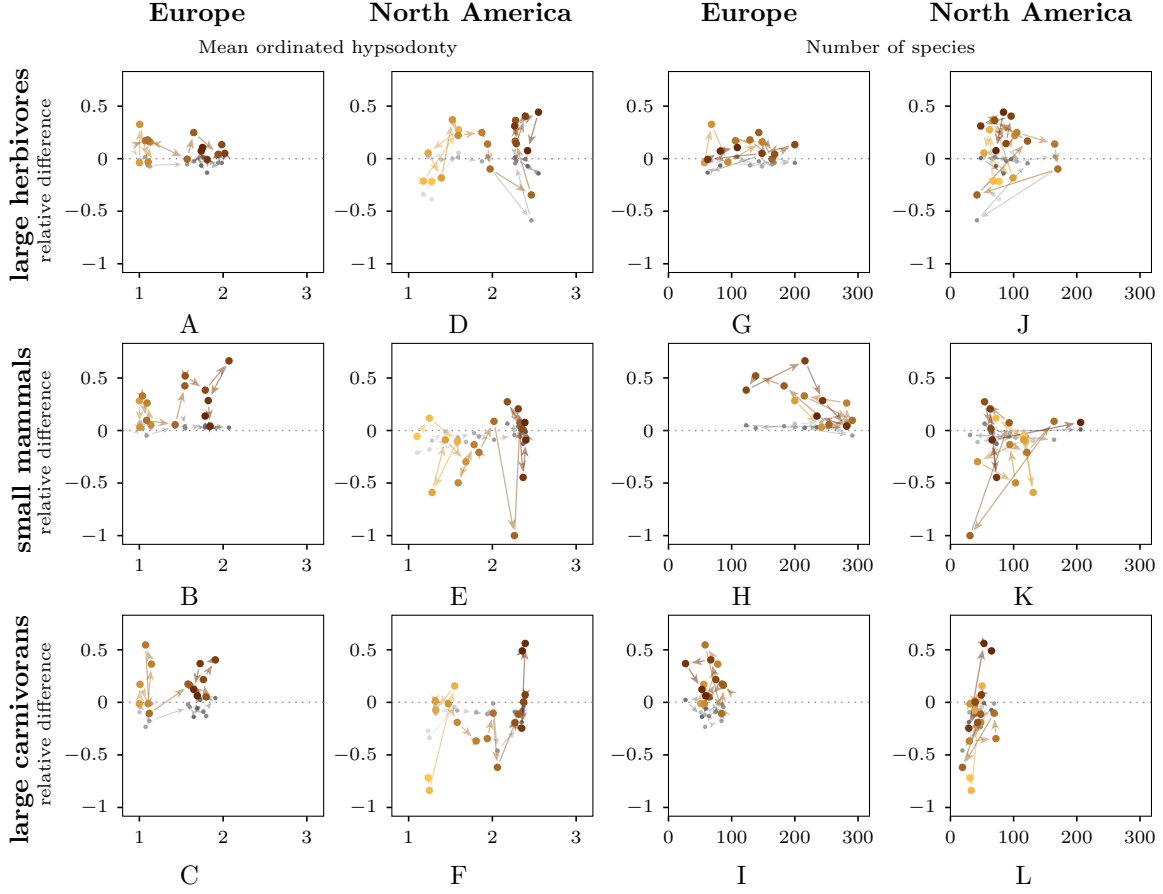


Figure 18: Relative difference between observed and expected average  $C$ -score versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of grey (pairs of species of different genera) or brown (pairs of species of the same genus), and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the average  $C$ -score computed from the original data and the  $UG$  null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger positive values indicate more segregated species pairs overall. Time bins span from 21.7 to 2.5 MA for Europe and from 33.9 to 2.6 Ma for North America.

