Using DNA to assess errors in tropical tree identifications: How often are ecologists wrong and when does it matter?

Ecological Monographs, 80(2), 2010, pp. 267–286

Using DNA to assess errors in tropical tree identifications: How often are ecologists wrong and when does it matter?

Kyle G. Dexter,[1,3] Terence D. Pennington,[2] and Clifford W. Cunningham[1]

[1]*Biology Department, University Program in Genetics and Genomics, Duke University, Box 90338, Durham, North Carolina 27708 USA*
[2]*Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AB United Kingdom*

Abstract. Ecological surveys of tropical tree communities have provided an important source of data to study the forces that generate and maintain tropical diversity. Accurate species identification is central to these studies. Incorrect lumping or splitting of species will distort results, which may in turn affect conclusions. Although ecologists often work with taxonomists, they likely make some identification errors. This is because most trees encountered in the field are not reproductive and must be identified using vegetative characters, while most species descriptions rely on fruit and flower characters. Because every tree has DNA, ecological surveys can incorporate molecular approaches to enhance accuracy. This study reports an extensive ecological and molecular survey of nearly 4000 trees belonging to 55 species in the tree genus *Inga* (Fabaceae). These trees were sampled in 25 community surveys in the southwestern Amazon. In a process of reciprocal illumination, trees were first identified to species using vegetative characters, and these identifications were revised using phylogenies derived from nuclear and chloroplast DNA sequences.

We next evaluated the effects of these revised species counts upon analyses often used to assess ecological neutral theory. The most common morphological identification errors involved incorrectly splitting rare morphological variants of common species and incorrectly lumping geographically segregated, morphologically similar species. Total error rates were significant (6.8–7.6% of all individuals) and had a measurable impact on ecological analyses. The revised identifications increased support for spatially autocorrelated, potentially neutral factors in determining community composition. Nevertheless, the general conclusions of community-level ecological analyses were robust to misidentifications. Ecological factors, such as soil composition, and potentially neutral factors, such as dispersal limitation, both play important roles in the assembly of *Inga* communities. In contrast, species-level analyses of neutrality with respect to habitat were strongly impacted by identification errors. Although this study found errors in morphological identifications, there was also strong evidence that a purely molecular approach to species identification, such as DNA barcoding, would be prone to substantial errors. The greatest accuracy in ecological surveys will be obtained through a synthesis of traditional, morphological and modern, molecular approaches.

Key words: biodiversity; DNA barcoding; genealogical species concept; Inga *(Fabaceae); Madre de Dios, southern Peru; neutral ecological theory; species delimitation; species identification; tropical trees.*

Introduction

In any field-based ecological study, an ecologist must identify individuals to species. This can be difficult in the species-rich tropics, with many undescribed species and often subtle, morphological distinctions between described species (see Plate 1). Furthermore, in studies of plants, only a subset of the characters used to describe a species is available for most individuals. Tropical plant taxonomists use both reproductive characters (i.e., the morphology of flowers, fruits, and their associated structures) and vegetative characters (i.e., the morphology of leaves, twigs, bark, and wood) to delimit species. Yet the paucity of reproductive individuals leads tropical woody plant ecologists to rely on vegetative characters alone for identification. This calls into question the accuracy of species identifications made by tropical ecologists. The corollary to this question is whether inaccurate species identifications are systematically biasing the results and conclusions of ecological studies. In this study, we used a procedure of reciprocal illumination between vegetative morphology and DNA sequence data to discover and correct mistakes in species identification and delimitation. Then we compare the results of ecological analyses using species identified with our method to results using species identified by vegetative characters alone.

The impact of potentially systematic errors in tropical tree species identification is important because much of

Manuscript received 15 February 2009; revised 28 July 2009; accepted 21 September 2009. Corresponding Editor: F. He.

