Supporting Information

Clough et al. 10.1073/pnas.1016799108

SI Methods

Study Area. All sites were situated around the Lore Lindu National Park, Central Sulawesi, Indonesia. The elevation in this region ranges from 400 m to about 1,000 m above sealevel. Natural vegetation is lowland to submontane rainforest, although only the latter remains (1). Being close to the equator, the seasonality is not very marked (1). Annual rainfall is around 2,500 mm (2), with a long-term pattern of 9 consecutive rainy months and 3 consecutive drier months (1). Cacao is the main cash crop grown in the study area. Hybridization between out-crossing genotypes is common in smallholder cacao, and cacao trees in this study, as well as most of the other trees in the area, originate from hybrids between Forastero and Trinitario types. The data were collected in Palolo and Kulawi valleys, respectively, on the northern and the western border of the Lore Lindu National Park.

Plot Selection. Biodiversity plots were selected to avoid strong correlations between the environmental variables shading, shade composition, and distance to forest, to allow their effects to be detected separately when considered jointly in statistical models. Random selection of plots was not used, as it would have caused multiple collinearity between explanatory variables. For example, natural forest shade trees tend to be more frequently found close to the forest edge. Instead, we selected 43 cacao plantations differing in shade intensity and shade tree diversity, as well as distance to forest and altitude from a subset of 80 plantations for which Y.C. recorded environmental data in 2006. Full-gradient range cover was achieved for all environmental variables in both valleys. In each plantation, we established a plot of 40 × 40 m, which was managed by specially hired and trained local assistants. Twenty-one plots were weeded manually every 3 mo, the other 22 plots every 6 mo. Half of the plots were fertilized with urea, the other half was not fertilized. To measure leaf-litter thickness, leaf litter was removed down to the soil from a 10-cm diameter patch at five points (center, and at the midpoints from the middle to each of the four corners) in each plot and the leaf litter measured with a ruler. The number of dead wood log piles were counted.

Plot Vegetation. Common planted shade tree species in the plots were the leguminous trees *Gliricidia sepium* (Jacq.) Walp. and *Erythrina subumbrans* Merr., candlenut *Aleurites moluccana* (L.) Willd., rambutan *Nephelium lappaceum* L., avocado *Persea americana* Mill., langsat (=longan) *Lansium domesticum* Correa, durian *Durio zibethinus* Merr. Species that were both planted and naturally occurring are sugar palm *Arenga pinnata* (Wurmb) Merr. and sago palm *Metroxylon sagu* Rottb. Common remaining forest tree species are *Ficus* sp., *Pterospermum celebicum* Miq., and *Bischofia javanica* Blume. Many other species occurred infrequently: 150 species of trees were recorded in total. In addition to the shade trees, other crops, such as coffee *Coffea* spp., or chili were grown within the cacao plots; however, in all our plots cacao was the most numerous crop tree.

Trees. Trees either standing in, or with the crown overlapping a 50 × 50-m plot containing the 40 × 40-m plot were identified to species level and measured for total height by R.P. Trees were considered forest trees when known to occur naturally in Sulawesi.

Herbaceous Plants. In October 2007, herb cover and species richness was recorded in two 5 × 5-m quadrats within each plot.

Endophytic Fungi. We collected 352 cacao leaf samples from 22 cacao plantations (*Theobroma cacao*) along the Kulawi Valley during the period from 10 to 15 March, 2007. From each tree we took four leaves from different branches in two different canopy layers. Considering the time foliar endophytes need to establish in a single young leaf, we harvested only matured leaves (3). After surface-sterilization (3% NaOH for 3 min, followed by 70% ethanol for 3 min, and rinsed in distilled water for further 3 min) leaf discs of 1 cm were plated on malt extract agar. Fungi growing out of leaf discs were isolated, purified, and grouped by morphospecies based on cultural characteristics (4, 5). Purified fungi were plated on thin layer malt extract agar with reduced nutrients and stored in a UV-chamber. After 1 mo, fungi were characterized by spores. Those fungi failing to sporulate were grouped by morphological characters of the mycelium.

Butterflies. Butterflies were caught using cylindrical live traps made out of a plastic plate and gauze, and baited with rotting banana. The four sampling phases ranged from March 4 to 19 and from March 24 until the April 8 in the valley of Kulawi, and from April 15 to 30 and from May 5 to 20 in the valley of Palolo. For each phase, during the first day, five traps were installed and baited in every one of a quarter of the valley's randomly selected cocoa plots, constituting a group of plots. On the second day, another quarter of the plots was prepared. We had four traps in the corners of the plot and one in the middle. For the remaining 14 d, butterflies were sampled and the baits refreshed by stirring the old mixture and adding a new one. We alternated each day between the groups and thus visited all of the plots every 2 d. Each plot was thus sampled seven times. Because of transport constraints, the same order of plot visitation was kept every day. In the following phase in the same valley, the other half of the valley's plots were assessed.