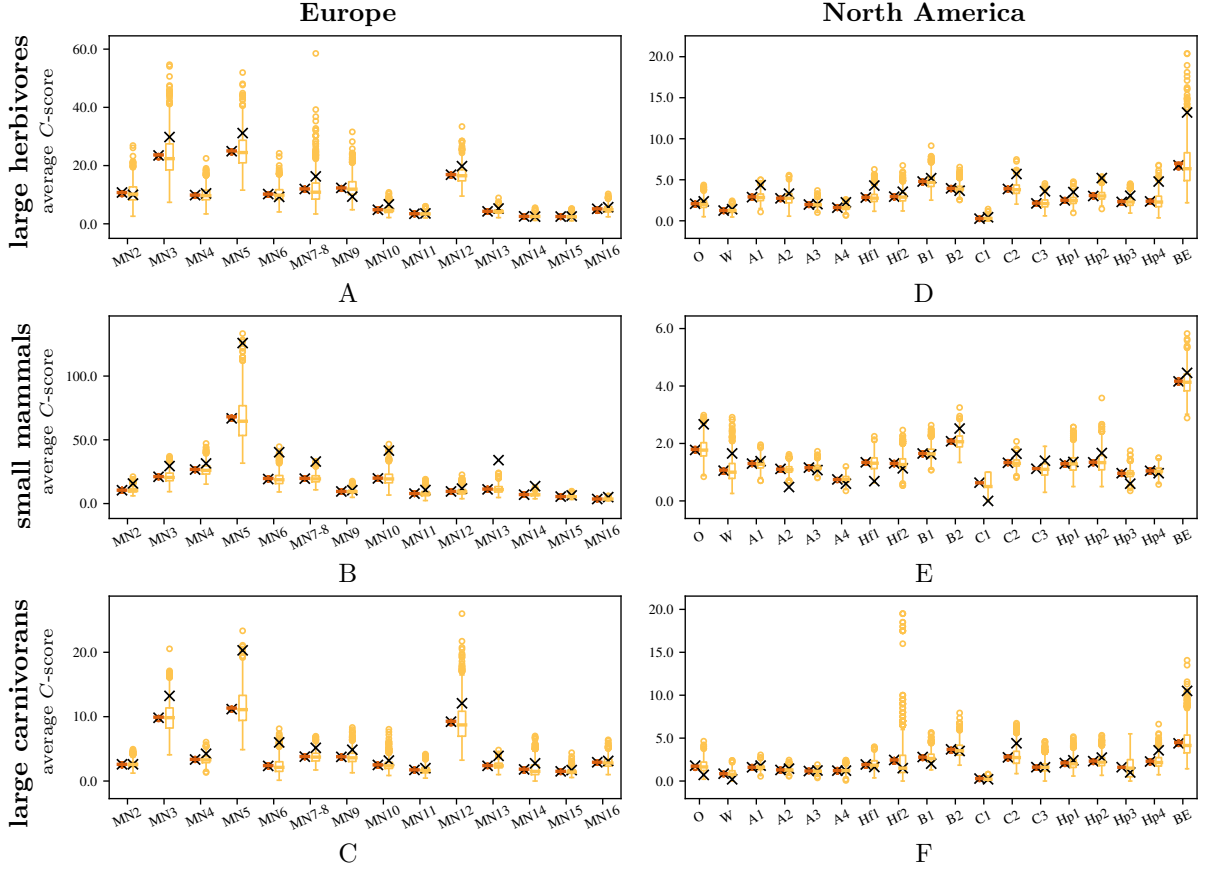


Figure 19: Comparing the average  $C$ -score in the original data (black crosses) to its distribution across one thousand `shuffle` null models, pairs of species of different genera (orange) and pairs of species of the same genus (yellow).

