



UNIVERSITY OF LEEDS

This is a repository copy of *Persistent effects of pre-Columbian plant domestication on Amazonian forest composition*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/114745/>

Version: Supplemental Material

---

**Article:**

Levis, C, Costa, FRC, Bongers, F et al. (150 more authors) (2017) Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science*, 355 (6328). pp. 925-931. ISSN 0036-8075

<https://doi.org/10.1126/science.aal0157>

---

© 2017, American Association for the Advancement of Science. This is an author produced version of a paper published in *Science*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



## Supplementary Materials for

### Persistent effects of pre-Columbian plant domestication on Amazonian forest composition

Carolina Levis\*,1,2 Flávia R.C. Costa,3 Frans Bongers,2 Marielos Peña-Claros,2 Charles R. Clement,4 André B. Junqueira,5 Eduardo G. Neves,6 Eduardo K. Tamanaha,7 Fernando O.G. Figueiredo,1 Rafael P. Salomão,8 Carolina V. Castilho,9 William E. Magnusson,3 Oliver L. Phillips,10 Juan Ernesto Guevara,11,12 Daniel Sabatier,13 Jean-François Molino,13 Dairon Cárdenas López,14 Abel Monteagudo Mendoza,15 Nigel C.A. Pitman,16,17 Alvaro Duque,18 Percy Núñez Vargas,19 Charles Eugene Zartman,3 Rodolfo Vasquez,15 Ana Andrade,20 José Luís Camargo,20 Ted R. Feldpausch,21,10 Susan G.W. Laurance,22 William F. Laurance,22 Timothy J. Killeen,23 Henrique Eduardo Mendonça Nascimento,3 Juan Carlos Montero,24,25 Bonifacio Mostacedo,26 Iêda Leão Amaral,3 Ima Célia Guimarães Vieira,8 Roel Brienen,10 Hernán Castellanos,27 John Terborgh,17 Marcelo de Jesus Veiga Carim,28 José Renan da Silva Guimarães,28 Luiz de Souza Coelho,3 Francisca Dionízia de Almeida Matos,3 Florian Wittmann,29 Hugo F. Mogollón,30 Gabriel Damasco,11,9 Nállarett Dávila,31 Roosevelt García-Villacorta,32,33 Euridice N. Honorio Coronado,34,10 Thaise Emilio,35,36 Diogenes de Andrade Lima Filho,3 Juliana Schietti,37 Priscila Souza,3 Natalia Targhetta,38 James A. Comiskey,39 Beatriz S. Marimon,40 Ben-Hur Marimon Jr.,40 David Neill,41 Alfonso Alonso,42 Luzmila Arroyo,43 Fernanda Antunes Carvalho,3 Fernanda Coelho Souza,3 Francisco Dallmeier,42 Marcelo Petrati Pansonato,36 Joost F. Duivenvoorden,44 Paul V.A. Fine,11 Pablo R. Stevenson,45 Alejandro Araujo-Murakami,43 Gerardo A. Aymard C.,46 Chris Baraloto,47,48 Dário Dantas do Amaral,8 Julien Engel,48 Terry W. Henkel,49 Paul Maas,50 Pascal Petronelli,47 Juan David Cardenas Revilla,3 Juliana Stropp,51,52 Doug Daly,53 Rogerio Gribel,54 Marcos Ríos Paredes,55 Marcos Silveira,56 Raquel Thomas-Caesar,57 Tim R. Baker,10 Naara Ferreira da Silva,58 Leandro Valle Ferreira,8 Carlos A. Peres,59 Miles R. Silman,60 Carlos Cerón,61 Fernando Cornejo Valverde,62 Anthony Di Fiore,63 Eliana M. Jimenez,64 Maria Cristina Peñuela Mora,65 Marisol Toledo,24 Edelcilio Marques Barbosa,3 Luiz Carlos de Matos Bonates,3 Nicolás Castaño Arboleda,14 Emanuelle de Sousa Farias,66 Alfredo Fuentes,67,68 Jean-Louis Guillaumet,69 Peter Møller Jørgensen,68 Yadvinder Malhi,70 Ires Paula de Andrade Miranda,3 Juan Fernando Phillips,71 Adriana Prieto,72 Agustín Rudas,72 Ademir R. Ruschel,73 Natalino Silva,74 Patricio von Hildebrand,75 Vincent A. Vos,76,77 Egleé L. Zent,78 Stanford Zent,78 Bruno Barçante Ladvoat Cintra,58 Marcelo Trindade Nascimento,79 Alexandre A. Oliveira,80 Hirma Ramirez-Angulo,81 José Ferreira Ramos,3 Jochen Schöngart,29 Rodrigo Sierra,82 Milton Tirado,82 Geertje van der Heijden,83,84 Emilio Vilanova Torre,81,85 Ophelia Wang,86 Kenneth R. Young,87 Cláudia Baider,88,80 Angela Cano,45 William Farfan-Rios,60 Cid Ferreira,3 Bruce Hoffman,89 Casimiro Mendoza,90,91 Italo Mesones,11 Armando Torres-Lezama,81 Maria Natalia Umaña Medina,92 Tinde R. van Andel,93 Daniel Villarroel,43 Roderick Zagt,94 Miguel N. Alexiades,95 Henrik Balslev,96 Karina Garcia-Cabrera,60 Therany Gonzales,97 Lionel

Hernandez,98 Isau Huamantupa- Chuquimaco,19 Angelo Gilberto Manzatto,99 William Milliken,100 Walter Palacios Cuenca,101 Susamar Pansini,102 Daniela Pauletto,103 Freddy Ramirez Arevalo,104 Neidiane Farias Costa Reis,105 Adeilza Felipe Sampaio,102 Ligia Estela Urrego Giraldo,18 Elvis H. Valderrama Sandoval,106,104 Luis Valenzuela Gamarra,15 César I.A. Vela,107 Hans ter Steege\*,93,108

\* Correspondence to: Carolina Levis (carollevis@gmail.com and carolina.levis@wur.nl) and Hans ter Steege (hans.tersteege@naturalis.nl)

**This PDF file includes:**

Materials and Methods  
Supplementary Text  
Figs. S1 to S13  
Tables S1 to S3  
Captions for Database S1 and S2  
References (42-79) [Note: The numbers refer to any additional references cited only within the Supplementary Materials]  
Custom R scripts

**Other Supplementary Material for this manuscript includes the following:**

Database S1 (separate file): List of species with populations that were likely domesticated, semi-domesticated or incipiently domesticated by pre-Columbian peoples in Amazonia and elsewhere in the Americas.

Database S2 (separate file): Plot meta data

## Materials and Methods

### Data collection: floristic data

We used 1091 forest inventory plots of the Amazon Tree Diversity Network (ATDN) database distributed across Amazonia (Fig. S2A). Only lowland (terra firme and white-sand podzol) plots were included in the analysis; wetland plots, as defined by ter Steege et al. (17), were excluded. The exclusion of wetlands is justified for four reasons: 1) floodplain trees have distinct and strong ecological requirements related to flood duration; 2) population turnover is probably faster in white-water floodplains and slower in blackwater floodplains than in non-flooded forests; 3) the effect of pre-Columbian domestication may be blurred as many floodplain archaeological sites have been destroyed by the annual floods; and 4) plots on floodplains are at zero distance from rivers, making it impossible to evaluate the effect of this variable.

Plots cover a wide range of soils and topographies (Table 1). Most plots (N = 819) measure 1 ha; others vary from 0.1 to 9 ha (Database S2). In each plot, ATDN scientists inventoried all woody species with  $\geq 10$  cm diameter at breast height; almost all individuals were identified to the species level (95 % of plots have less than 5 % of individuals without botanical identification). Plots with more than 25 % of trees unidentified to species were excluded from the analyses of the absolute and relative richness of domesticated species. Although identification problems exist in Amazonian tree inventories, domesticated species are widely used and cultivated, and are therefore better-known to botanists and local parataxonomists.

### Data collection: historical human factors

The distribution of archaeological sites was obtained from a database of 3795 archaeological sites and eco-archaeological regions in lowland South America that includes pre-Columbian habitation sites (with and without anthropogenic soils), earthworks (mounds, causeways, raised fields, terraces) and rock art (paintings and petroglyphs) updated from 3318 sites presented by a previous study (1). Only archaeological sites that fall within the Amazon River basin were used in this study (3348 sites shown in Fig. S2B). Archaeological sites are places where material remains of pre-Columbian human activities are still visible and eco-archaeological regions are environmental settings with large and abundant pre-Columbian earthworks (25). To illustrate the biased sampling of archaeological sites across Amazonia, we created a map with a background color showing the density of archaeological sites in 1°-grid cell and forest plots in open circles indicating values of the relative abundance of domesticated species (Fig. S6). To quantify the gradient of historical human influence, we measured the distance from each forest plot to the nearest archaeological site and to the nearest navigable river margin (Fig. S2B), using near distance tool of ArcMap version 9.3. For plots located in 1°-grid cells with zero density of archaeological sites we measured distances from plots to the nearest eco-archaeological regions (if they exist) using near distance tool of ArcMap version 9.3. Plots located at zero distance from archaeological sites occur in south-western and eastern Amazonia (Table 1) and those located within eco-archaeological regions occur in south-western Amazonia and the Guiana Shield (Fig. S6). Equal weighting was given to all the different types of archaeological sites or eco-archaeological regions because we did not have the description of all sites in the database we used for this study.

Distance from rivers was also chosen as a proxy of pre-Columbian settlements, because this is a good predictor of the probability of finding sedentary pre-Columbian occupation sites in Amazonia (26), which reflects peoples' preferences for living along rivers. Distance from the river is not strongly correlated with the distance to archaeological sites or any of the environmental variables we tested (Fig. S13), allowing the use of both variables in the analyses. Archaeological sites along tributaries in interfluvial forests are under-sampled compared to sites in more accessible areas (Fig. S2), which also justifies using both distance measures (i.e., distance from archaeological sites and from rivers) in the analyses. The river network was obtained from the HydroSHEDS dataset (available at <http://hydrosheds.cr.usgs.gov>) (43). 'Upcell' values are features of the HydroSHEDS dataset that represent the maximum flow accumulation at any location in the river network. We used HydroSHEDS data to define perennial and navigable rivers by selecting cells with upcell values greater than 15,000, following the study of McMichael et al. (26). For larger rivers (more than 1 km wide) we used river polygons obtained from ANA/BRASIL (44).

#### Data collection: regional and local environmental data

To account for the effect of regional environmental conditions across Amazonia we used the geological regions delimited by Fittkau (45) and analyzed by ter Steege et al. (17), who showed that six different geological regions are dominated by different suites of tree species. To account for the effect of local environmental conditions, soil fertility (Cation Exchange Capacity), soil pH, rainfall seasonality and the Height Above the Nearest Drainage (HAND) were included in the analyses. Cation Exchange Capacity (CEC) and soil pH were obtained from SoilGrids 250 m for all plots using the mean values of 5 cm of soil depth (46, <http://www.isric.org/content/soilgrids>); rainfall seasonality was calculated as the maximum cumulative number of months with < 100 mm of rainfall using the monthly data from 1998 to 2004 of the Tropical Rainfall Measuring Mission (TRMM) satellite product 3B43 V6 at a 0.25° resolution (about 28 km at the equator) (47); HAND was obtained from the Ambdata database (48). Mean, median, maximum and minimum values of all variables are presented in Table 1.

#### Data collection: control groups for testes of how environment and human factors affect distribution of hyperdominant non-domesticated species

To understand whether non-human primates may promote dominance of certain tree species in forests closer to archaeological sites and rivers, we analyzed the effect of distance from archaeological sites and rivers on the abundance of non-domesticated species that are dispersed by primates and probably by other vertebrates. Although all vertebrates disperse forest seeds, large non-human primates have similar fruit preferences to humans and their actions in the forest can be compared with human behavior (2). We identified 20 non-domesticated species that had estimated population sizes comparable to those of the 20 hyperdominant domesticated species, and that are primarily dispersed by non-human primates. We also selected two control groups of non-domesticated species: the first group consists of 20 hyperdominant species based on specific criteria and the second of 20 hyperdominant species selected at random. The criteria used to select the first control group were: (1) species with estimated population sizes comparable to those of the 20 hyperdominant domesticated species; (2) species that belong to the same botanical families as the 20 hyperdominant domesticated species (if possible); (3) species that are not mainly dispersed by primates. In the second control group of 20 hyperdominant species selected at

random, three are dispersed mainly by primates and four were selected in the first control group. The lists of the control groups of hyperdominant species are presented in the Table S1.