Ants. We used standardized plates (25-cm diameter), which were equipped with two baits of 2 cm^3 of tuna in oil and two sponges saturated with 70% sugar solution, to determine the abundance of ant morphospecies. One plate was placed in the main ramification of each experimental cacao tree (if the main ramification was too high, it was fixed at breast height on the stem with elastic rubber band; approximately 1 of 10 cases) and one was placed in a distance of 1.70 m on the ground, to the north of the stem. The plates were observed for 1 h. Every 15 min, the abundance of all ant species occurring at the baits feeding on fish or sugar was counted. Of every ant species appearing on the plate 5 to 10 specimens were caught with forceps and conserved in 70% ethanol for later identification. This survey was conducted twice in all 43 cacao plots at two daytimes (0900–1100 hours and 1100–1500 hours) with 2-wk intermission between the two samples. We identified the samples to morphospecies level.

Spiders. Within 22 cacao agroforestry systems in the Kulawi valley, ground-dwelling spiders were sampled using roofed pitfall traps (20-cm diameter) with four traps per plot each. We used a 1:1 mixture of ethylene glycol and water as preservation liquid. Spiders were sampled monthly from May 2007 to May 2008. All adult spiders were identified to species or morphospecies, but juveniles were excluded from the statistical analyses.

Birds. Each plot was visited once in 2007 (Kulawi: April–May, Palolo: August–October) and a second time in 2008 (Kulawi: April–May, Palolo: May–July). Plots were visited from 0530 to 1030 hours. No rain events ever occurred during that time of the day. Birds were recorded visually and acoustically, and by systematic tape recordings, by the same observer in all plots. Individuals flying above the canopy were excluded from the analysis. Further details can be found in ref. 6.

Rats. From August to December 2007, in each of two trapping sessions, locally made cage-traps (23 × 12 × 10 cm) baited with dried fish and slightly roasted coconut were set for three consecutive nights. During the first night traps were left open without trapping. The second trapping session was conducted 4 to 6 wk after the first. Twenty-five traps were placed in a 5 × 5-m grid system on the ground. Fifteen arboreal traps were placed on cocoa trees on horizontal branches and fixed with elastic bands. We baited traps in the afternoon and checked them the next morning. After data collection, we released the individuals near the point of capture. Further details can be found in ref. 7.

Amphibians and Reptiles. All plots were sampled six times between December 2007 and July 2008 (258 total sampling sessions). Sampling was conducted three times during day and night between 0600 and 1800 hours, and 1800 and 0600 hours, respectively. We used both diagonals of the plots as a single transect (113 m) with a width of 3 m on each side (i.e., 43.4% of the total plot area). Although transects were sampled in a time-constrained manner (25 min per plot), we thoroughly searched the leaf litter and turned logs, branch piles, and stones (8). We randomized sampling time of each plot and treatment category to avoid repeated sampling of the same plot at the same time. Animals found were photographed, measured, and toe-clipped, the latter to avoid pseudoreplication (8).

Cacao Management Survey. The cacao agroforestry survey was conducted in 2007 in five villages of the project region situated in the same valleys (Palolo, Kulawi) as the experimental plots. In each village, a sample of one cacao plot of each of 12 cocoa-producing households was selected, resulting in a total sample size of 60 cacao plots. The plots were not randomly selected but were systematically chosen to represent the entire intensification gradient of high to low canopy closure values. The percentage of cover by shade trees was monitored in 3 consecutive years from 2006 to 2008, with 8 to 16 densiometer measurements per plot using a hemispherical convex densitometer (Model-C, Robert E. Lemmon, Forest Densiometers, Bartlesville, OK 74006). Plot owner perceptions about soil fertility, patterns of input use, and adoption of agricultural innovations, as well as the impact of pests, diseases, dryness, and tree age on cocoa production were also recorded. All plot owners were contracted for a 1-y period to prepare weekly reports on yields and several yield determining factors from January to December 2007. Surveyed parameters included capital and labor used for plot management activities, for cacao pod and bean processing, (changes in) plot structure and intercropping, fertilizer input, pesticide input, fungicide and herbicide input, yield of fresh pods, and dry-bean marketing.