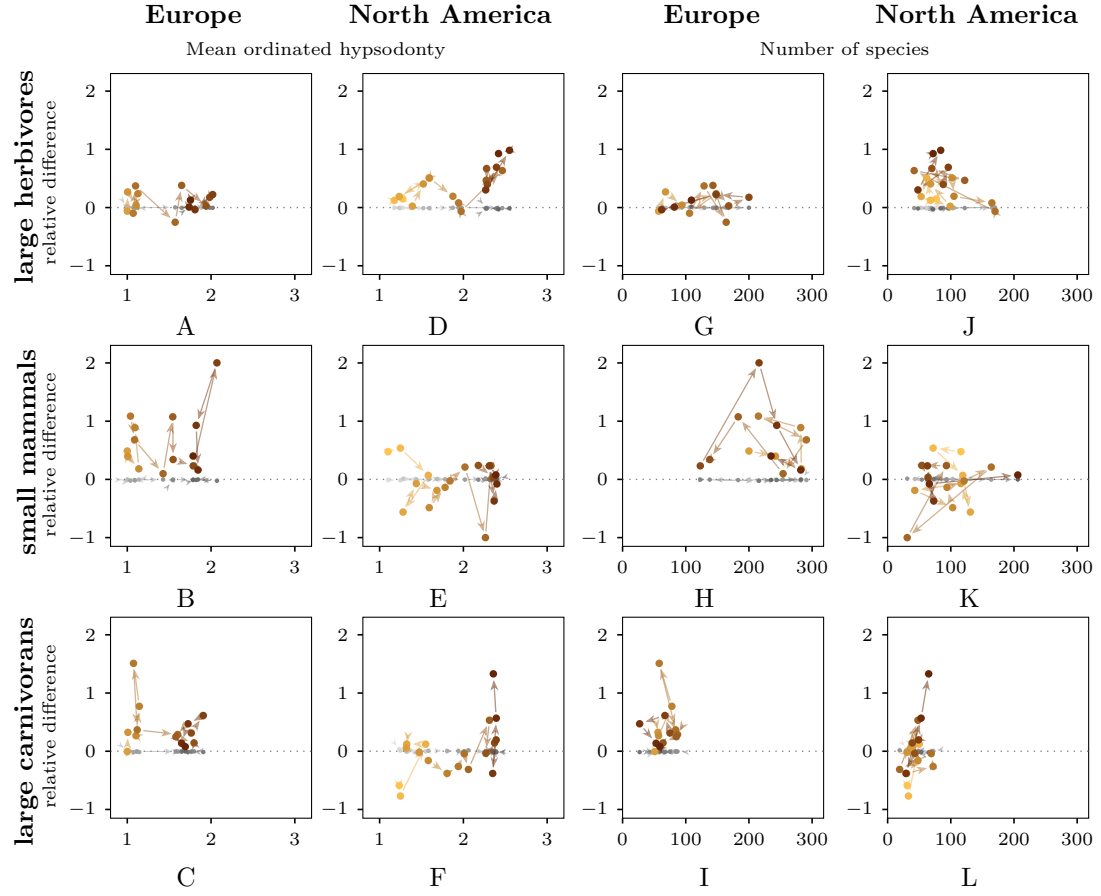


Figure 20: Relative difference between observed and expected average  $C$ -score versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of grey (pairs of species of different genera) and brown (pairs of species of the same genus), and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the average  $C$ -score computed from the original data and the `shuffle` null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger positive values indicate more segregated species pairs overall. Time bins span from 21.7 to 2.5 Ma for Europe and from 33.9 to 2.6 Ma for North America.

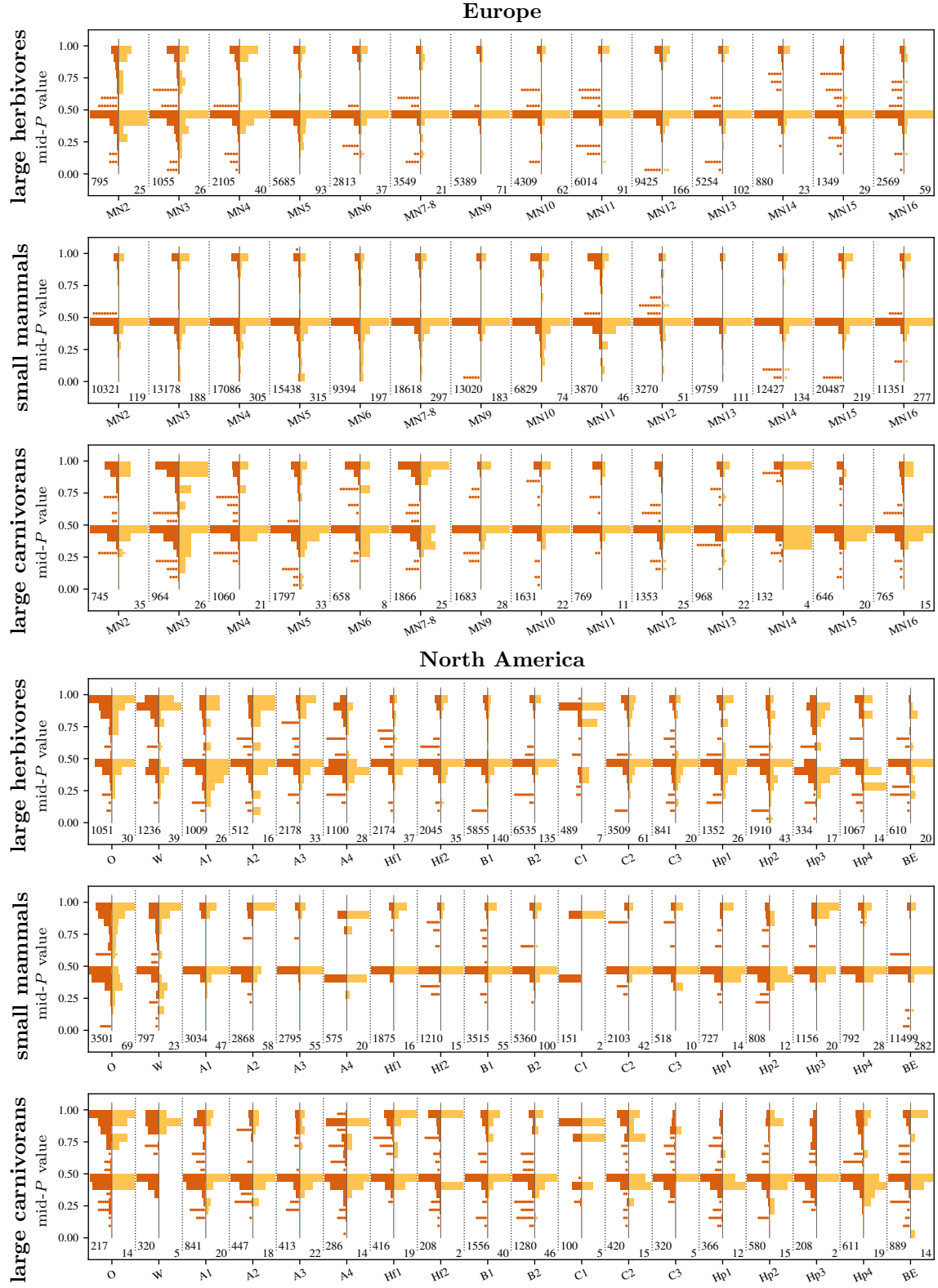


Figure 21: Distribution of mid- $P$  value for pairs of species of different genera (orange) and of the same genus (yellow).