#### List of domesticated species in ATDN plots

We created a list of woody species with some evidence of selection and propagation by humans in Amazonia and elsewhere in the Americas to which we refer hereafter as “domesticated species”. First, we considered the list of domesticated species at European contact compiled by Clement (6). We excluded herbs and species identified to the genus level (e.g., *Hevea* spp.), because herbs were not sampled in the ATDN plots and only a few species of species-rich genera were shown to be domesticated. To this first list, we added two species with evidence of past human selection that have been studied by Clement’s group (*Euterpe precatoria*) (49) and *Caryocar brasiliense* (50), and one species considered to be the wild progenitor of a species with domesticated populations (*Bixa urucurana*) (51).

We then conducted a bibliographical search for recent articles on domesticated plants using “domestication in Amazon\*” and “domestication in Brazil\*” (and all other Amazonian countries) as keywords in Web of Science and “domestication in Amazon\*” in Google Scholar. Three palm species (*Attalea phalerata*, *Phytelephas macrocarpa* and *Astrocaryum chambira*) were incorporated in the list based on two recent papers (52, 53).

We also used Mansfeld’s World Database of Agricultural and Horticultural Crops (54, <http://mansfeld.ipk-gatersleben.de/>) to add new species to the list. First, from this database we obtained a list of cultivated species in all Amazonian countries. Then, information about the natural distribution, cultivation, uses and domestication was obtained for all cultivated species that occur in the forest plots to classify the degree of domestication (DD) of each species. Cultivation is defined here as the process of growing plants, while plant domestication involves cultivation and also selection and propagation of specific populations by humans. The degree of domestication was based on the following indicators adapted from Clement (6), Dempewolf et al. (55) and Hammer & Khoshbakht (56):

(A) Any degree of phenotypic differentiation between the domesticated taxon and its wild progenitor (including evidence of a smaller variance of traits subjected to selection than that of the original wild populations, as this may represent a founder effect; DD = 2).

(B) The extent of cultivation in terms of geographical area (if the geographical area of cultivation is outside its natural range of distribution within the Americas - North, Central and South America; DD = 1). The natural range of distribution of each species was obtained from Mansfeld’s World Database (54).

(C) Evidence of cultivation since AD 1492 (DD = 1) and before AD 1492 (DD = 2), both suggesting a long history of selection.

All species from the Mansfeld’s World Database (10 species) with a summed degree of domestication  $\geq 2$  were included in the new list, resulting in a list with 85 domesticated species. Species with evidence of extensive and long-term cultivation (i.e., indicator B: the extent of cultivation and C: evidence of cultivation before AD1492) are likely to have been subjected to a long history of selection and propagation even if no studies were done to investigate the degree of phenotypic variation in cultivated populations. If we look for evidence of phenotypic differentiation between cultivated and wild populations of these species, there is a very high chance of finding a signal of a reduction in phenotypic variability (for incipiently domesticated populations) or an increase in phenotypic variability (for semi-domesticated and domesticated populations) (6). For more information, we

presented references for evidence of domestication and cultivation of each species listed in the Database S1. Finally, we validated all names with the Taxonomic Name Resolution Service using the Tropicos® database (57, accessed in May, 2015). We cross-checked all scientific names of domesticated species with the list of all species present in ATDN plots.

### Data analyses

All analyses were conducted in the R environment (58). We used a spatial loess model to produce distribution maps for 11 domesticated species for which there is reasonably good information about their origins of domestication (Fig. 1 and Fig. S1). Using the same approach as ter Steege et al. (17), Amazonia was divided into 1° latitude and longitude grid cells, and the interactions between latitude and longitude were used as variables in the loess regression models to estimate average density of individuals in each grid cell. These densities were mapped and compared to the origin of domestication predicted by previous studies (1, 13) that analyzed the geographic distribution of genetic and morphological diversity found in cultivated and wild populations of domesticated species (see Supplementary Text for information on the origin of domestication of each species). Higher genetic diversity often indicates the location of the origin of domestication, and the genetic variability found in the cultivated populations is often a subset of the genetic variability found in the wild population (6).

We used the list of domesticated species to quantify five domestication measures in each plot (59): (1) the abundance of domesticated species (the number of individuals of domesticated species per hectare); (2) the relative abundance of domesticated species (the number of individuals of domesticated species divided by the total number of individuals found in the plot); (3) the richness of domesticated species (the number of domesticated species per plot); (4) the relative richness of domesticated species (the number of domesticated species divided by the total number of species found in the plot); and (5) the relative abundance of hyperdominant domesticated species (the number of individuals of domesticated species that are hyperdominants divided by the total number of individuals found in the plot). The lists of hyperdominant non-domesticated species (control groups) were used to quantify their relative abundance in the plot as described above in (5).

We calculated the spatial variation of the absolute and the relative abundance and richness of domesticated species to understand how the proportion of domesticated species varies across Amazonia. We also used a loess regression model to interpolate the measures of domestication for the entire Amazon. The model was used to estimate the absolute and the relative abundance and richness of domesticated species for each grid cell (Fig. 2). We used an exponential model to fit the relationship between the absolute and relative abundance of 85 domesticated species and the absolute and relative richness of 85 domesticated species in forest plots presented in the Fig. S4.

To evaluate the relationship between the measures of domestication and the measures of abundance for control groups (response variables) and historical human and environmental conditions (explanatory variables) we used mixed-effects models and multiple linear regressions. Distance to archaeological sites and eco-archaeological regions, distance to rivers and HAND were log transformed ( $\log_{10} + 1$ ) before the analysis to normalize these three variables. In the Amazonia-wide mixed-effects model, geological regions were incorporated as random factors and explanatory variables as fixed factors. Within each geological region, we analyzed the effects of environmental and human factors on the response variables using separate multiple linear regression models. Mixed-effects and

multiple regression models were implemented with the 'lmer' and 'lm' functions of the R 'lme4' (60) and R 'sjstats' packages (61), respectively. We included all explanatory variables in the models, because only one strong correlation (higher than 0.8) was found in south-western, but was not present in other regions nor over all Amazonia (Fig. S13). We also used the variation inflation factor (VIF) to identify multicollinearity among explanatory variables used in the multiple regression models and we did not detect signs of strong multicollinearity in any of the models. The results of the mixed-effects and multiple regression models corrected for multiple testing (Bonferroni correction) are presented in Table S2. Although Bonferroni correction leads to conservative test results, the main results presented in Figure 3 remain significant after this correction.

After removing plots located at zero distance to archaeological sites or eco-archaeological regions from the multiple regression models presented in Fig. 3 the relations with archaeological sites are still visible and significant within south-western and eastern regions, but are not visible for the Amazonia-wide models (Table S3). Simple plots of the response variables against distance from archaeological sites (and eco-archaeological regions) and rivers are presented in Figs. S7-S11 and show actual units of distance. The visualization of the regression models was performed using the visreg package (62). We used a log transformation ( $\log_{10} + 1$ ) of the explanatory variables to fit a nonlinear relationship between the relative abundance and richness of domesticated species and the distance to archaeological sites and rivers. Conditional plots were used for simple models within geological regions and contrast plots were used for mixed-effect models in the Amazonia-wide analyses (geological regions incorporated as random factors), because it is only possible to obtain confidence intervals for contrast plots in mixed-effect models.

We used variation partitioning (63) to determine how much of the variation in the response variables can be explained by historical human factors, by environmental conditions, and by human and environmental factors together. The fractions of variation were based on the results of three multiple regression models (adjusted  $r^2$ ): a model with only human factors included as predictors; a model with only local environmental predictors; and a human + environment model, including both sets of predictors. Variation partitioning was implemented using the varpart function of the R vegan package (64).

## **Supplementary Text**

### Description of the 11 domesticated species, their distribution across Amazonia and their likely origins of domestication

We analyzed 11 domesticated species for which information about their origins of domestication had been described by previous studies (13), and the molecular and morphological studies used to identify their origins are described below. We divided the species in two groups: five domesticated species that are hyperdominants of the Amazonian Flora (Fig. 1) and six that are not hyperdominants (Fig. S1).

#### Five domesticated species that are hyperdominants:

1. *Bertholletia excelsa* Bonpl., Lecythidaceae

The Brazil nut or Amazon nut tree (*Bertholletia excelsa*) produces oil-rich seeds that are the most important non-timber forest product extracted from Amazonian forests. Populations with large seeds probably resulted from human selection and stands with high abundances of Brazil nut trees have been associated with past human dispersal and cultivation (6, 65). The dispersal-mediated effect of large caviomorph rodents is considered another factor

determining the existence of dense stands of Brazil nut in Amazonia (66). Brazil nut trees are commonly planted and/or favored in swiddens, but the species survives into mature forests when abandoned (67). There are three hypotheses about Brazil nut's origin of domestication: it may have originated in south-eastern Amazonia, according to Scott Mori (13); a northern/central Amazonian origin was suggested based on an historical linguistic analysis (65); a south-western origin is a new hypothesis (42), supported by population genetics (68). Brazil nut trees occur in all Amazonian regions except north-western Amazonia, and it is particularly abundant in forests of the eastern and western parts of southern Amazonia.

2. *Inga ynga* (Vell.) J.W. Moore, synonym of *Inga edulis* Mart., Fabaceae

The icecream bean tree or inga (*Inga ynga*) is widely cultivated in homegardens and swiddens for its edible fruits and is the most important useful species of the genus. Inga trees were selected by Amerindians, resulting in semi-domesticated populations with long and thick pods that die out in mature forests after abandonment (13, 69). Clement et al. (13) hypothesized that the species originated in western Amazonia, where it shows the most dramatic phenotypic variability. However, genetic evidence shows that cultivated inga in Peru is not associated with wild populations in the same area (70). Inga trees occur in all Amazonian regions. New studies are needed to confirm the origin of domestication of *Inga ynga*.

3. *Pourouma cecropiifolia* Mart., Urticaceae

Amazon tree grape (*Pourouma cecropiifolia*) is a pioneer tree cultivated inside and outside its natural distribution range (54). The presence of tree grape in forests often indicates recent human disturbance, because semi-domesticated populations cannot survive long after human abandonment (6). Clement (69) suggested an origin in western Amazonia because of the phenotypic contrast between populations from western Amazonia and wild populations. Fruits with more pulp and a smooth rind have also been observed along the Vaupés River in Colombia. Tree grape occurs in all Amazonian regions, but we found the highest abundance of this species in south-western forests, where no cultivated populations are known.

4. *Pouteria caimito* (Ruiz & Pav.) Radlk., Sapotaceae

Known as abiu or caimito (*Pouteria caimito*), it is cultivated as a fruit tree in tropical America (54). It is commonly found in indigenous agroforestry systems in the Peruvian Amazon with various fruit shapes. Its fruits can vary from 50 to 1000 g (69). Fully domesticated populations of this species cannot survive in mature forests (6). Individuals with large and spherical fruits have a restricted distribution in eastern Peru, whereas small and ovoid fruits also occur in the same location, suggesting a western origin of the domesticated populations of the species (69). Although *P. caimito* occurs in all Amazonian regions, we frequently found it in plots in eastern Amazonia, and its highest abundance was found in forests of the northern part of the Guiana Shield. None of these regions are hypothetical origins of domestication.

5. *Theobroma cacao* L., Malvaceae

The cocoa tree (*Theobroma cacao*) has a long history of cultivation, particularly in Mesoamerica, where seeds were used to produce a stimulant beverage called chocolātl (54). Genetic evidence indicates that cocoa is native to western Amazonia and it was taken to Mesoamerica before European conquest. In Amazonia, the cocoa trees may have been selected for its edible pulp and it was first domesticated in the north-western region (13, 39). Although cocoa trees occur in almost all Amazonian regions, its highest frequency and abundance was found in forests of the southern part of western Amazonia. Semi-

domesticated populations of cocoa are common in swiddens and the species survives in tropical evergreen forests after abandonment (69), which could explain their high abundance in the southern part of the Amazon basin.

Six domesticated species that are not hyperdominants:

6. *Anacardium occidentale* L., Anacardiaceae

The cashew tree (*Anacardium occidentale*) is a medium to large tree and was found in low abundances in a few plots of central and eastern Amazonia. The probable origin of domestication of cashew is in open forest ecosystems of the Northeast of Brazil, where the greatest diversity of cultivated varieties has been found (71). The cashew tree was domesticated for its edible (pseudo)fruits for direct consumption and for making beverages (60). The species is now cultivated across the tropics (54).

7. *Bixa orellana* L. var. *urucurana* (Willd.) Kuntze ex Pilg., synonym of *Bixa urucurana* Willd., Bixaceae

The annatto tree (*Bixa orellana*) produces a red colorant used since pre-Columbian times. It is cultivated throughout tropical America, and was introduced early into nearly all tropical regions of the Old World during the first century after European contact (54). The origin of domestication is still not clear, although *Bixa urucurana* was recently identified as the wild progenitor of *B. orellana* (51). *B. urucurana* occurs mainly along rivers in forest-savanna-transitions and domesticated landscapes, never cultivated. The probable origin of domesticated annatto is in south-western Amazonia (13). Intriguingly, we found a high abundance of *B. urucurana* in the upper Solimões River and especially in Ecuadorian forests.