[3] Present address: Laboratoire Evolution et Diversité Biologique, UMR 5174, CNRS/Université Paul Sabatier, Bâtiment 4R3, 31062, Toulouse, France.
E-mail: kgdexter@gmail.com

our current theory and understanding of ecology comes from tropical tree communities. One notable example is the neutral theory of biodiversity and biogeography, which was inspired by, and is still largely tested, using data from tropical tree communities (Hubbell 1979, 2001, Pitman et al. 2001, McGill 2003, Volkov et al. 2003). Other examples include the intermediate disturbance hypothesis (Connell 1978, Denslow 1987, Hubbell et al. 1999), ecological niche conservatism (Pitman et al. 1999, Webb 2000), and the role of density dependence in structuring communities (Janzen 1970, Connell 1971, Wills et al. 1997).

It is well established that ecologists (Goldstein 1997, Vecchione et al. 2000, Bortolus 2008), including tropical tree ecologists (Sheil 1995, Condit 1998), make errors in species identification. Determining how these errors affect the results and conclusions of ecological studies is more difficult. One must first determine when and how many identification errors have occurred. This can be accomplished by having a taxonomic expert review all of the ecologists' identifications (cf. Oliver and Beattie 1993, Derraik et al. 2002, Scott and Hallam 2002, Barratt et al. 2003) or through repeated surveys with different observers (Condit 1998, Archaux et al. 2006).

Alternatively, genetic data in the form of DNA sequences can be used to assess identification errors (Knowlton et al. 1992, Caesar et al. 2006, Bickford et al. 2007, Stuart and Fritz 2008). For example, the genealogical species concept (Baum and Shaw 1995) argues that evolutionarily coherent units can be identified when gene genealogies reveal monophyletic groups of individuals. In practice this means that individuals of a given species should be genetically more closely related to one another than to individuals of other species. If a sufficient number of individuals of the focal species are sequenced for a given gene, then morphological species can be compared to genealogical species to help assess identification accuracy. The large-scale sequencing efforts necessary for this type of study have previously been considered cost-prohibitive, but as sequencing costs drop, such studies have entered the realm of possibility (e.g., Hebert et al. 2003, 2004, Janzen et al. 2005, Lahaye et al. 2008).

Ecologists have long used genetic information to aid in the identification of individuals (Nanney 1982, Pace 1997, Brown et al. 1999). In fact, in studies of animal taxa, DNA sequence data alone have been used to delimit and identify species (termed DNA barcoding; Hebert et al. 2003, 2004, Janzen et al. 2005). However, this approach has been criticized because it can fail to delimit recently diverged species (Hickerson et al. 2006, Knowles and Carstens 2007). There are even fewer reasons to expect a purely genetic approach to be successful in plants.

First, most of the genetic markers commonly used for plants, including the ones used in this study, evolve much more slowly than the mitochondrial cytochrome oxidase I gene commonly used for barcoding animals (Kress et al. 2005, Newmaster et al. 2006). This increases the number of cases in which DNA sequences may not separate closely related species that clearly possess multiple, distinguishing morphological characters. Second, plants may engage in interspecific gene flow more often than animals (Chase et al. 2005, Cowan et al. 2006). Limited past hybridization can obscure patterns of species monophyly even when it has not affected species cohesion as determined by morphology (e.g., *Quercus*; Burger 1975). Nevertheless, when used in concert with morphological analyses, DNA sequence data have the potential to increase the accuracy of plant species delimitation and identification.

As part of a conventional ecological study, we surveyed communities of the tropical tree genus *Inga* (Mimosoideae: Fabaceae) at 25 sites across 30 000 km^2 of the lowland Amazonian Peru. We encountered nearly 4000 individual trees belonging to 63 putative *Inga* species. As in many ecological studies, individuals were first identified using vegetative morphological characters. These identifications were confirmed by careful consultation of herbarium specimens and with the aid of the recognized taxonomic authority in the genus *Inga* (co-author T. D. Pennington). Many tropical tree ecology studies incorporate advice from plant taxonomists (e.g., Pitman et al. 2001, Phillips et al. 2003, Tuomisto et al. 2003), and our morphological identification accuracy should be as good or better than that of these other studies.