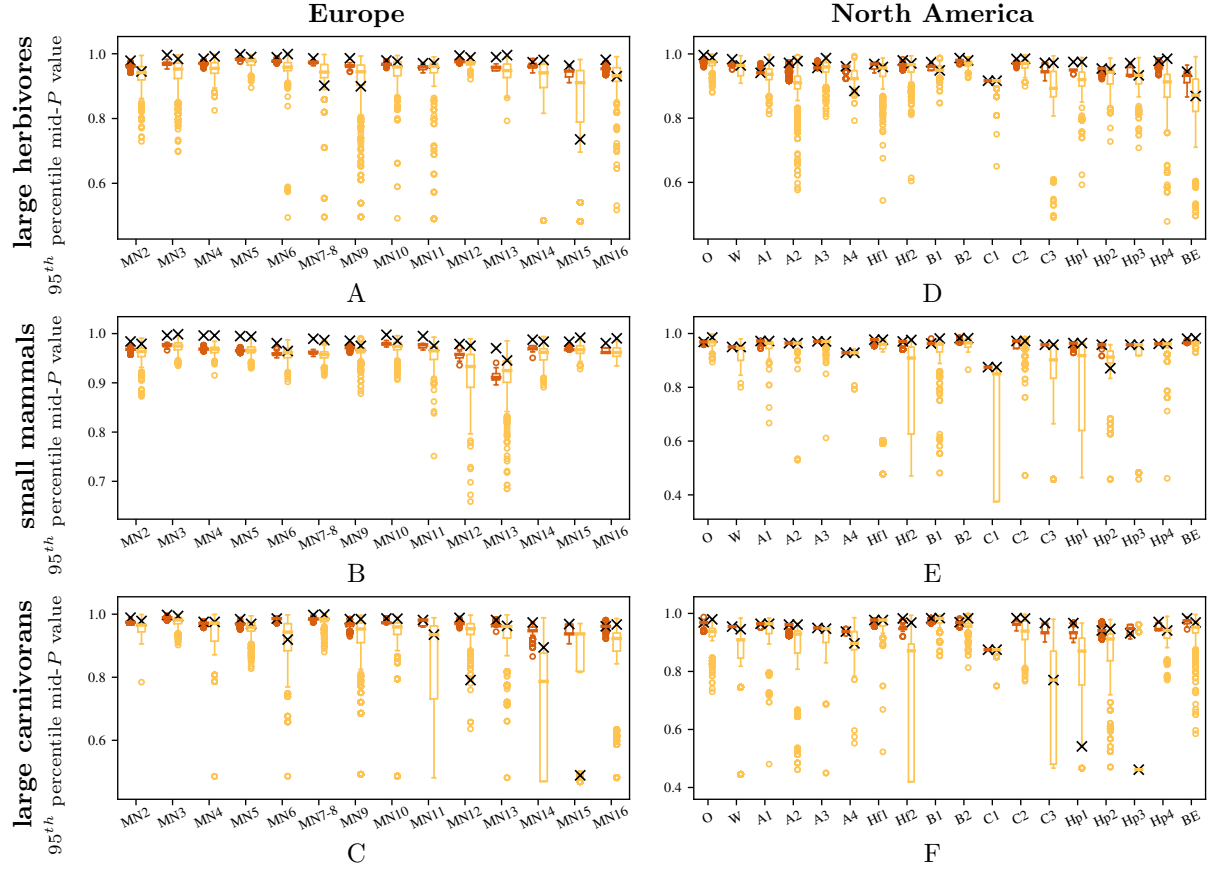


Figure 22: Comparing the 95<sup>th</sup> percentile mid-*P* value in the original data (black crosses) to its distribution across one thousand **Curveball** null models, pairs of species of different genera (orange) and pairs of species of the same genus (yellow).