8. *Genipa americana* L., Rubiaceae

The genipap tree (*Genipa americana*) is commonly cultivated in home-gardens and swiddens of tropical America for its edible fruits and colorant properties. The species is widespread in tropical America, and it occurs in different environments in Amazonia (54). Spontaneous and cultivated genipap trees occur especially in floodplains of white water rivers in Amazonia, but we found a high abundance of the species in south-western forests. The high abundance of genipap in south-western Amazonia is curious, probably associated with riverine vegetation or anthropogenic forests within the Llanos de Mojos (16). The genipap distribution along the Guiana coast fits with its probable origin of domestication (13).

9. *Matisia cordata* Bonpl., Malvaceae

South American sapota (previously recognized as *Quararibea cordata*) is an emergent tree with edible fruits, cultivated in the eastern Andean foothills and lowlands in Peru, Ecuador and Colombia, and along the middle and upper Solimões River in Brazil. The cultivated trees always have larger fruits (300-1000 g) with more pulp and less fiber than those of wild populations (150-400 g) (69). Trees cultivated in swiddens on fertile anthropogenic soils can have even larger fruits (> 1000 g). Western Amazonia was hypothesized as the probable origin of semi-domesticated populations of South American sapota (69), based on morphological analysis of fruit sizes. Sapota trees were found in highest abundances in the southern part of western forests, where wild populations also exist.

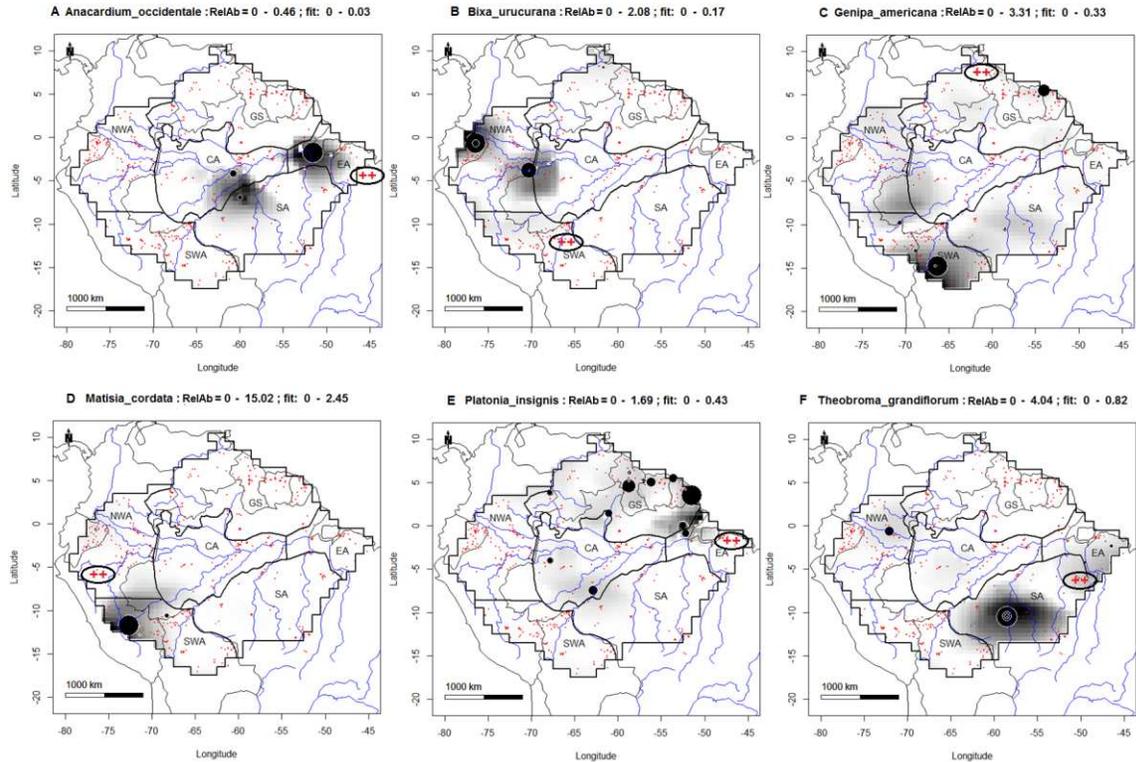
10. *Platonia insignis* Mart., Clusiaceae

Known in Brazil as bacurí (*Platonia insignis*), it is cultivated for its edible fruits, predominantly in the estuary of the Amazonas River and eastern Amazonia into Maranhão and Piauí (54). The species occurs naturally in degraded and sandy areas, and occasionally it occurs in old-growth forests. In some indigenous agroforestry systems bacurí is very

abundant (72), and it can also be favored in fallows through intensive management. The species commonly occurs in eastern Amazonia, whereas it rarely occurs in western Amazonia. Although we did not find any tree in the ATDN forest plots of eastern Amazonia, native populations of bacurí occur mainly in eastern Pará and Maranhão States of Brazil. It was found in forests of the Guiana Shield, Central and Southern Amazonia, with highest abundance on the Guiana coast.

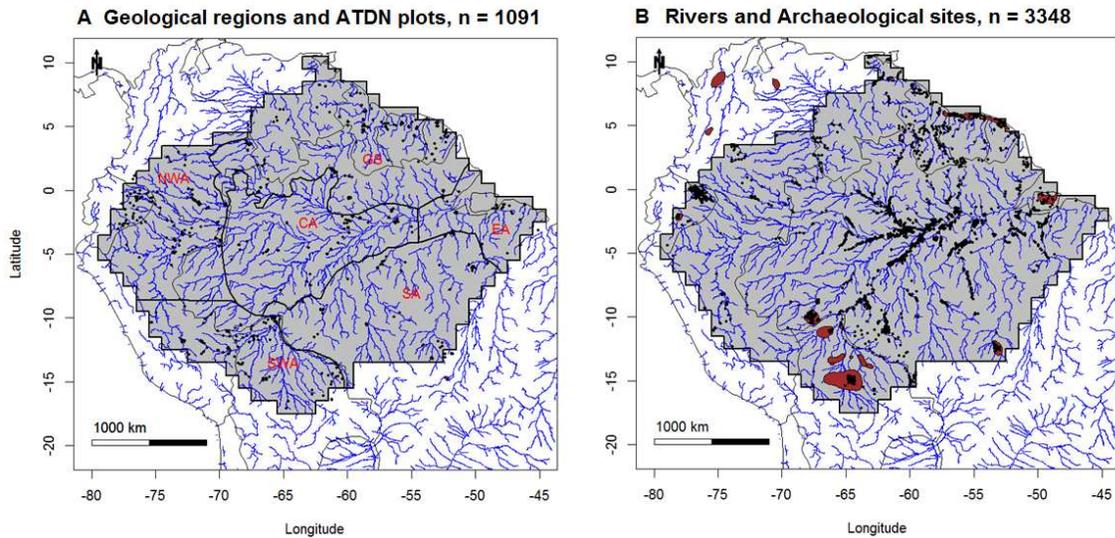
11. *Theobroma grandiflorum* (Willd. ex Spreng.) K. Schumm., Malvaceae

The cupuaçu tree (*Theobroma grandiflorum*) is native in the Brazilian Amazon and is now widely cultivated in Colombia, Venezuela, Ecuador and Costa Rica (54). Cupuaçu fruits are appreciated for their rich flavorful pulp. Cultivated plants tend to produce larger fruits than those collected from forests, certainly due to selection by humans. South-eastern Amazonia was hypothesized as the origin of domestication of the species, because cultivated and native populations are frequent in this region (13). However, genetic analysis was unable to relate wild populations from eastern Amazonia with cultivated cupuaçu, so the origin is still unknown (73). Cupuaçu occurs in forests of almost all Amazonian regions, but it rarely occurs in western Amazonia, although it is one of the most important species cultivated in agroforestry systems in northern Bolivia (74). The high abundance in southern forests is unexpected; no one has ever suggested this region as an origin of cupuaçu.



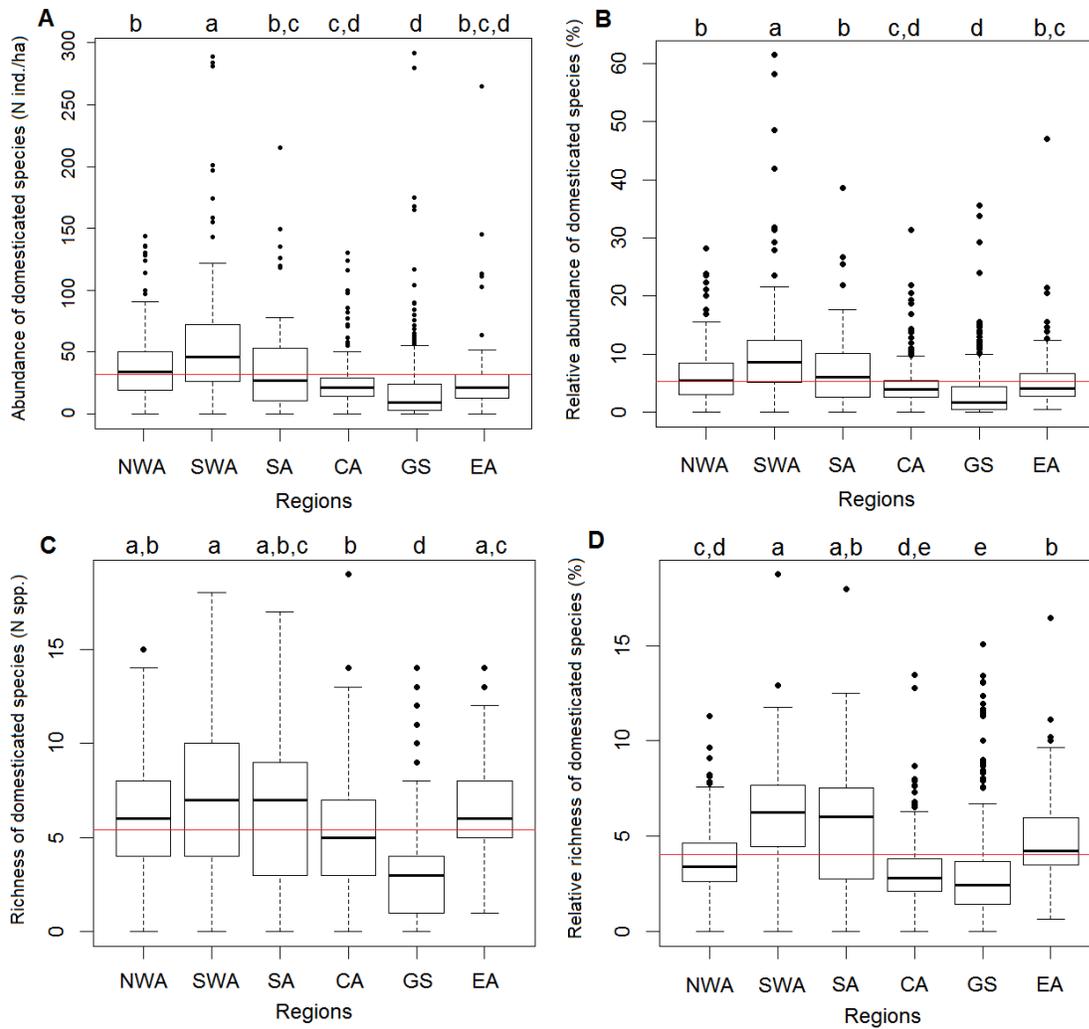
**Fig. S1.**

Distribution maps of domesticated species that are not hyperdominants in Amazonian forests and their probable origins of domestication (9). Distribution maps were estimated by the spatial loess model for six domesticated species that are not hyperdominants: *Anacardium occidentale* (A); *Bixa urucurana* (B); *Genipa americana* (C); *Matisia cordata* (D); *Platonia insignis* (E); *Theobroma grandiflorum* (F). The suspected origin of domestication is indicated in the maps by the symbol (++). The size of black dots indicates the abundance of the species in the plots where the species has been recorded. Red dots are plots where each domesticated species has not been recorded. Shading in maps shows the loess spatial average. The range of the relative abundance in plots (RelAb) and the loess spatial average in individual grid cells (fit) is reported in percentage on the top of each specific map. Amazonia is divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Maps created with custom R scripts. Base map source (country.shp, rivers.shp): ESRI (<http://www.esri.com/data/basemaps>, © Esri, DeLorme Publishing Company).