We inferred phylogenies using nuclear and chloroplast DNA sequences for one-quarter of the surveyed individuals (946 total), encompassing all 63 putative species and multiple locations for widespread species. We use this phylogeny to propose hypotheses about possible mistakes in species delimitations and identifications. In cases of conflict between this phylogeny and the original species designations, we reexamined the original morphological material and identified three kinds of mistakes: (1) mistakes in individual identifications; (2) incorrect lumping of morphologically distinct species; and (3) incorrect splitting of a single species.

We evaluated the effect of species identification errors on several analyses that are conventionally used to test neutral ecological theory (cf. Hubbell 2001). We conducted analyses that compared the relative importance to community assembly of environmental factors, such as soil characteristics, and spatially autocorrelated, potentially neutral factors, such as dispersal limitation (Duivenvoorden et al. 2002, Legendre et al. 2005, Tuomisto and Ruokolainen 2006). Furthermore, we analyzed distance decay in the compositional similarity of communities (Condit et al. 2002, Morlon et al. 2008). Finally, we conducted species-level analyses of neutrality with respect to habitat preference (Phillips et al. 2003, John et al. 2007).

Materials and Methods

Field sites and survey methods

This study focused on distinguishing species within the genus *Inga* (Mimosoideae: Fabaceae) (Pennington 1997). This is the most diverse and abundant tree genus within the study area of Madre de Dios, in southern Peru (N. Pitman, *unpublished data*; K. Dexter, *personal observation*). We surveyed *Inga* communities in the two principal habitat types found in the region: terra firme (upland) and floodplain (bottomland) forest.

In community surveys, we first censused all *Inga* individuals that reached breast height (1.3 m) in a 50 × 50 m plot. If there were fewer than 80 individuals in the plot, we sampled additional individuals in 2 m wide transects until that number was reached (again including all *Inga* individuals that reached breast height). Transects were run in straight lines from the plot and restricted to the same habitat as the plot (mature floodplain or terra firme forest). In some locations (Cocha Cashu, Los Amigos, Las Piedras, Tambopata), we surveyed *Inga* individuals in additional 25 × 25 m plots. All transects and plots within a given community survey were restricted to a 2 × 2 km or smaller area. Within plots or transects, all *Inga* individuals were measured (for diameter at breast height and absolute height), identified, and collected for a genetic specimen and, in most cases, an herbarium voucher. When walking trails at study sites, we collected additional *Inga* individuals if they were morphologically unusual and also to boost the sample size of individual species. These additional individuals were not included in community survey totals as they were not randomly sampled.

For each community survey, we obtained a soil sample from the 50 × 50 m plot by bulking soil cores taken at five random points throughout the plot. Soil samples were also collected from the additional 25 × 25 m plots installed at some sites. All soil samples were sent to the Clemson University Agricultural Services Laboratory (Clemson, South Carolina, USA) for analyses. Soil pH was measured using an AS-3000 Dual pH Analyser (LabFit, Perth, Australia). The pH of a buffer solution was also measured to quantify total acidic cations (stored acidity; in milliequivalents per 100 g of soil). Extractable cations (B, Ca, Cu, K, Mn, Mg, Na, Zn) and phosphorous were quantified (in ppm) using the Mehlich 1 procedure. Nitrate nitrogen (NO_3^-) was quantified (in ppm) by cadmium reduction using a Flow Injection FIALab 2500 instrument (FIAlab, Bellevue, Washington, USA). Cation exchange capacity (CEC; in milliequivalents per 100 g soil) was calculated as the sum of the total acidic cations and base cations (Ca, K, Mg, and Na; first converted to milliequivalents per 100 g soil). The base saturation percentage (BS) was also calculated (in total and separately for K, CA, Mg, and Na). Furthermore, samples were sent to the North Carolina State Soil Services Laboratory, Raleigh, North Carolina, USA, for analysis of particle size distribution (percentage of mass of sand, silt, and clay).