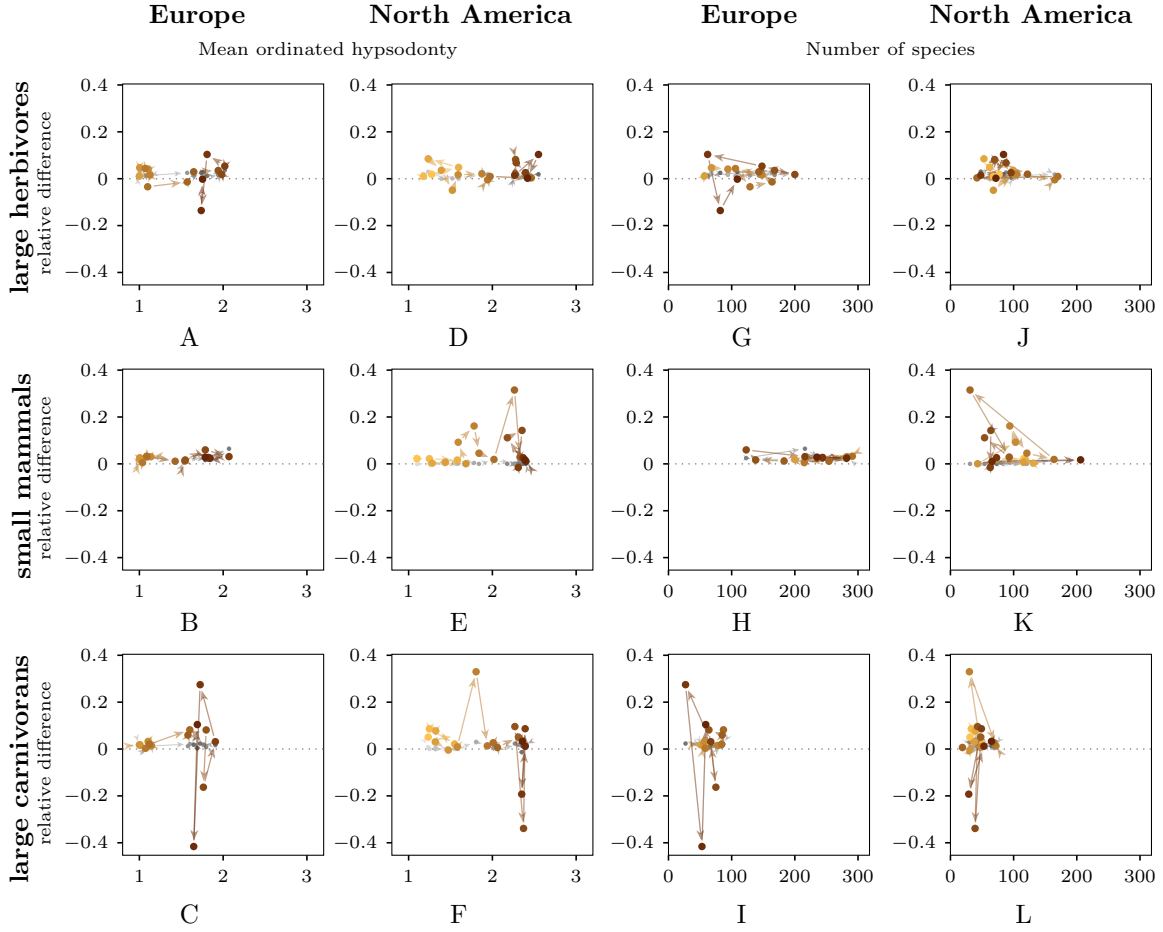


Figure 23: Relative difference between observed and expected 95<sup>th</sup> percentile mid- $P$  value versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of grey (pairs of species of different genera) or brown (pairs of species of the same genus), and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the 95<sup>th</sup> percentile mid- $P$  value computed from the original data and the **Curveball** null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger positive values indicate the presence of strongly aggregated species pairs. Time bins span from 21.7 to 2.5 Ma for Europe and from 33.9 to 2.6 Ma for North America.