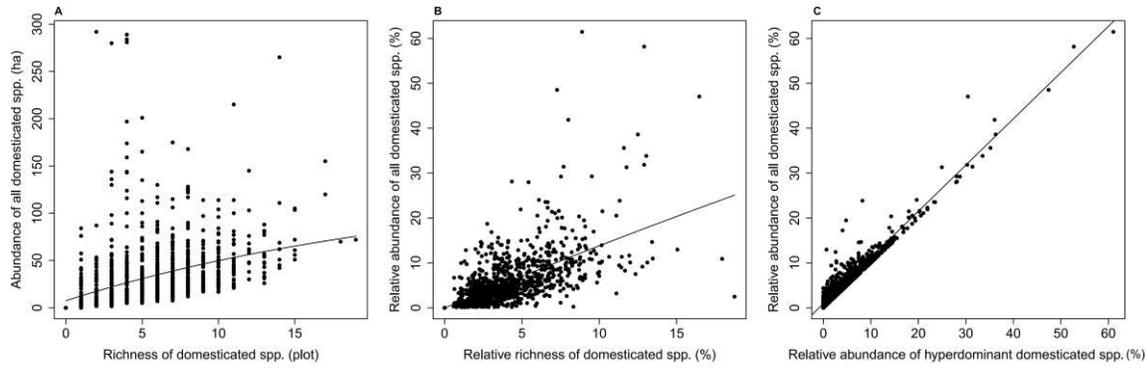
**Fig. S2.**

Maps of ATDN forest plots, geological regions, and archaeological sites overlaying the river network. Black circles show the location of forest plots (A) and black circles and red polygons show the location of archaeological sites and eco-archaeological regions in Amazonia, respectively (B). Red polygons are eco-archaeological regions encompassing numerous earthworks. Amazonia is divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). River network was obtained from the HydroSHEDS dataset with upcell values greater than 15000. Maps created with custom R scripts. Base map source (country.shp): ESRI (<http://www.esri.com/data/basemaps>, © Esri, DeLorme Publishing Company).



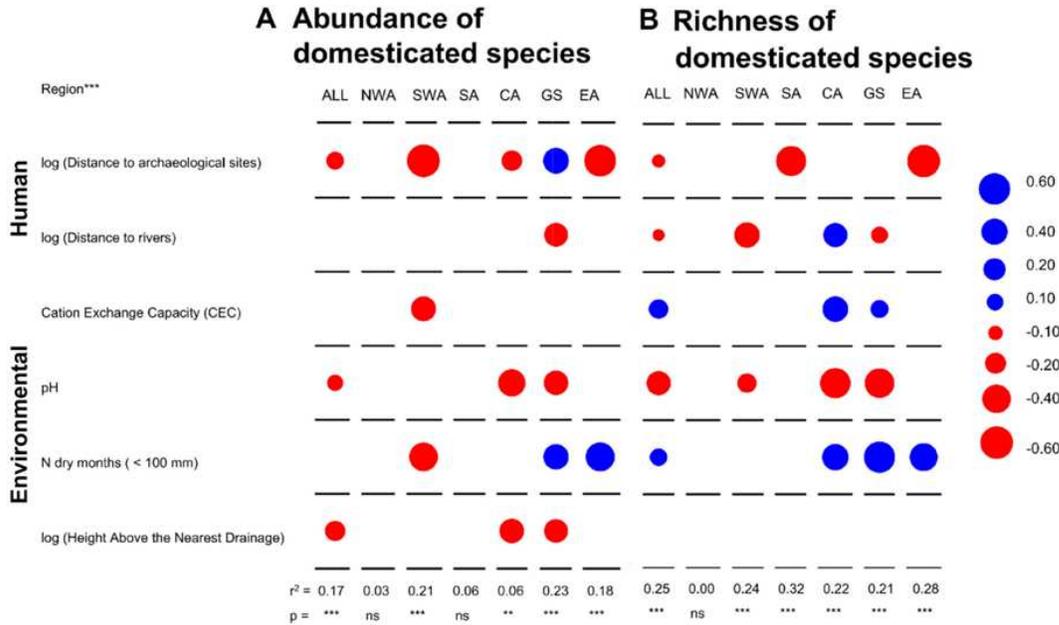
**Fig. S3.**

Abundance and richness of domesticated species in different geological regions. Box plots showing the abundance (A), the relative abundance (B), the richness (C) and the relative richness of domesticated species (D) in the six geological regions of Amazonia (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia; see Fig. S2 for the map of the regions). Different letters indicate significant differences (Tukey post hoc test,  $p \leq 0.05$ ) and the red line indicates the mean value across the whole Amazon.



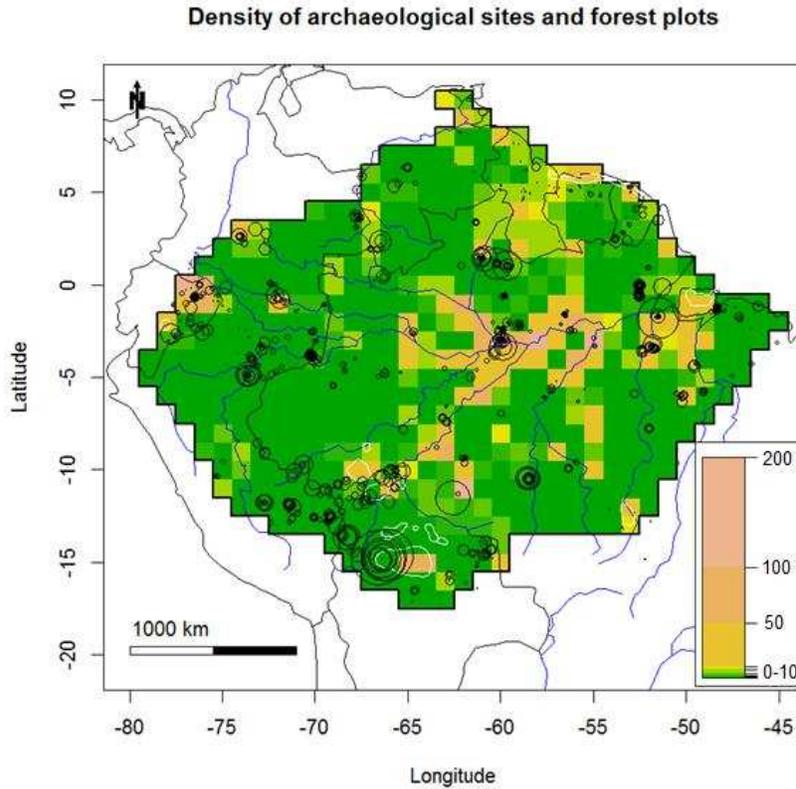
**Fig. S4.**

Abundance of domesticated species and their relationship to the richness and hyperdominance of domesticated species. Relationship between the abundance of 85 domesticated species per hectare (ha) and the richness of domesticated species in forest plots ( $r^2 = 0.15$ , A); relationship between the relative abundance of 85 domesticated species in forest plots and relative richness of domesticated species ( $r^2 = 0.39$ , B); and relationship between the relative abundance of 85 domesticated species and the relative abundance of 20 hyperdominant domesticated species in forest plots ( $r^2 = 0.94$ , C). Non-linear models created with custom R scripts.



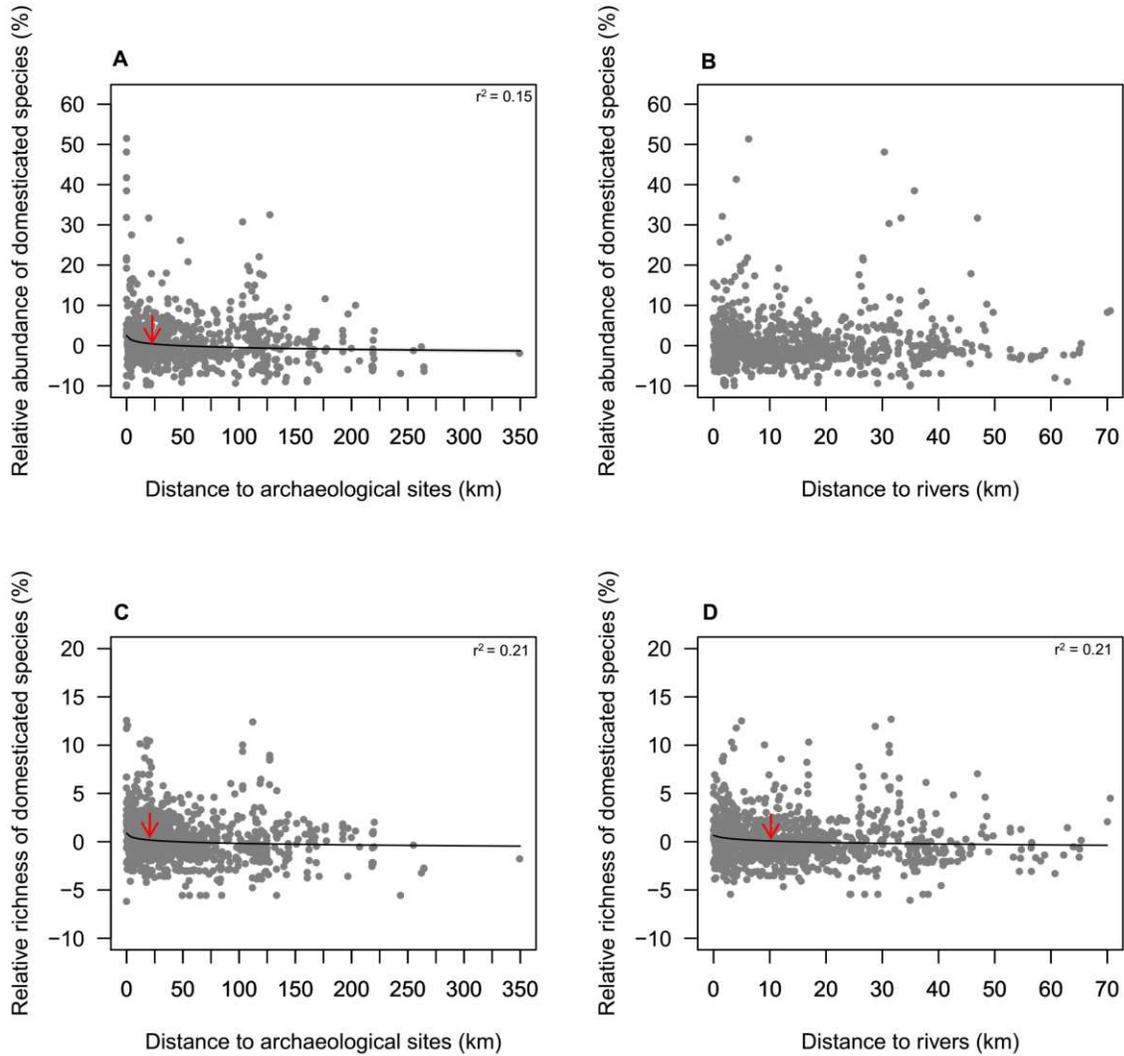
**Fig. S5.**

The abundance and richness of domesticated species as a function of human and environmental variables. Standardized regression coefficients for the abundance (A) and the richness of 85 domesticated species (B) as a function of human factors (distance to archaeological sites, distance to navigable rivers) and environmental conditions (soil Cation Exchange Capacity, soil pH, number of dry months and Height Above the Nearest Drainage). Circle size represents the relative contribution of the predictors, shown by standardized coefficients at the Amazonia-wide level (All) and region-level regression models (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Red circles indicate negative effects and blue circles positive effects. Standardized coefficients are presented only for significant relations analyzed in the models ( $p \leq 0.05$ ). Adjusted  $r^2$  and significance codes ( $p$  values:  $\leq 0.001$  ‘\*\*\*’;  $\leq 0.01$  ‘\*\*’;  $\leq 0.05$  ‘\*’;  $> 0.05$  ‘ns’) are presented for the effect of human factors and environmental conditions at the Amazonia-wide level (All; using mixed-effect models and region as random factors), and at the region level (using multiple regression models).