1. Whitten AJ, Mustafa M, Henderson GS (2002) *The Ecology of Sulawesi* (Gajah Mada University Press, Yogyakarta).
2. Gutzler C, Koehler S, Gerold G (2010) *Tropical Rainforests and Agroforests Under Global Change*, eds Tscharntke T, et al. (Springer, Heidelberg), pp 309–326.
3. Arnold AE, et al. (2003) Fungal endophytes limit pathogen damage in a tropical tree. *Proc Natl Acad Sci USA* 100:15649–15654.
4. Arnold AE, Maynard Z, Gilbert G, Coley PD, Kursar TA (2000) Are tropical fungal endophytes hyperdiverse? *Ecol Lett* 3:267–274.
5. Brown KB, Hyde KD, Guest DI (1998) Preliminary studies on endophytic communities of *Musa acuminata* species complex in Hong Kong and Australia. *Fungal Divers* 1:27–51.
6. Clough Y, Dwi Putra D, Pitopang R, Tscharntke T (2009) Local and landscape factors determine functional bird diversity in Indonesian cacao agroforestry. *Biol Conserv* 142: 1032–1041.
7. Weist M, Tscharntke T, Sinaga MH, Maryanto I, Clough Y (2010) Effect of distance to forest and habitat characteristics on endemic versus introduced rat species in agro-forest landscapes of Central Sulawesi, Indonesia. *Mamm Biol* 75:567–571.
8. Wanger TC, et al. (2009) Conservation value of cacao agroforestry for amphibians and reptiles in Southeast Asia: Combining correlative models with follow-up field experiments. *J Appl Ecol* 46:823–832.

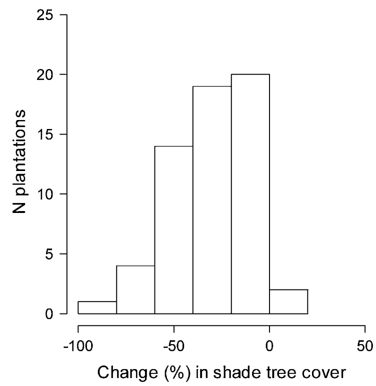

Fig. S1. Shade tree removal is ongoing. Change in percentage cover by shade trees in 60 cacao plantations of Central Sulawesi, Indonesia between 2007 and 2008 (survey component).

Table S1. Effects of dry yield, forest distance, region, and the interaction between dry yield and forest distance on species richness of different species groups in cacao agroforests of Central Sulawesi

Group	Parameter	Estimate	SE	*t*	*P*	Lower CI	Upper CI
Trees	(Intercept)	14.58	5.97			2.88	26.28
	log(yield)	−1.05	1.12	−0.94	0.35	−3.25	1.15
	sqrt(forest distance)	−0.14	0.06	−2.42	0.02	−0.26	−0.027
Herbs	(Intercept)	30.08	7.24			15.90	44.27
	log(yield)	−2.91	1.28	−2.27	0.03	−5.42	−0.40
Endophytic fungi	(Intercept)	59.71	29.81			1.28	118.14
	log(yield)	−2.50	5.09	−0.49	0.63	−12.47	7.47
Butterflies	(Intercept)	4.64	4.23			−3.66	12.93
	log(yield)	0.87	0.72	1.20	0.24	−0.54	2.27
	Region	−1.29	0.87	−1.48	0.15	−3.00	0.42
Ants	(Intercept)	4.47	11.31			−17.71	26.64
	log(yield)	3.02	2.05	1.48	0.15	−0.99	7.03
	sqrt(forest distance)	−0.26	0.10	−2.59	0.01	−0.45	−0.063
	Region	4.97	2.24	2.22	0.03	0.58	9.35
Spiders	(Intercept)	8.77	8.57			−8.02	25.56
	log(yield)	1.09	1.46	0.75	0.46	−1.77	3.96
Birds	(Intercept)	10.54	6.28			−1.77	22.85
	log(yield)	0.16	1.11	0.14	0.89	−2.02	2.34
Rats	(Intercept)	−2.44	2.55			−7.44	2.55
	log(yield)	0.83	0.49	1.71	0.10	−0.12	1.78
	sqrt(forest distance)	−0.05	0.02	−2.23	0.04	−0.09	−0.01
Amphibians and reptiles	(Intercept)	5.47	2.32			0.93	10.02
	log(yield)	−0.41	0.41	−0.99	0.33	−1.21	0.40
Endemic birds	(Intercept)	6.55	2.45			1.75	11.36
	log(yield)	−0.35	0.46	−0.76	0.45	−1.25	0.55
	sqrt(forest distance)	−0.055	0.024	−2.29	0.03	−0.10	−0.008
Endemic butterflies	(Intercept)	−1.04	2.26			−5.48	3.39
	log(yield)	0.70	0.43	1.64	0.11	−0.13	1.53
	sqrt(forest distance)	−0.017	0.022	−0.79	0.44	−0.061	0.026

Sqrt, Square-root. Maximum-likelihood estimates of intercepts and parameter coefficients from the best generalized linear model; uncertainty estimates are 95% confidence intervals.

