

Supplementary material

Appendix S1. Detailed description on sampling

In each of the four forest types, we established 20 evenly spaced plots along a transect line of 400 m according to the ALL-protocol (Agosti et al. 2000). In limestone and alluvial forest, we established a second transect line of 200 m length in these forest types, where ten additional samples were collected. The second alluvial transect was established 2 km away from the first alluvial transect, while the second limestone transect was 14 km away from the first one; for the all locations see Fig. 1 in this article. We used a metal frame with the size of 1.0 to 1.0 m to mark the plot and to reduce the numbers of fleeing ants during sampling. At each sample site, we collected leaf litter and soil inside the frame separately and concentrated the collected material by sifting it with a metal sieve (mesh size 12 mm) thus removing larger particles from the sampled material. We collected the soil to the depth where a colour change of the soil signalized the end of the top soil layer.

The sieved matter was put into canvas bags for transport and extracted with Winkler-bags, separately for each plot and layer. The high air humidity in Gunung Mulu NP made it necessary to hang up the Winkler-bags in an air-conditioned place where they remained for seven days. Arthropods leaving the soil were automatically collected and stored in 70% ethanol. As a control of extraction efficiency we checked ten percent of the soil samples for remaining arthropods after processing; in none of them more than one percent of the ant individuals remained.

Table S1. Spatial autocorrelation of environmental parameters given as values of Moran's I calculated with the "gearymoran" function of R. Shown are Moran's I coefficients, measured for pooled data, as well as for single transects. Especially pooled data (ALL) showed significant negative spatial autocorrelations. Parameters that were measured only on a regional basis were not investigated at local level. Stars give the significant levels after false discovery rate correction.

Tested parameters	ALL		Alluvial 1		Alluvial 2		Limestone 1		Limestone 2		Kerangas		Dipterocarp	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
PCA topsoil (2)	-0.22**	-21.18	-0.08	-2.14	-0.12	-0.25	-0.05	-0.07	-0.13	-1.47	-0.07	-1.11	-0.05	0.17
Living trees	-0.08**	-6.71	-0.07	-1.71	-0.08	1.30	-0.06	-0.53	-0.15	-1.87	-0.05	1.20	-0.06	-0.22
Leaf litter coverage	-0.04	-2.29	-0.05	0.35	-0.11	-0.04	-0.07	-1.80	-0.10	0.67	-0.06	-0.30	-0.06	-0.66
Vegetation points	-0.11**	-8.98	-0.05	-0.11	-0.10	0.31	-0.06	-0.89	-0.12	-0.40	-0.06	0.30	-0.05	0.47
Carbon/nitrogen content	-0.21**	-18.54	-0.07	-1.80	-0.09	1.04	-0.04	2.08	-0.11	-0.13	-0.06	-0.64	-0.05	0.81
PCA roots (3)	-0.09**	-7.67	-0.08	-2.08	-0.14	-1.45	-0.07	-2.24	-0.09	1.18	-0.05	0.79	-0.08	-2.11
PCA stone (1)	-0.07**	-5.60	-0.07	-1.82	-0.11	1.20	-0.04	1.37	-0.10	0.42	-0.06	0.53	-0.05	0.41
PCA decay (4)	-0.31**	-27.34												
Dead trees	-0.01	-0.13	-0.05	0.50	-0.13	-0.66	-0.06	-0.83	-0.13	-1.44	-0.07	-0.74	-0.05	0.01
Number of invertebrates	-0.03	-1.83	-0.04	1.17	-0.11	-0.14	-0.05	0.46	0.00	0.00	-0.07	-1.38	-0.06	-0.60
Phosphate	-0.04	-3.16												
Slope angle	-0.07**	-5.73	-0.06	-0.82	-0.12	-1.27	-0.06	-1.59	-0.09	0.99	-0.10**	-3.58	-0.06	-0.41
Temperature topsoil	-0.47**	-42.52												
Ph value	-0.40**	-36.90	-0.06	-2.99	-0.11	0.01	-0.09**	-5.66	-0.12	-1.18	-0.04	1.93	-0.07	-1.95
Leaf litter thickness	-0.03	-2.08	-0.07	-1.58	-0.09	1.14	-0.06	-1.16	-0.14	-1.35	-0.10**	-4.42	-0.06	-0.35
Canopy openness	-0.01	-0.35	-0.04	0.93	-0.08	1.32	-0.06	-0.58	-0.10	0.55	-0.05	1.13	-0.05	0.77
Dead wood < 5 cm	-0.03	-1.86	-0.09	-3.36	-0.10	0.54	-0.04	1.77	-0.17	-3.60	-0.05	0.59	-0.09	-2.95

Table S3. Partition of variation in RDA for different data sets. Given are the adjusted R squared for the testable fractions, the DF and F-, and p-values for all data, the 60 and 20 most abundant species, the indicator species (n = 53) and for the species richness and individual numbers of all plots. Negative r^2 values may arise due to the process of adjustment (Oksanen et al. 2008). All models, but the number of individuals, were significant at the $p < 0.01$ level. Fraction [a] = spatial variation independent of environment; fraction [b] = variation dependent on the interaction of space and environment; fraction [c] = environmental variation independent of space; fraction [d] = residuals; fraction [a+b+c] = total explained variation.

Data set	[a + b + c]	[a + b]	[b + c]	[a]	[b]	[c]	[d]	DF	F	p
All species	0.16242	0.02662	0.15538	0.00704	0.01958	0.1358	0.83758	18	1.8647	0.005 **
60 most abundant species	0.2002	0.03291	0.19328	0.00692	0.02599	0.16729	0.7998	18	2.1155	0.005 **
20 most abundant species	0.22538	0.03461	0.21175	0.01364	0.02098	0.19077	0.77462	18	2.3135	0.005 **
Indicator species	0.24385	0.04695	0.24075	0.00311	0.04384	0.1969	0.75615	18	2.3888	0.005 **
Species richness (all)	0.47753	0.03357	0.43169	0.04583	-0.01226	0.44396	0.52247	18	5.5319	0.005 **
Number of individuals (all)	0.09895	0.02976	0.10738	-0.00843	0.03819	0.06919	0.90105	18	1.4095	0.11