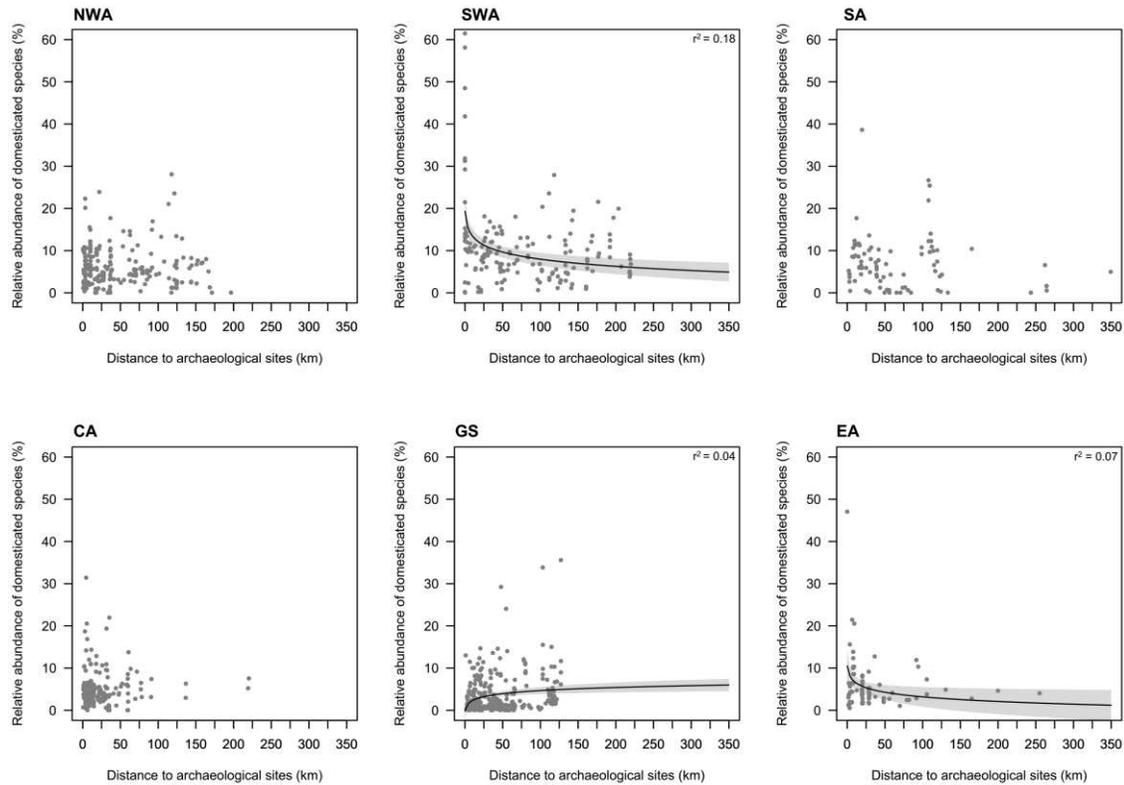
**Fig. S6.**

A heat map of archaeological sites in Amazonia and the variation of the relative abundance of domesticated species in forest plots. The pink-green background shows the density of archaeological sites in a 1°-grid cell scale, ranging from 0-200 sites per cell. White polygons show the location of eco-archaeological regions encompassing numerous earthworks. Circle sizes represent the variation of the relative abundance of domesticated species in forest plots shown in Fig. S4B. Maps created with custom R scripts. Base map source (country.shp, rivers.shp): ESRI (<http://www.esri.com/data/basemaps>, © Esri, DeLorme Publishing Company).



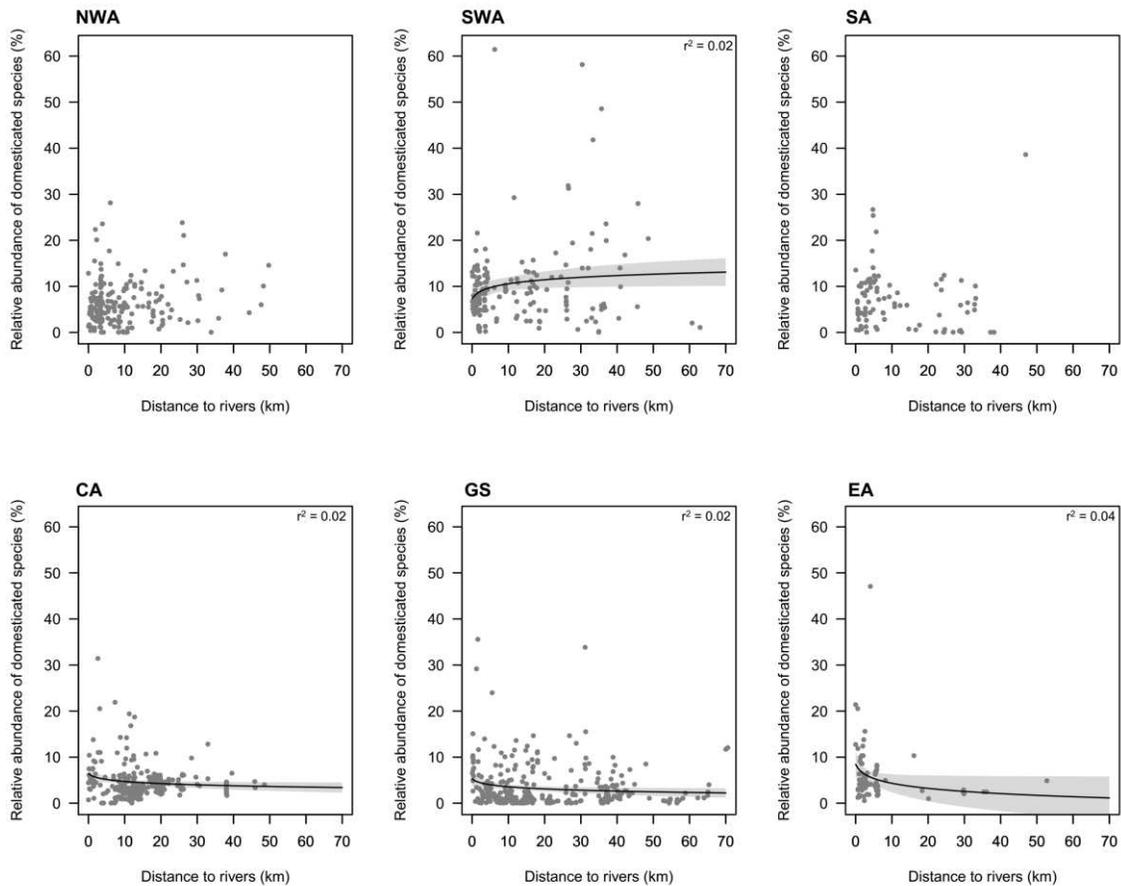
**Fig. S7.**

Relationships between the relative abundance and richness of domesticated species and the distance to archaeological sites (and eco-archaeological regions) and rivers in Amazonia. Note that some of these relationships may not match with the results obtained by the mixed-effect models (Fig. 3), as these take into account the effects of environment. Black lines presented only for the significant relationships ( $p \leq 0.05$ ). Red arrows are presented to show the distance for which a pronounced decrease occurs. Mixed models fit by t-tests use Satterthwaite approximations:  $\text{lmer}(\text{domestication} \sim \log_{10}(\text{distance} + 1) + (1 \mid \text{Region}))$ . Contrast plots were created with `visreg` function in R.



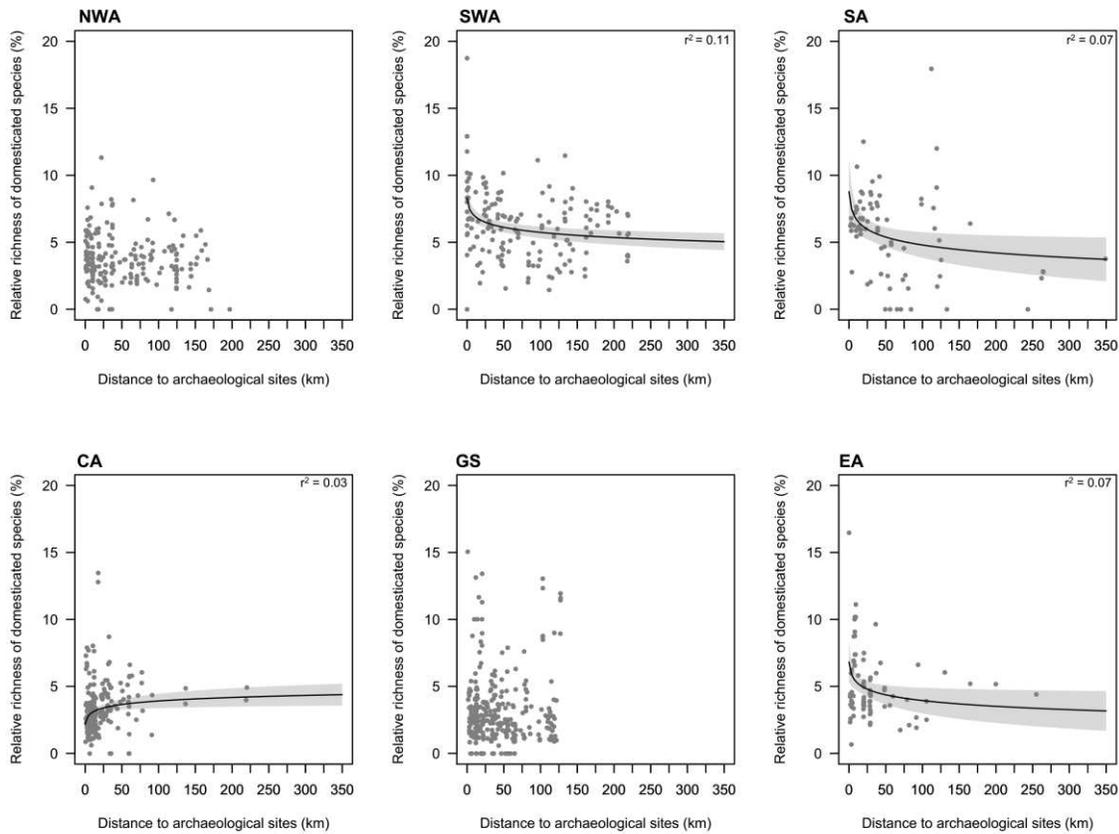
**Fig. S8.**

Relationships between the relative abundance of domesticated species and the distance to archaeological sites or eco-archaeological regions within Amazonian regions. Amazonia was divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Note that some of these relationships may not match with the results obtained by the multiple regression models (Fig. 3), as these take into account the effects of environment. Black lines presented only for the significant relationships ( $p \leq 0.05$ ). Simple models were used for each geological region:  $\text{lm}(\text{domestication} \sim \log_{10}(\text{distance} + 1))$ . Conditional plots were created with `visreg` function in R.



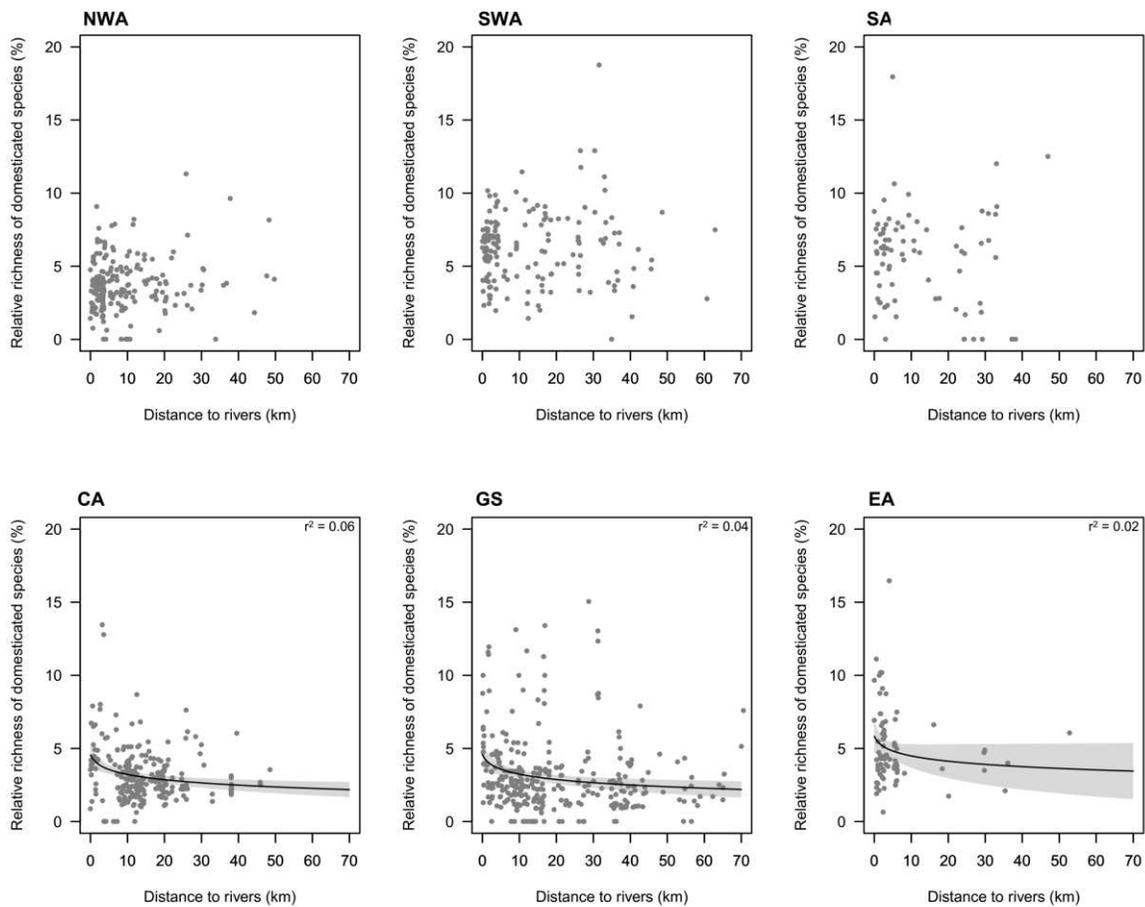
**Fig. S9.**

Relationships between the relative abundance of domesticated species and the distance to rivers within Amazonian regions. Amazonia was divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Note that some of these relationships may not match with the results obtained by the multiple regression models (Fig. 3), as these take into account the effects of environment. Black lines presented only for the significant relationships ( $p \leq 0.05$ ). Simple models were used for each geological region:  $\text{lm}(\text{domestication} \sim \log_{10}(\text{distance} + 1))$ . Conditional plots were created with visreg function in R.