Initial morphological species identification

All surveyed individuals were initially identified to morphospecies based on vegetative characters. *Inga* species are not highly variable in trunk and bark characters, so we relied largely on leaf characters to delimit species. This has the added advantage that collected vouchers can later be compared with one another and with identified herbarium specimens. *Inga* species vary greatly in the number and size of leaflets (all have compound leaves), presence and size of stipules, presence and nature of pubescence, secondary and tertiary venation, the presence and form of wings on the rachis, as well as many other characters, all of which were used to identify individuals. After vegetative morphospecies were delimited, we reviewed specimens from southern Peru in various herbaria to determine their taxonomic identity. Representative vouchers of all species, including unidentified morphospecies, have been deposited at Kew Botanic Gardens (K), Duke University Herbarium (DUKE), and the herbarium of the forestry department of La Universidad Agraria La Molina in Lima, Peru (MOL).

DNA sequences

We selected three sets of individuals for verification, via DNA sequencing, of the initial morphological identifications. First, we selected ~100 individuals at random from community surveys in floodplain and terra firme at our two principal field sites, Cocha Cashu and Los Amigos (a total of 442 randomly selected individuals). Second, we sequenced at least two individuals of all putative species that were not covered by the above sequencing. Third, for most widespread species (*I. alata*, *alba*, *auristellae*, *bourgonii*, *cayennensis*, *chartacea*, *cinnamomea*, *edulis*, *marginata*, *poeppigiana*, *ruiziana*, *sapindoides*, *thibaudiana*, and *umbellifera*), we sequenced individuals from additional sites that spanned the breadth of their distribution within the study area. We also collected and sequenced individuals of the genus *Zygia* (Mimosoideae: Fabaceae), the likely sister genus to *Inga* (Jobson and Luckow 2007), for use as an outgroup in phylogenetic analyses.

We used a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle 1990) to extract DNA from silica-gel-dried leaf specimens from all selected individuals. We used polymerase chain reaction (PCR) to amplify one nuclear region, the internal transcribed spacers (ITS 1 and 2) and 5.8S gene of the nuclear ribosomal DNA, and one chloroplast region, the trnD-trnT intergenic spacer (which spans multiple intergenic spacers). We selected the ITS region based on its previously demonstrated variability in the genus (Richardson et al. 2001; K. G. Dexter, *unpublished data*) and used the following primers for amplification: ITS 4 (White et al. 1990) and ITS 5a (Stanford et al.

2000). We selected the trnD-trnT region because it contained the highest number of polymorphic sites among multiple chloroplast intergenic regions that were initially screened in the genus (Coley et al. 2005; R. T. Pennington, *unpublished data*). We amplified the trnD-T region using the following primers: trnH (trnT in Demesure et al. 1995) and trng2 (Oh and Potter 2003). Reaction conditions were as follows: one cycle of 94°C for 2 min; 40 cycles of 94°C for 30 s, 52°C for 1 min (55°C here for trnD-T region), and 72°C for 2 min; 1 cycle of 72°C for 7 min. The 25-μL reaction mix consisted of 12.3 μL H_2O, 5 μL Q reagent (Qiagen, Valencia, California, USA), 2.5 μL *Taq* Buffer, 0.5 μL dNTP mix (10 mmol/L concentration for each nucleotide), 1.25 μL primer 1 (10 μmol concentration), 1.25 μL primer 2 (10 μmol concentration), 1 μL $MgCl_2$, 0.2 μL *Taq* polymerase, and 1 μL of DNA template.

Cleaned PCR products were sequenced, using the amplification primers, on an ABI 3730 XL capillary sequencer (Applied Biosystems, Foster City, California, USA). Sequences were assembled using Sequencher version 4.5 (Gene Codes, Ann Arbor, Michigan, USA) and manually aligned in MacClade version 4.06 (Maddison and Maddison 2003).

Phylogenetic analyses

For the ITS and trnD-T regions, Modeltest version 3.7 (Posada and Crandall 1998) found the same model of sequence evolution using the Akaike information criterion (AIC) (general time reversible model with a proportion of sites invariant and a gamma distribution for variable sites, GTR + I + G). Preliminary phylogenetic analyses indicated few strongly supported topological differences between the two regions. Therefore, we concatenated the two regions for additional phylogenetic analyses. Once concatenated, we used COLLAPSE version 1.2 (Posada 2006) to reduce the data set to unique sequences. Thus, individuals with identical ITS and trnD-T sequences were represented by only one placeholder in phylogenetic analyses. If individuals of two or more species were determined to have the same unique sequence, we retained an individual sequence from each species to aid in assessing the monophyly or lack thereof of putative species.