Table S2. Effects of dry yield, forest distance, and region on rarefied species richness of different species groups in cacao agroforests of Central Sulawesi

Group	Parameter	Estimate	SE	*t*	*P*	Lower CI	Upper CI
Trees	(Intercept)	0.88	1.09			−1.26	3.03
	log(yield)	0.23	0.21	1.12	0.35	−1.77	0.63
	sqrt(forest distance)	−0.01	0.01	−0.963	0.02	−0.03	0.01
Butterflies	(Intercept)	4.59	1.69			1.27	7.91
	log(yield)	0.13	0.32	0.41	0.68	−0.49	0.76
	Region	−0.02	0.02	−1.42	0.16	−0.06	0.01
Spiders	(Intercept)	8.84	4.46			0.09	17.58
	log(yield)	0.54	0.80	0.67	0.51	−1.03	2.12
	sqrt(forest distance)	−0.05	0.04	−1.12	0.28	−0.13	0.04
Amphibians and reptiles	(Intercept)	2.71	1.26			0.24	5.18
	log(yield)	0.13	0.24	0.56	0.58	−0.34	0.60
	sqrt(forest distance)	−0.02	0.01	−1.62	0.12	−0.05	0.00
Endemic butterflies	(Intercept)	2.32	1.89			−1.38	6.03
	log(yield)	0.13	0.35	0.37	0.71	−0.56	0.82
	sqrt(forest distance)	0.00	0.02	0.25	0.81	−0.03	0.04

Maximum-likelihood estimates of intercepts and parameter coefficients from the best generalized linear model; uncertainty estimates are 95% confidence intervals. The interaction between yield and forest distance was included in the full models. Samples were rarefied to 5 individuals for trees, 13 individuals for all butterflies, 29 for spiders, 5 for trees, 5 for amphibians and reptiles (six low-abundance sites excluded), and 5 for endemic butterflies (21 low-abundance sites excluded).

Table S3. Effects of region, altitude, distance to forest, shade tree cover, number of forest tree species, number of trees higher than 10 m, forest distance, weeding frequency, density of dead wood, and leaf-litter thickness on species richness of different species groups in cacao agroforests of Central Sulawesi

Group	Parameter	Estimate	SE	*t*	*P*	Lower CI	Upper CI
Trees	(Intercept)	7.19	1.53			4.20	10.18
	n Tall trees	0.13	0.079	1.64	0.11	−0.026	0.28
	sqrt(forest distance)	−0.17	0.052	−3.31	0.002	−0.27	−0.070
Herbs	(Intercept)	19.63	4.15			11.50	27.76
	Low weeding	4.61	1.43	3.22	0.003	1.80	7.42
	Dead wood	0.59	0.23	2.60	0.01	0.14	1.03
	Shade tree cover	−0.14	0.05	−2.62	0.01	−0.24	−0.03
	sqrt(forest distance)	−0.29	0.11	−2.54	0.02	−0.51	−0.07
	n Forest tree species	−0.072	0.27	−0.27	0.79	−0.60	0.46
	sqrt(fd) × *n* forest tree species	0.033	0.014	2.33	0.03	0.0052	0.060
Endophytic fungi	(Intercept)	45.13	2.82			39.60	50.67
Butterflies	(Intercept)	11.63	1.10			9.46	13.79
	n Tall trees	−0.11	0.05	−2.03	0.05	−0.22	−0.0039
	Region	−2.25	0.83	−2.70	0.01	−3.88	−0.61
Ants	(Intercept)	7.23	5.17			−2.90	17.37
	Altitude	0.019	0.0077	2.46	0.02	0.0038	0.034
Spiders	(Intercept)	21.50	3.71			14.24	28.77
	Altitude	−0.010	0.0057	−1.76	0.09	−0.021	0.0012
Birds	(Intercept)	9.17	1.85			5.53	12.80
	n Tall trees	0.31	0.07	4.38	0.00	0.17	0.46
	Low weeding	−4.10	1.15	−3.57	0.00	−6.35	−1.85
	Dead wood	0.27	0.28	0.98	0.34	−0.28	0.82
	sqrt(forest distance)	0.052	0.064	0.80	0.43	−0.075	0.18
	Dead wood × sqrt(fd)	−0.040	0.015	−2.72	0.01	−0.070	−0.011
Rats	(Intercept)	−0.36	1.29			−2.89	2.16
	Altitude	0.0033	0.0019	1.78	0.09	−0.00035	0.0070
	Region	−0.88	0.50	−1.75	0.09	−1.86	0.10
Amphibians and reptiles	(Intercept)	6.81	1.30			4.25	9.36
	Altitude	−0.01	0.0016	−4.00	0.00030	−0.01	0.00
	Low weeding	0.32	0.67	0.47	0.64	−1.00	1.64
	Leaf litter thickness	0.46	0.20	2.32	0.03	0.07	0.85
	Dead wood	0.12	0.06	1.93	0.06	0.00	0.25
	sqrt(forest distance)	0.00	0.03	0.02	0.99	−0.05	0.05
	Low weeding × sqrt(forest distance)	−0.08	0.03	−2.35	0.02	−0.14	−0.01
Endemic birds	(Intercept)	0.81	1.25			−1.64	3.25
	Leaf litter thickness	0.42	0.23	1.87	0.07	−0.02	0.86
	Shade tree cover	0.048	0.017	2.85	0.01	0.015	0.080
	sqrt(forest distance)	−0.039	0.021	−1.83	0.07	−0.081	0.0027
Endemic butterflies	(Intercept)	3.91	0.63			2.67	5.15
	n Tall trees	−0.06	0.032	−1.91	0.06	−0.12	0.0016
	Region	−0.84	0.48	−1.75	0.09	−1.77	0.10