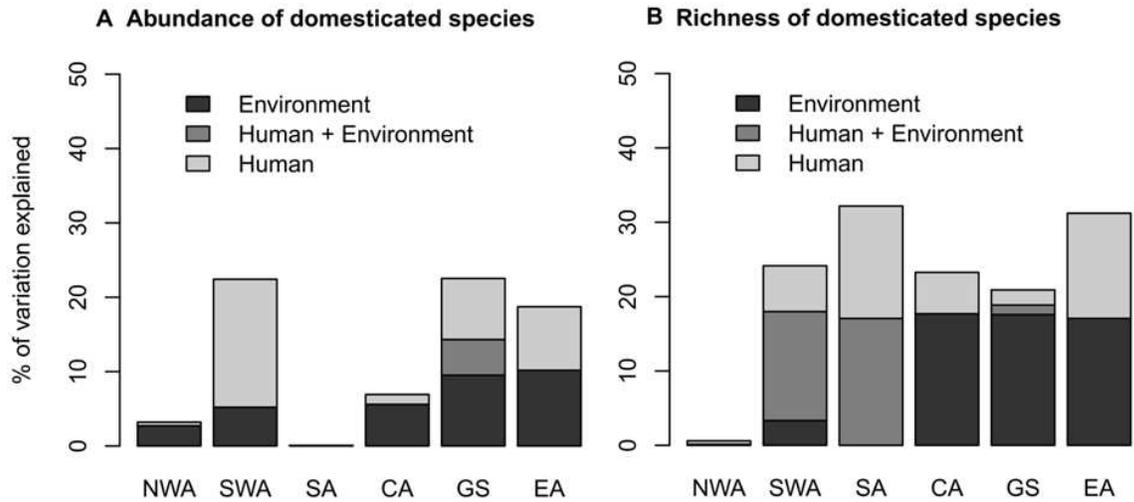
**Fig. S10.**

Relationships between the relative richness of domesticated species and the distance to archaeological sites and eco-archaeological regions within Amazonian regions. Amazonia was divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Note that some of these relationships may not match with the results obtained by the multiple regression models (Fig. 3), as these take into account the effects of environment. Black lines presented only for the significant relationships ( $p \leq 0.05$ ). Simple models were used for each geological region:  $\text{lm}(\text{domestication} \sim \log_{10}(\text{distance} + 1))$ . Conditional plots were created with `visreg` function in R.



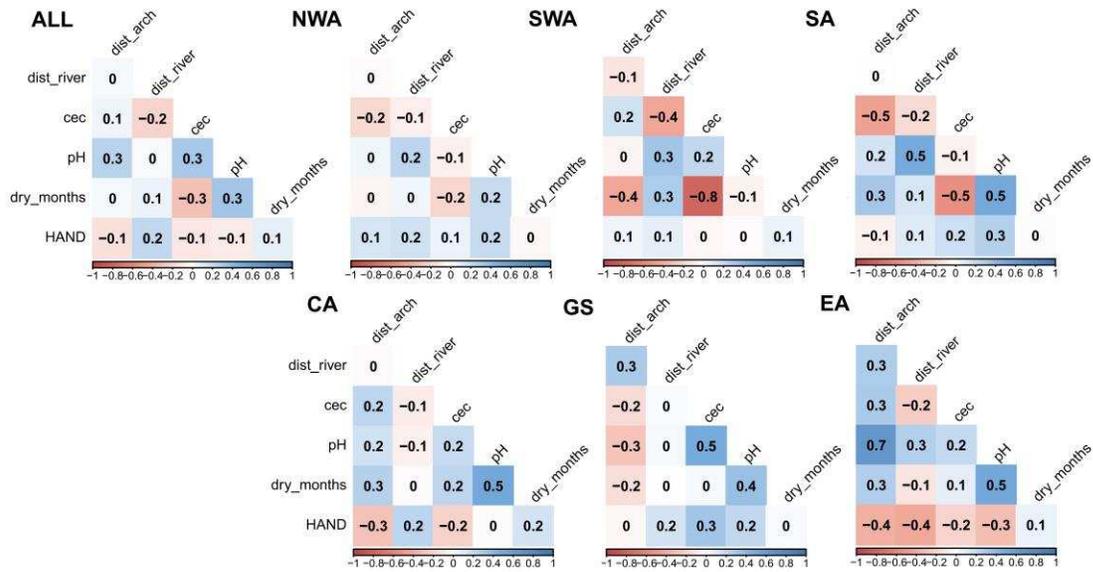
**Fig. S11.**

Relationships between the relative richness of domesticated species and the distance to rivers within Amazonian regions. Amazonia was divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Note that some of these relationships may not match with the results obtained by the multiple regression models (Fig. 3), as these take into account the effects of environment. Black lines presented only for the significant relationships ( $p \leq 0.05$ ). Simple models were used for each geological region:  $\text{lm}$  (domestication  $\sim \log_{10}(\text{distance} + 1)$ ). Conditional plots were created with `visreg` function in R.



**Fig. S12.**

Relative contributions of human and environmental variables for explaining variation in abundance and richness of domesticated species in Amazonian forests. The figure shows the partitioning of variation in abundance (A) and richness (B) of domesticated species uniquely explained by environmental (dark gray) or human factors (light gray), and the variation jointly explained by both (gray). Variance partitioning was conducted over the results of multiple regression analyses presented in Fig. 3. Amazonia was divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia).



**Fig. S13.**

Matrices of Spearman's correlation coefficients between pairs of explanatory variables used in the multiple regression models in Amazonia and each region. Amazonia was divided in six geological regions (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Red indicates negative correlations and blue positive correlations. The intensity of the color indicates the strength of the correlation. dist\_arch = log-transformed distance to archaeological sites, dist\_river = log-transformed distance to navigable rivers, cec = soil cation exchange capacity, pH = soil pH, dry\_months = number of dry months, HAND = log-transformed Height Above the Nearest Drainage.

**Table S1.**

List of the 73 hyperdominant species studied. Three groups of 20 non-domesticated species with estimated population sizes (17) comparable to those of the 20 hyperdominant domesticated species are presented. The first group contains species primarily dispersed by non-human primates, the second consists of species selected based on specific criteria described in the materials and methods, and the third of species selected at random.

Species	Estimated population <sup>27</sup>	Domesticated species	Dispersed by primates	Specific selection	Random selection
<i>Euterpe precatória</i>	5.21 x 10 <sup>9</sup>	yes	no	no	no
<i>Euterpe oleracea</i>	3.78 x 10 <sup>9</sup>	yes	no	no	no
<i>Oenocarpus bataua</i>	3.71 x 10 <sup>9</sup>	yes	no	no	no
<i>Astrocaryum murumuru</i>	2.41 x 10 <sup>9</sup>	yes	no	no	no
<i>Hevea brasiliensis</i>	1.91 x 10 <sup>9</sup>	yes	no	no	no
<i>Mauritia flexuosa</i>	1.43 x 10 <sup>9</sup>	yes	no	no	no
<i>Theobroma cacao</i>	1.32 x 10 <sup>9</sup>	yes	yes	no	no
<i>Theobroma subincanum</i>	1.26 x 10 <sup>9</sup>	yes	yes	no	no
<i>Oenocarpus bacaba</i>	1.24 x 10 <sup>9</sup>	yes	no	no	no
<i>Theobroma speciosum</i>	1.20 x 10 <sup>9</sup>	yes	yes	no	no
<i>Attalea maripa</i>	9.65 x 10 <sup>8</sup>	yes	no	no	no
<i>Attalea phalerata</i>	5.91 x 10 <sup>8</sup>	yes	no	no	no
<i>Pouteria caimito</i>	5.79 x 10 <sup>8</sup>	yes	yes	no	no
<i>Astrocaryum aculeatum</i>	5.39 x 10 <sup>8</sup>	yes	no	no	no
<i>Caryocar glabrum</i>	5.22 x 10 <sup>8</sup>	yes	no	no	no
<i>Spondias mombin</i>	4.95 x 10 <sup>8</sup>	yes	yes	no	no
<i>Garcinia macrophylla</i>	4.65 x 10 <sup>8</sup>	yes	yes	no	no
<i>Inga ynga</i>	4.29 x 10 <sup>8</sup>	yes	yes	no	no
<i>Pourouma cecropiifolia</i>	4.25 x 10 <sup>8</sup>	yes	yes	no	no
<i>Bertholletia excelsa</i>	4.17 x 10 <sup>8</sup>	yes	no	no	no
<i>Pseudolmedia laevis</i>	4.30 x 10 <sup>9</sup>	no	yes	no	no
<i>Brosimum lactescens</i>	2.28 x 10 <sup>9</sup>	no	yes	no	yes
<i>Helicostylis tomentosa</i>	1.79 x 10 <sup>9</sup>	no	yes	no	no
<i>Micropholis guyanensis</i>	1.35 x 10 <sup>9</sup>	no	yes	no	no
<i>Ecclinusa guianensis</i>	1.18 x 10 <sup>9</sup>	no	yes	no	no
<i>Brosimum guianense</i>	1.04 x 10 <sup>9</sup>	no	yes	no	no
<i>Brosimum rubescens</i>	1.03 x 10 <sup>9</sup>	no	yes	no	yes
<i>Chrysophyllum sanguinolentum</i>	1.02 x 10 <sup>9</sup>	no	yes	no	no
<i>Leonia glycyarpa</i>	1.02 x 10 <sup>9</sup>	no	yes	no	no
<i>Minquartia guianensis</i>	9.87 x 10 <sup>8</sup>	no	yes	no	no
<i>Pourouma minor</i>	9.68 x 10 <sup>8</sup>	no	yes	no	no
<i>Quararibea wittii</i>	5.94 x 10 <sup>8</sup>	no	yes	no	no
<i>Inga thibaudiana</i>	5.77 x 10 <sup>8</sup>	no	yes	no	no
<i>Manilkara bidentata</i>	5.59 x 10 <sup>8</sup>	no	yes	no	no