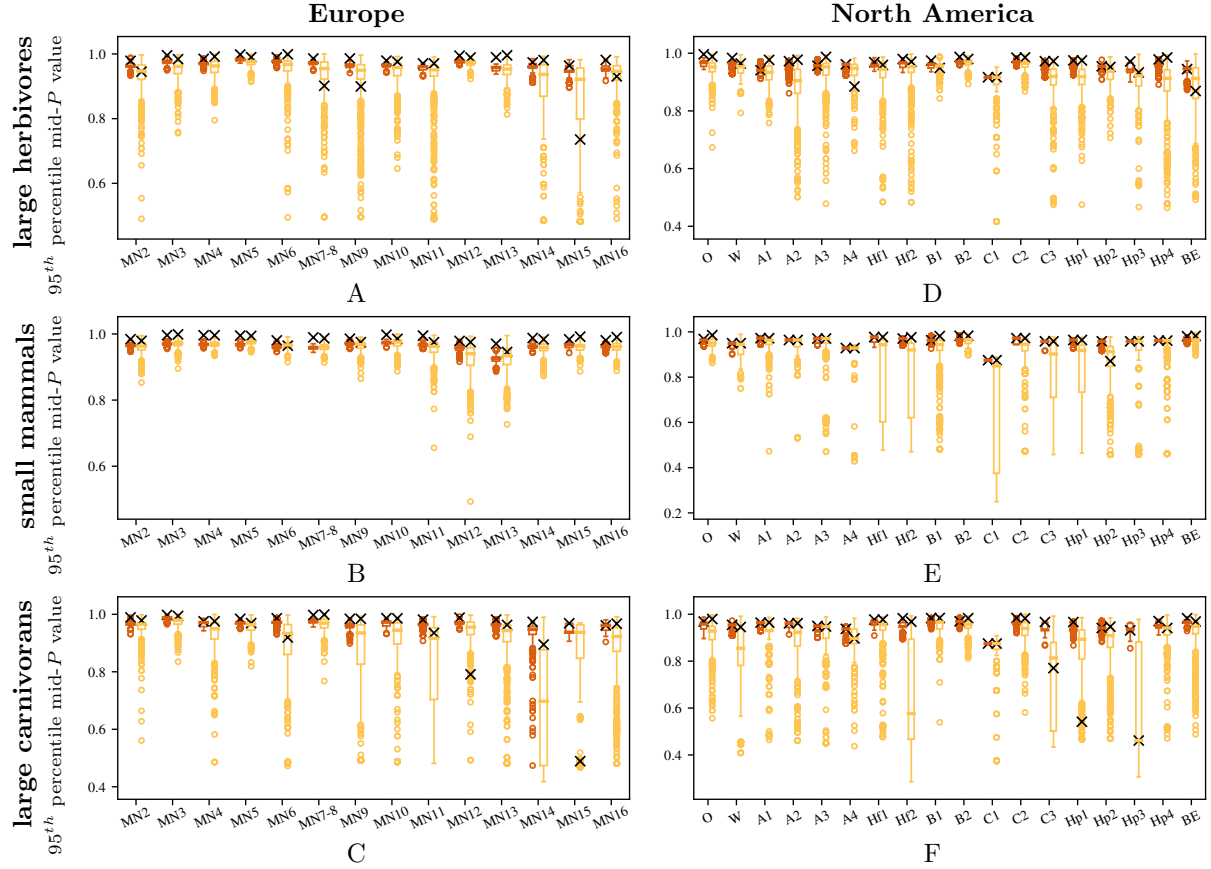


Figure 24: Comparing the 95<sup>th</sup> percentile mid- $P$  value in the original data (black crosses) to its distribution across one thousand UG null models, pairs of species of different genera (orange) and pairs of species of the same genus (yellow).