Maximum-likelihood estimates of intercepts and parameter coefficients for the best generalized linear model; uncertainty estimates are 95% confidence intervals.

Table S4. Determinants of dry yield (log-transformed) in the 43 cacao plots of the agroecology experiment

Parameter	Estimate	SE	*t*	*P*	Lower CI	Upper CI
(Intercept)	7.09	0.54	13.21	0.00	6.04	8.15
Altitude	−0.0010	0.0006	−1.64	0.11	−0.0023	0.0002
Dead wood	−0.039	0.023	−1.71	0.10	−0.083	0.006
Shade tree cover	−0.012	0.0050	−2.45	0.02	−0.022	−0.0025
sqrt(forest distance)	0.012	0.0070	1.75	0.09	−0.0015	0.026
Region	−0.34	0.15	−2.29	0.03	−0.63	−0.048

Maximum-likelihood estimates of intercepts and parameter coefficients from the best generalized linear model; uncertainty estimates are 95% confidence intervals.

Table S5. Determinants of yield in 60 cacao plantations monitored for socioeconomic data

Parameter	Estimate	SE	*t*	*P*	Lower CI	Upper CI
(Intercept)	5.34	0.56	9.50	0.0000	4.24	6.44
Shade tree cover (%)	−0.0089	0.0039	−2.26	0.028	−0.017	−0.0012
log(inputs)	0.10	0.030	3.34	0.0016	0.04	0.16
log(labor)	0.24	0.071	3.37	0.0014	0.10	0.38
n Forest trees	−0.12	0.037	−3.19	0.0024	−0.19	−0.045
Disease incidence	−0.018	0.0079	−2.23	0.030	−0.033	−0.0021
sqrt(forest distance)	0.014	0.0094	1.51	0.138	−0.0043	0.033
log(inputs) × sqrt(fd)	−0.0023	0.0010	−2.29	0.0261	−0.0042	−0.00033

Maximum-likelihood estimates of intercepts and parameter coefficients for the best generalized linear model; uncertainty estimates are 95% confidence intervals. Inputs include expenses for herbicides, insecticides and fungicides, but not fertilizer.

Table S6. Range in drivers of yield in 60 plantations monitored for shade, management and yield data

Variable	Maximum	Minimum
Shade tree cover (%)	88.8%	1.6%
Inputs	1 Mio Rp·y^{-1}	0 Rp·y^{-1}
Labor	3200 h·ha^{-1}·y^{-1}	20 h·ha^{-1}·y^{-1}
n Forest trees·ha^{-1}	9 Trees	0 Trees
Losses to disease (%)	50%	3%
Distance to forest (m)	4,000	0

Rp stands for Indonesian Rupiah (9,200 Rp to 1 US dollar in the period between June 2006 and June 2008).