Species	Estimated population <sup>27</sup>	Domesticated species	Dispersed by primates	Specific selection	Random selection
<i>Pouteria cuspidate</i>	5.31 x 10 <sup>8</sup>	no	yes	no	no
<i>Brosimum utile</i>	4.89 x 10 <sup>8</sup>	no	yes	no	no
<i>Bocageopsis multiflora</i>	4.62 x 10 <sup>8</sup>	no	yes	no	no
<i>Pouteria reticulata</i>	4.51 x 10 <sup>8</sup>	no	yes	no	no
<i>Pourouma bicolor</i>	4.47 x 10 <sup>8</sup>	no	yes	no	yes
<i>Apeiba tibourbou</i>	4.14 x 10 <sup>8</sup>	no	yes	no	no
<i>Protium altissimum</i>	5.21 x 10 <sup>9</sup>	no	no	yes	no
<i>Iriarteia deltoidea</i>	4.07 x 10 <sup>9</sup>	no	no	yes	no
<i>Trattinnickia burserifolia</i>	2.78 x 10 <sup>9</sup>	no	no	yes	no
<i>Socratea exorrhiza</i>	2.68 x 10 <sup>9</sup>	no	no	yes	no
<i>Attalea butyracea</i>	1.78 x 10 <sup>9</sup>	no	no	yes	no
<i>Eperua leucantha</i>	1.84 x 10 <sup>9</sup>	no	no	yes	yes
<i>Clathrotropis macrocarpa</i>	1.35 x 10 <sup>9</sup>	no	no	yes	yes
<i>Pentaclethra macroloba</i>	1.34 x 10 <sup>9</sup>	no	no	yes	no
<i>Dicymbe corymbosa</i>	1.26 x 10 <sup>9</sup>	no	no	yes	no
<i>Virola calophylla</i>	1.22 x 10 <sup>9</sup>	no	no	yes	no
<i>Micrandra spruceana</i>	9.57 x 10 <sup>8</sup>	no	no	yes	yes
<i>Protium decandrum</i>	5.87 x 10 <sup>8</sup>	no	no	yes	no
<i>Cenostigma tocantinum</i>	5.76 x 10 <sup>8</sup>	no	no	yes	no
<i>Ocotea aciphylla</i>	5.19 x 10 <sup>8</sup>	no	no	yes	no
<i>Conceveiba guianensis</i>	5.17 x 10 <sup>8</sup>	no	no	yes	no
<i>Protium trifoliolatum</i>	4.93 x 10 <sup>8</sup>	no	no	yes	no
<i>Eschweilera tessmannii</i>	4.68 x 10 <sup>8</sup>	no	no	yes	no
<i>Ocotea cernua</i>	4.31 x 10 <sup>8</sup>	no	no	yes	no
<i>Trichilia pleeana</i>	4.25 x 10 <sup>8</sup>	no	no	yes	no
<i>Cedrelinga cateniformis</i>	4.17 x 10 <sup>8</sup>	no	no	yes	yes
<i>Aspidosperma excelsum</i>	1.13 x 10 <sup>9</sup>	no	no	no	yes
<i>Goupia glabra</i>	9.88 x 10 <sup>8</sup>	no	no	no	yes
<i>Lecythis idatimon</i>	9.09 x 10 <sup>8</sup>	no	no	no	yes
<i>Sagotia brachysepala</i>	8.67 x 10 <sup>8</sup>	no	no	no	yes
<i>Inga alba</i>	7.82 x 10 <sup>8</sup>	no	no	no	yes
<i>Iryanthera laevis</i>	6.82 x 10 <sup>8</sup>	no	no	no	yes
<i>Aparisthium cordatum</i>	6.18 x 10 <sup>8</sup>	no	no	no	yes
<i>Scleronema micranthum</i>	6.12 x 10 <sup>8</sup>	no	no	no	yes
<i>Eperua grandiflora</i>	5.41 x 10 <sup>8</sup>	no	no	no	yes
<i>Leonia crassa</i>	4.77 x 10 <sup>8</sup>	no	no	no	yes
<i>Laetia procera</i>	4.73 x 10 <sup>8</sup>	no	no	no	yes
<i>Hura crepitans</i>	4.21 x 10 <sup>8</sup>	no	no	no	yes
<i>Pouteria procera</i>	3.61 x 10 <sup>8</sup>	no	no	no	yes

**Table S2.**

Results of the multiple regression models of the relative abundance and richness of domesticated species as functions of human and environmental variables using Bonferroni correction. Standardized regression coefficients for the relative abundance of domesticated species and the relative richness of domesticated species as a function of human factors (dist\_arch = log-transformed distance to archaeological sites, dist\_river = log-transformed distance to navigable rivers) and environmental conditions (cec = soil cation exchange capacity, pH = soil pH, dry\_months = number of dry months, HAND = log-transformed Height Above the Nearest Drainage). Standardized coefficients are shown at the Amazonia-wide level (All) and region-level regression models (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia) and presented only for significant relations analyzed in the models before Bonferroni correction ( $p \leq 0.05$ ). Significant codes are presented for each variable using Bonferroni correction: ( $p \leq 0.0001$  ‘\*\*\*’;  $\leq 0.0014$  ‘\*\*’;  $\leq 0.0071$  ‘\*’) are presented for the effect of human factors and environmental conditions at the Amazonia-wide level (All; using mixed-effect models and region as random factors), and at the region level (using multiple regression models).

Region (number of plots)	Variables	Relative abundance		Relative richness	
		Standardized coefficient	Correction-value	Standardized coefficient	Correction-value
All (1091)	dist_arch	-0.13	***	-0.10	**
	cec				
	pH			-0.12	*
	dry_months	0.11	*	0.25	***
	HAND	-0.17	***	-0.14	***
NWA (197)	dist_arch				
	cec				
	pH				
	dry_months			0.22	*
SWA (158)	dist_arch	-0.49	***	-0.35	***
	dist_rivers				
	pH				
	dry_months	-0.38	**		
SA (86)	HAND				
	dist_arch				
	dist_rivers				
	cec				
	pH				
CA (250)	dry_months				
	HAND				
	dist_arch				
	cec			0.17	*
	pH	-0.41	***	-0.62	***
	dry_months	0.29	**	0.58	***
	HAND	-0.27	**	-0.32	***

<b>Region (number of plots)</b>	<b>Variables</b>	<b>Standardized coefficient</b>	<b>Correction-value</b>	<b>Standardized coefficient</b>	<b>Correction-value</b>
GS (317)	dist_arch	0.22	***		
	cec				
	pH			-0.28	***
	dry months			0.51	***
	HAND			-0.17	*
EA (83)	dist_arch	-0.50	*	-0.63	***
	dist_rivers				
	cec				
	pH				
	dry months			0.39	*
	HAND				

**Table S3.**

Results of the multiple regression models with data from the plots on archaeological sites and eco-archaeological regions and without these data. Standardized coefficients (Beta coefficients) and p values are presented for the relative abundance of domesticated species and the relative richness of domesticated species as a function of human factors (dist\_arch = log-transformed distance to archaeological sites, dist\_river = log-transformed distance to navigable rivers) and environmental conditions (cec = soil cation exchange capacity, pH = soil pH, dry\_months = number of dry months, HAND = log-transformed Height Above the Nearest Drainage) at the Amazonia-wide level (All) and region-level regression models (SWA, south-western Amazonia and EA, eastern Amazonia) where plots on eco-archaeological regions are located. Standardized coefficients are presented only for significant relations analyzed in the models ( $p \leq 0.05$ ). Significant codes (p values:  $\leq 0.001$ ; '\*\*\*';  $\leq 0.01$  '\*\*';  $\leq 0.05$  '\*';  $> 0.05$  'ns') are presented for the effect of human factors and environmental conditions at the Amazonia-wide level (All; using mixed-effect models and region as random factors), and at the region level (using multiple regression models).

Region (number of plots)	Variables	Relative abundance				Relative richness			
		Beta coefficient with plots	p	Beta coefficient without plots	p	Beta coefficient with plots	p	Beta coefficient without plots	p
All (1091)	dist_arch	-0.13	***			-0.10	***		
	dist_river			-0.06	*	-0.09	**		
	cec								
	pH			-0.16	***	-0.12	**	-0.15	***
	dry_months	0.11	**	0.14	***	0.25	***	0.27	***
	HAND	-0.17	***	-0.16	***	-0.14	***	-0.13	***
SWA (158)	dist_arch	-0.49	***	-0.21	*	-0.35	***	-0.23	*
	dist_river			0.22	*				
	cec	-0.31	**			-0.33	**	-0.37	*
	pH								
	dry_months	-0.38	***	-0.31	*				
	HAND			-0.26	**				
EA (83)	dist_arch	-0.50	**	-0.36	*	-0.63	***	-0.54	***
	dist_river			-0.49	***			-0.23	*
	cec			-0.22	*				
	pH			0.61	***	0.39	*	0.55	**
	dry_months	0.32	*			0.39	**	0.29	*
	HAND								

**Database S1 (separate file)**

List of 85 species with populations that were likely domesticated, semi-domesticated or incipiently domesticated by pre-Columbian peoples in Amazonia and elsewhere in the Americas. The main use of each species, the degree of domestication (6), the rank of dominance according to ter Steege et al. (17), the relative frequency of the species in each region (%), the number of regions where the species occurs and information about cultivation are provided. Numbers provided below each region correspond to the number of forest plots inventoried in each region (NWA, north-western Amazonia; SWA, south-western Amazonia; SA, southern Amazonia; CA, central Amazonia; GS, Guiana Shield; EA, eastern Amazonia). Species that were not present in Clement (6) were classified here as incipiently domesticated species. References for evidence of domestication of each species are presented in this table. The information about cultivation was obtained from Mansfeld's World Database of Agricultural and Horticultural Crops (54) and other sources described in the table. Species are listed based on the rank of dominance according to ter Steege et al. (17). All species with ranking  $\leq 227$  are considered "hyperdominant" species.

**Database S2 (separate file): Plot meta data**

ATDNNR: number in ATDN database

Country: country in which plot is located

Subdivision: mostly province

Site: site name

PlotCode: Unique ATDN plot code

Region: Geological region in which plot is located

Latitude, Longitude

PlotSize: plot size in ha.

PlotType: single: 1 single contiguous area; combi; few plots very close added together; pcq: plots built from point center quarter data.

DBHmin: min dbh cut off

Year\_est: Year in which the plot was established (not necessarily the census year)

Owner/contact: Owner of plot data

Reference from Endnote: literature reference of plot data. This source does not always contain the full data set.

## References:

43. B. Lehner, K. Verdin, A. Jarvis, New global hydrography derived from spaceborne elevation data. *Eos, Trans. Amer. Geophys.* **89**, 93-94 (2008).
44. BRASIL, “Manual de Construção da Base Hidrográfica Ortocodificada” (Brasília, ANA, SGI, 2007); available online at [www.ana.gov.br](http://www.ana.gov.br).
45. E. J. Fittkau, Esboço de uma divisão ecológica da região amazônica. *Proc. Symp. Biol. Trop. Amaz.*, 363–372 (1971).
46. T. Hengl et al., SoilGrids1km-Global Soil Information based on automated mapping. *PLoS One* **9**, e105992 (2013). doi: [info:doi/10.1371/journal.pone.0114788](https://doi.org/10.1371/journal.pone.0114788).
47. C. Kummerow, W. Barnes, T. Kozu, J. Shiue, J. Simpson, The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Ocean Tech.* **15**, 809–817 (1998).
48. A. Nobre et al., Height above the nearest drainage - a hydrologically relevant new terrain model. *J. Hydrol.* **404**, 13-29 (2011).
49. P. Perrut de Lima, “Caracterização da variabilidade genética, sistema de cruzamento e parâmetros de germinação e emergência de *Euterpe precatoria* Martius em populações do baixo rio Solimões” thesis, Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil (2014).
50. M. Smith, C. Fausto, Socialidade e diversidade de pequis (*Caryocar brasiliense*, *Caryocaraceae*) entre os Kuikuro do alto rio Xingu (Brasil). *Bol. Mus. Para. Emílio Goeldi. Cienc. Hum.* **11**, 87-113 (2016).
51. P. A. Moreira, J. Lins, G. Dequigiovanni, E. A. Veasey, C. R. Clement, The domestication of Annatto (*Bixa orellana*) from *Bixa urucurana* in Amazonia. *Econ. Bot.* **69**, 127-135 (2015).
52. J. Sosnowska, A. Walanus, H. Balslev, Asháninka palm management and domestication in the Peruvian Amazon. *Hum. Ecol.* **43**, 1-16 (2015).
53. N. García et al., Management of the palm *Astrocaryum chambira* Burret (*Arecaceae*) in northwest Amazon. *Acta Bot. Bras.* **29**, 45-57 (2015).
54. P. Hanelt, R. Büttner, R. Mansfeld, Mansfeld's Encyclopedia of Agricultural and Horticultural Crops: Except Ornamentals, (Berlin, Springer, 2001); available online at <http://mansfeld.ipk-gatersleben.de/>.
55. H. Dempewolf, L. H. Rieseberg, Q. C. Cronk, Crop domestication in the *Compositae*: a family-wide trait assessment. *Genet. Resour. Crop Ev.* **55**, 1141-1157 (2008).
56. K. Hammer, K. Khoshbakht, A domestication assessment of the big five plant families. *Genet. Resour. Crop Ev.* **62**, 665–689 (2015).
57. B Boyle et al., The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **16**, (2013). doi:10.1186/1471-2105-14-16.
58. R Development Core Team, R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2012); available online at [www.R-project.org](http://www.R-project.org).
59. C. Levis et al., Historical human footprint on modern tree species composition in the Purus-Madeira interfluve, central Amazonia. *PLoS One* **7**, e48559 (2012). doi:10.1371/journal.pone.0048559.
60. J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R Core Team, nlme: Linear and Nonlinear Mixed Effects Models. (Comprehensive R Archive Network, 2016); available online at <http://cran.r-project.org/package=nlme>.