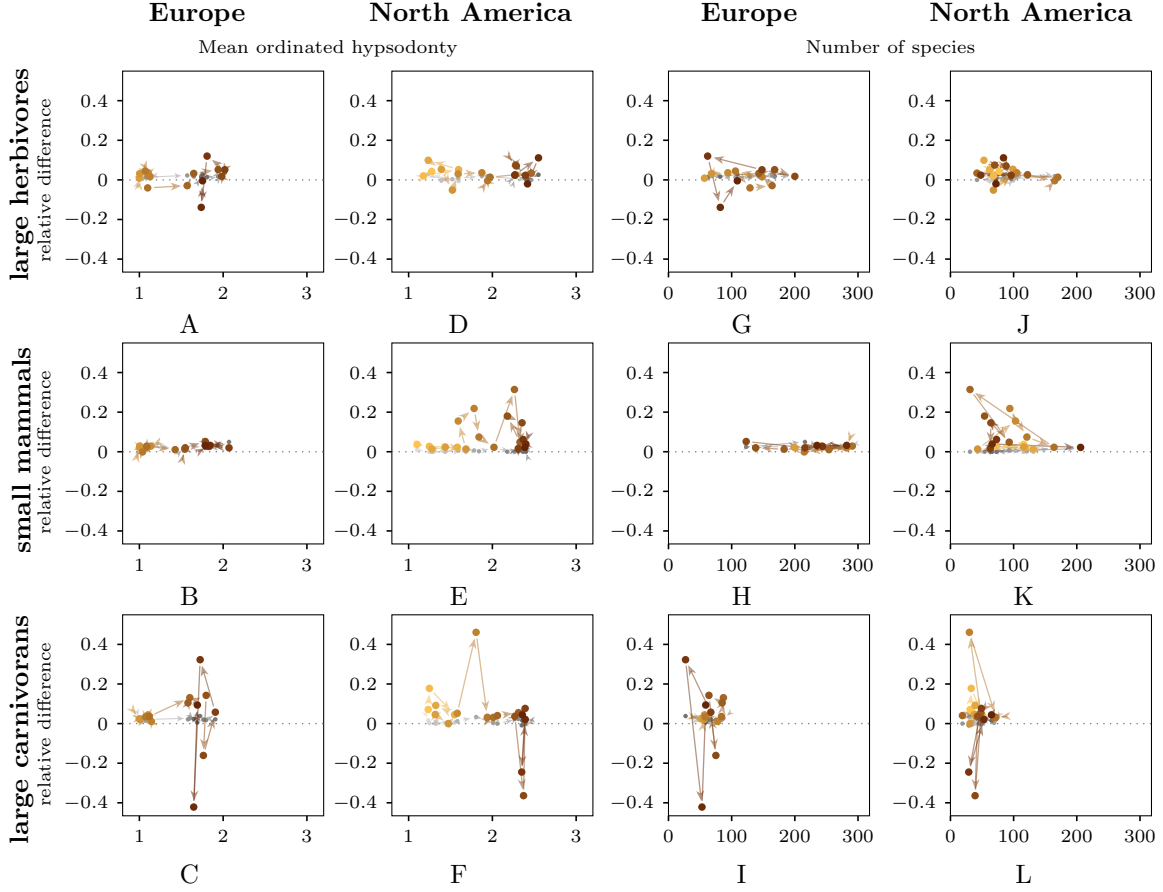


Figure 25: Relative difference between observed and expected 95<sup>th</sup> percentile mid- $P$  value versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of grey (pairs of species of different genera) or brown (pairs of species of the same genus), and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the 95<sup>th</sup> percentile mid- $P$  value computed from the original data and the **UG** null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger positive values indicate the presence of strongly aggregated species pairs. Time bins span from 21.7 to 2.5 Ma for Europe and from 33.9 to 2.6 Ma for North America.



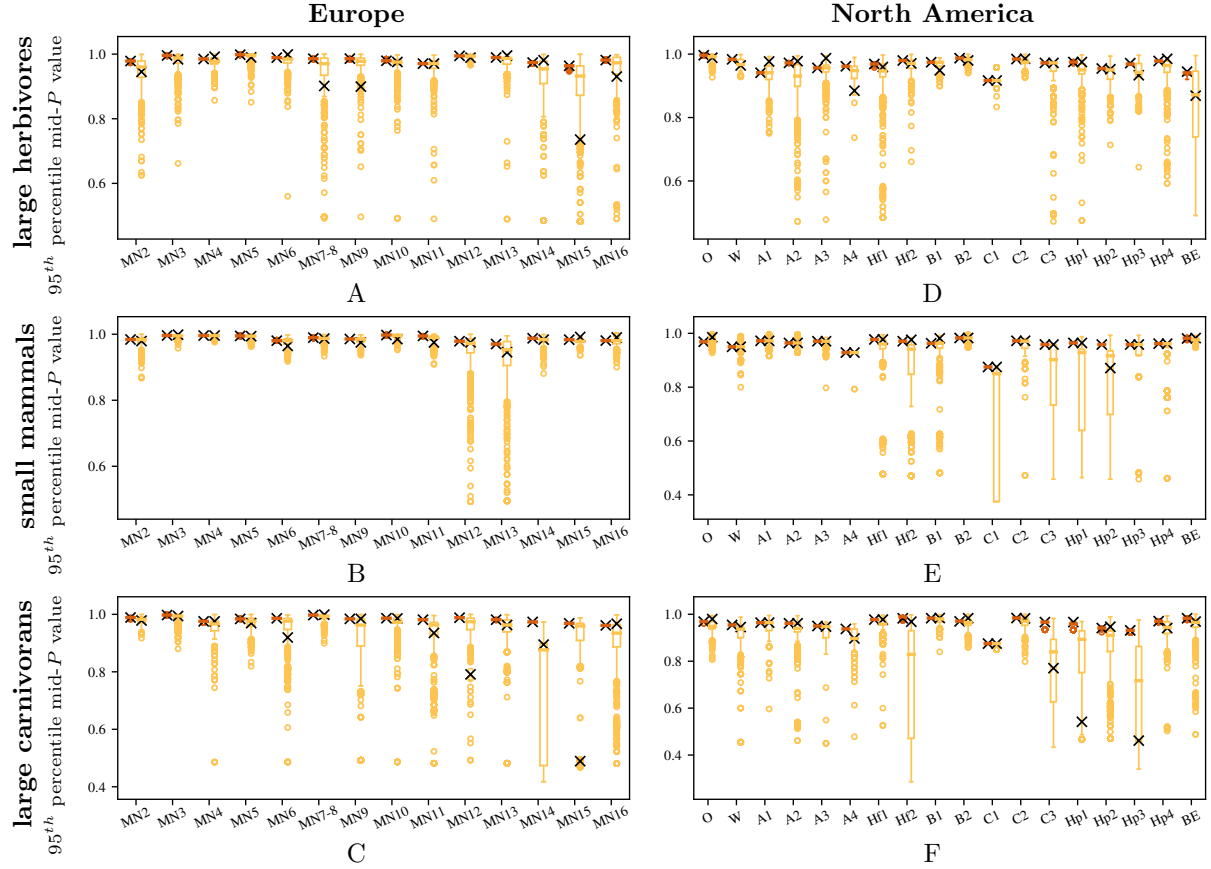


Figure 26: Comparing the 95<sup>th</sup> percentile mid-*P* value in the original data (black crosses) to its distribution across one thousand **shuffle** null models, pairs of species of different genera (orange) and pairs of species of the same genus (yellow).