61. D. Lüdecke, sjPlot: Data Visualization for Statistics in Social Science. (The Comprehensive R Archive Network, 2016); available online at <https://cran.r-project.org/package=sjPlot>
62. P. Breheny, W. Burchett, Visualization of Regression Models Using visreg. (Comprehensive R Archive Network, 2013); available online at <https://cran.r-project.org/web/packages/visreg/index.html>
63. P. Legendre, Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *J. Plant Ecol.* **1**, 3-8 (2008).
64. J. Oksanen et al., '*vegan*': *Community ecology package*. (The Comprehensive R Archive Network, 2016); available online at <https://github.com/vegandevs/vegan>.
65. G. H. Shepard, H. Ramirez, "Made in Brazil": Human dispersal of the Brazil Nut (*Bertholletia excelsa*, Lecythidaceae) in Ancient Amazonia. *Econ. Bot.* **65**, 44-65 (2011).
66. C. A. Peres, C. Baider, Seed dispersal, spatial distribution and population structure of Brazil nut trees (*Bertholletia excelsa*) in southeastern Amazonia. *J. Trop. Ecol.* **13**, 595-616 (1997).
67. P. M. Paiva, M. C. Guedes, C. Funi, Brazil nut conservation through shifting cultivation. *For. Ecol. Manage.* **261**, 508-514 (2011).
68. P. S. Sujii, K. Martins, L. H. de Oliveira Wadt, V. n. C. R. Azevedo, V. N. Solferini, Genetic structure of *Bertholletia excelsa* populations from the Amazon at different spatial scales. *Conserv. Genet.* **16**, 955-964 (2015).
69. C. R. Clement, A center of crop genetic diversity in western Amazonia. *BioScience* **39**, 624-631 (1989).
70. I. K. Dawson et al., Origins and genetic conservation of tropical trees in agroforestry systems: a case study from the Peruvian Amazon. *Conserv. Genet.* **9**, 361-372 (2008).
71. J. de Paiva, L. d. M. Barros, J. Cavalcanti, in *Breeding Plantation Tree Crops: Tropical Species*, S. M. Jain, P. M. Priyadarshan, Eds. (Springer, 2009), pp. 287-324.
72. W. Balée, Indigenous transformation of Amazonian forests: an example from Maranhão, Brazil. *L'Homme* **33**, 231-254 (1993).
73. R. M. Alves, A. M. Sebbenn, A. S. Artero, C. R. Clement, A. Figueira, High levels of genetic divergence and inbreeding in populations of cupuassu (*Theobroma grandiflorum*). *Tree Genet. Genomes* **3**, 289-298 (2007).
74. V. A. Vos, O. Vaca, A. Cruz, "Sistemas Agroforestales en la Amazonía Boliviana" (Centro de Investigación y Promoción del Campesinado (CIPCA), 2015).
75. N. Smith, Ed., *Palms and People in the Amazon* (Springer Series in Geobotany Studies, 2015).
76. G. Morcote-Ríos, R. Bernal, Remains of palms (Palmae) at archaeological sites in the New World: A review. *Bot. Rev.* **67**, 309-350 (2001).
77. R. J. Seibert, The uses of *Hevea* for food in relation to its domestication. *Ann. Missouri Bot. Gard.* **35**, 117-121 (1948).
78. A. C. Roosevelt et al., Paleoindian cave dwellers in the Amazon: the peopling of the Americas. *Science* **272**, 373-384 (1996).
79. P. B. Cavalcante, Ed., *Frutas comestíveis na Amazônia* (Museu Paraense Emílio Goeldi, 2010).

## Custom R scripts:

```
##### map scripts #####
##### function to add countries and rivers to maps of Amazon #####
#load the shapefiles
countries = readOGR("D:/Documents/GIS Data/ESRIDATA/WORLD", "CNTRY92")
rivers = readOGR("D:/Documents/GIS Data/ESRIDATA/WORLD", "RIVERS")
#if higher resolution needed
#countries = readOGR("D:/Documents/GIS Data/ESRIDATA/WORLD", "CNTRY98")
#rivers = readOGR("D:/Documents/GIS Data/ESRIDATA/WORLD", "rivers98")
forestborder = readOGR("D:/Documents/GIS Data/neotropics", "forestborder")
#regions
regions = readOGR("D:/Documents/GIS Data/neotropics/RAINFOR", "itsct1")
add.geography = function(draw.countries = T, draw.rivers = T, draw.forestborder = F,
                          draw.regions = F, r.color = "black", border = "black",
                          add.arrow = T, add.scale = T){
  force(border)
  if (draw.countries == T) plot(countries, xlim = c(-80,-45), ylim = c(-20,10), border =
border, add = T, asp = 1)
  if (draw.rivers == T) plot(rivers, xlim = c(-80,-45), ylim = c(-20,10), col = 'blue',
add = T, asp = 1)
  if (draw.forestborder == T) plot(forestborder, xlim = c(-80,-45), ylim = c(-20,10),
col = '0', border = border,add = T, lwd = 2, asp = 1)
  if (draw.regions == T){
    plot(regions, xlim = c(-80,-45), ylim = c(-20,10), col = '0',
border = border, add = T, lwd = 2)
    text(-58,2.5,"GS", col = r.color)
    text(-48,-3,"EA", col = r.color)
    text(-55,-8,"SA", col = r.color)
    text(-63,-2.5,"CA", col = r.color)
    text(-65,-13.5,"SWA", col = r.color)
    text(-73.5, 1,"NWA", col = r.color)
  }
  if(add.arrow == T) SpatialPolygonsRescale(layout.north.arrow(), offset = c(-80, 9) ,
scale = 2,
fill = c("black", "black"), plot.grid = F)
  if(add.scale == T){
    SpatialPolygonsRescale(layout.scale.bar(), offset = c(-80, -20) , scale = 10/1.11,
fill = c("transparent", "black"), plot.grid = F)
    text(-77.8, -18.5 ,"1000 km")
  }
}
##### End function add countries and rivers to maps of Amazon #####
##### map characteristics in loess map #####
```

```

map.loess = function(z, Longitude, Latitude, res = 1,
                    span = 0.75, degree = 2, se = T,
                    predict = T, surface = "direct",
                    co = 0, draw.regions = F, r.color = "black",
                    name = "", draw.legend = T,
                    blocks = T, dots = T, c.col = "white",
                    grid.color = c("white", "black"),
                    n.colors = 256, pal = 0){
  force(span); force(degree); force(se);
  force(predict); force(surface)
  force(draw.regions); force(r.color)
  grid.pal = colorRampPalette(grid.color)(n.colors ) ## (n)
  if(pal == 1) grid.pal = heat.colors(n.colors, alpha = 1)
  if(pal == 2) grid.pal = terrain.colors(n.colors, alpha = 1)
  if(pal == 3) grid.pal = topo.colors(n.colors, alpha = 1)
  if(pal == 4) grid.pal = cm.colors(n.colors, alpha = 1)
  if (res == 0.1){
    cex_pred = 0.1; data2pred = AmazonForestGrid
  } else if (res == 0.5){
    cex_pred = 1.25; data2pred = data_to_pred05
  } else {
    cex_pred = 2.5; data2pred = data_to_pred
  }
  z.loess = loess(z ~ Longitude * Latitude,
                 span = span, degree = degree, se = se,
                 normalize = TRUE, family = "gaussian",
                 surface = surface) #!surface is direct to be able to extrapolate
  #calculate explained variation
  SSq = sum((z-mean(z))^2)
  SSqres = sum((z - z.loess$fit)^2)
  expl_var = 100*(SSq-SSqres)/SSq
  #give output for loess regression model and expl variation
  cat("explained variation :",expl_var,"%","\n")
  #calculate the predicted values for the Amazon grid
  grid.z.predict = predict(z.loess, data2pred, se = T)
  #replace all fits < co by zero
  grid.z.predict$fit[grid.z.predict$fit < co] = 0
  if (blocks != T){
    plot(data2pred$Longitude,data2pred$Latitude,
         main = name,
         xlab = "Longitude", ylab = "Latitude",
         xlim = c(-80, -45), ylim = c(-20,10), asp = 30/30,
         xaxp = c(-80, -45, 7), yaxp = c(-20, 10, 6),
         pch = 22, cex = 3,
         col = rgb(0.85,0.95,0.85),
         bg = rgb(0.85,0.95,0.85))
  }
}

```

```

}
#show map of expected DCA scores and actual plot locations
if (blocks == T){
  grid.col = vector(length = length(data2pred$Longitude))
  grid.min = min(grid.z.predict$fit, na.rm = TRUE)
  grid.max = max(grid.z.predict$fit, na.rm = TRUE)
  grid.range = grid.max - grid.min
  grid.col = 1 -(grid.z.predict$fit - grid.min)/grid.range
  grid.col = grid.pal[1+round((n.colors-1)*(grid.z.predict$fit - grid.min)/grid.range)]
  plot(data2pred$Longitude,data2pred$Latitude,
        main = name,
        xlab = "Longitude", ylab = "Latitude",
        xlim = c(-80, -45), ylim = c(-20, 10), asp = 30/30,
        xaxp = c(-80, -45, 7), yaxp = c(-20, 10, 6),
        pch = 22, cex = cex_pred,
        col = grid.col, bg = grid.col)
}
if (dots == T){
  zmin = min(z)
  zmax = max(z)
  zrange = zmax - zmin
  zcex = 0.1 + round((4*(z - zmin)/zrange),1)
  points(Longitude, Latitude, cex = zcex, pch = 21, bg = "black", col = c.col)
}
if (draw.regions == T){
  add.geography(draw.forestborder = T, draw.regions = draw.regions, r.color = r.color)
} else {
  add.geography(draw.forestborder = T)
}
if (draw.legend == T){
  if (blocks == T){
    n.round = 0
    if (grid.max <=10) n.round = 1
    if (grid.max <= 1) n.round = 2
    legend.n = seq(grid.min, grid.max, by = grid.range/4)
    legend.pch = rep(2,length(legend.n))
    legend.col = grid.pal[1+round((n.colors-1)*(legend.n - grid.min)/grid.range)]
    legend.fill = legend.col
    legend(x = -49, y = -13,
           legend = round(legend.n, n.round),
           fill = legend.fill,
           bg = "white")
  }
  if (dots == T){
    legend.n = round(seq(zmin,zmax,zrange/4),0)
    legend.pch = rep(21,length(legend.n))
  }
}

```

```

legend.cex = 0.1 + round((4*(legend.n - zmin)/zrange),1)
legend(x = -48, y = 11,
      legend = legend.n,
      pch = legend.pch,
      pt.bg = "black",
      col = c.col,
      pt.cex = legend.cex,
      bg = "white")
  }
}
return(z.loess)
}
##### end map characteristics in loess map #####
##### map archaeological density#####
dens.arq<-rasterize(outp[,c(10,11)], amaz.r, fun='count', background=0)
dens.arq<-dens.arq*amaz.r
plot(dens.arq,col = terrain.colors(10), breaks = c(0,1,2,4,8,10,50,100,200))
##### end map archaeological density#####

##### generate balloon plot with default scaling #####
library(gplots)
balloonplot(dframe1$Region,dframe1$Variables, abs(tm),
            cum.margins=FALSE, rowmar=10.0, colmar=0.5, scale.range="absolute",
            zlab = "", xlab = "", ylab = "", label=FALSE, dotsize=4,
            dotcolor = c("blue", "red")[(c(tm5) < 0) + 1],
            show.margins=FALSE, sorted=FALSE, label.lines=FALSE,
            main=" ", cex.main=1)

##### non-linear relationships between abundance and richness #####
ytemp<-log((y - min(y)) + 0.1)
resu<-lm(ytemp~x)
summary(resu)
coef<-resu$coefficients
b<-as.vector(coef[2])
a<-as.vector(exp(coef[1]))
yo<-as.vector(min(y))
var<-c(yo,a,b);
names(var)<-c("yo", "a", "b")
# non-linear models
eq<-as.formula(y ~ yo + a*exp(b*x))
nlmod.2<-nls(eq, start=var,trace=TRUE)
resid<-residuals(nlmod.2)
# output
resu2<-summary(nlmod.2)
# summary
coef<-as.matrix(resu$coefficients)

```

```

df<-as.matrix(resu$df)
# statistics
dfregr<-df[1]-1
dfresid<-df[1]
dftotal<-length(y)-1
yest<-as.vector(fitted.values(nlmod.1)) # y estimado
ymed<-mean(y)
Rsqr<-sum((yest-ymed)^2)/sum((y-ymed)^2)
Rsqr<-round(Rsqr,2)
Radj<-1-(1-Rsqr)*(dftotal/dfresid)
Radj<-round(Radj,2)
F<-(sum((yest-ymed)^2)/dfregr)/(sum((y-yest)^2)/dfresid)
F<-round(F,2)
p<-df(F, dfregr, dfresid)
p<-round(p,2)
# plot
plot(y,x, pch=16, cex.lab=1.5, cex.axis=1.5, ylab="Relative abundance of all
domesticated spp. (%)", xlab = "Relative richness of domesticated spp. (%)")
xest<-seq(round(min(x),1),round(max(x),2),length=101)
lines(xest,predict(nlmod.1,list(x=xest)))

```