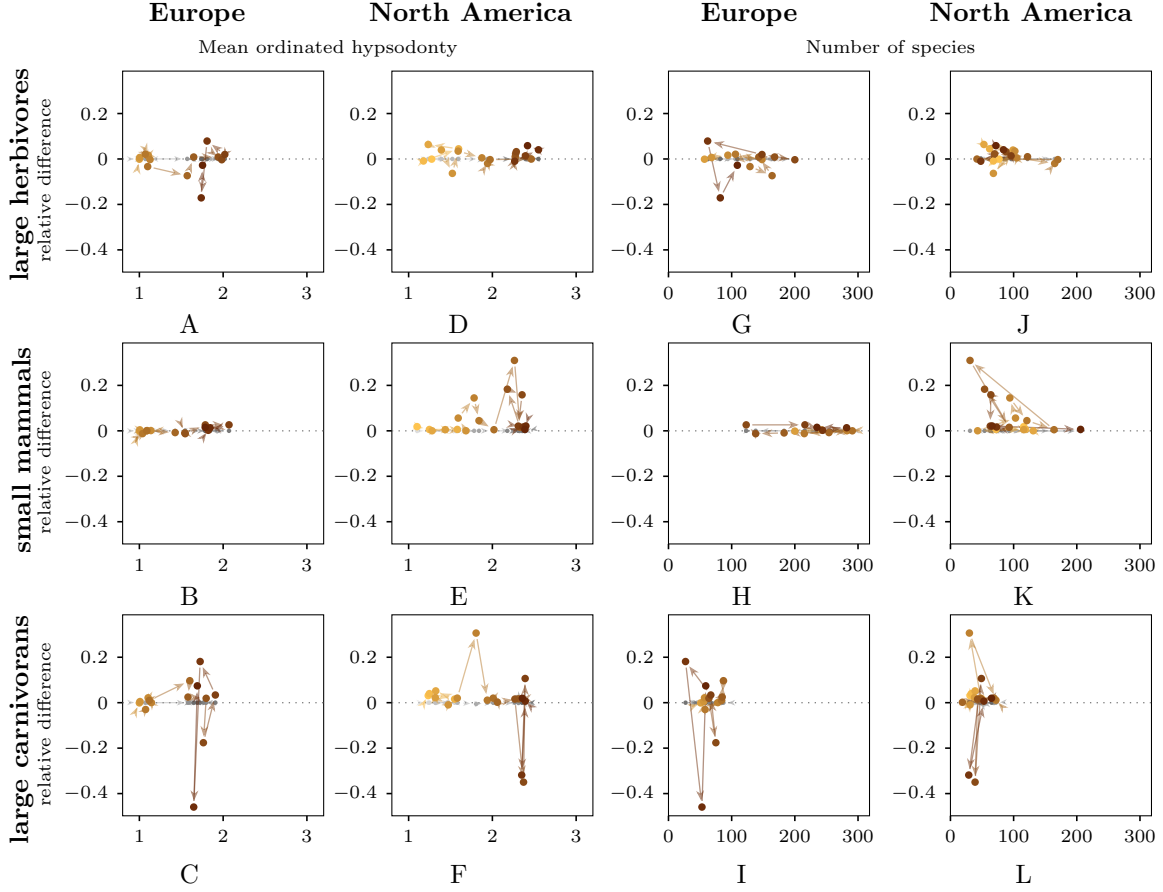


Figure 27: Relative difference between observed and expected 95<sup>th</sup> percentile mid- $P$  value versus environmental context variables, respectively mean ordinated hypsodonty (panels A–F) and number of species (panels G–L), over time. Each dot represents a time bin, with more recent times depicted in darker shades of grey (pairs of species of different genera) or brown (pairs of species of the same genus), and arrows going from the oldest to most recent time bins. The vertical axis represents the relative difference between the 95<sup>th</sup> percentile mid- $P$  value computed from the original data and the `shuffle` null model expected value. The dashed horizontal line in each panel corresponds to equality between the observed and expected values, indicating no segregation pattern. Larger positive values indicate the presence of strongly aggregated species pairs. Time bins span from 21.7 to 2.5 Ma for Europe and from 33.9 to 2.6 Ma for North America.