We conducted a maximum-likelihood analysis in Garli version 0.96 (Zwickl 2006) with 100 random addition sequence replicates. We used the best-fit GTR + G + I model, allowing Garli to estimate all parameters while searching for the optimal phylogeny. We also conducted a maximum-likelihood bootstrap analysis with 1000 bootstrap replicates with Garli version 0.96, importing bootstrap trees into PAUP* 4.0b10 (Swofford 2002) to produce 50% majority rule consensus trees.

We also conducted a Bayesian phylogenetic analysis of the concatenated data set, using Mr. Bayes version 3.1.2 (Ronquist and Huelsenbeck 2003). We used the GTR + I + G model, but we unlinked the partitions so that parameter values and overall rate of substitution could differ for the two genetic regions. We ran two independent runs of four Markov chains for 20 million generations with a heating scheme, selected through preliminary analyses, to minimize the average standard deviation of split frequencies (Temp = 0.15, Swapfreq = 1, Nswaps = 1). Trees were sampled every 1000 generations. The average standard deviation of split frequencies reached 2% after 11 million generations and fluctuated around this value thereafter. We discarded trees from the first 11 million generations of each run as the burn-in. As before, we used PAUP beta version 10.4 (Swofford 2002) to produce 50% majority rule consensus trees reflecting posterior probabilities for each node.

In addition to the aforementioned preliminary analyses for each marker, we conducted full phylogenetic analyses separately for each partition, ITS and trnD-T, of the concatenated data set.

Assessing identification errors and revising species delimitations

We used a two-step process of reciprocal illumination to assess errors in the initial morphology-based delimitations and identifications. Potential errors were first identified through examination of the generated molecular phylogenies. We then reviewed the morphology of the species involved as well as relevant herbarium material to confirm whether an error had been made. We classified errors into three categories.

1. *Mistakes in individual identifications.*—In considering thousands of individuals in sometimes difficult field conditions, some individual trees may be mistakenly identified. These instances were revealed when an accession of a given species was placed phylogenetically within another species. If the ambiguous individual better matched the morphology of the new phylogenetic placement, this was considered a simple identification mistake due to human error and not due to the vagaries of species delimitation.

2. *Incorrect lumping of distinct species.*—We may also have made errors in species delimitation. For example, if a putative species falls out as two or more divergent groups in the phylogeny (as opposed to just one individual being divergent), this could indicate that multiple species were incorrectly lumped together as one species. In these cases, we reexamined the morphology of vouchers to determine whether there were any consistently segregating morphological characters between the phylogenetically divergent groups. If any such characters were found, the distinguishing characters were noted, and we considered the two groups to be two separate species.

3. *Incorrect splitting of a single species.*—If two or more putative species were mixed together within a monophyletic group, this could indicate that they actually comprise only one species. If close reexamination of vouchers found no consistently segregating morphological differences between the putative species, we lumped the original species together as one species. If

the species did possess segregating, morphological characters, we continued to treat them as separate species, despite the inability to distinguish them genetically.

Evolutionarily significant units.—When the phylogeny placed an unidentified morphospecies as sister to a known species in the phylogeny, we reexamined the morphology of these individuals and also herbarium records of the known species across its entire range (not just in southern Peru as was done in the first pass). This review may indicate that the unidentified morphospecies falls within the morphological limits of the known species. However, reciprocal monophyly of our samples of the two putative species indicates that they may be evolving independently. In these cases, we labeled the originally delimited species as two different evolutionarily significant units (ESUs; Moritz 1994) of the same species. When we tabulated error rates, we did so twice, both lumping ESUs as single species and treating them as distinct species.

Ecological analyses

We conducted the following ecological analyses using three sets of species delimitations: (1) the original delimitations based on vegetative morphology; (2) delimitations revised using the reciprocal illumination procedure and treating ESUs as distinct species; and (3) revised delimitations with lumping ESUs of a given species together.

Partitioning variation in community composition.—To assess the relative importance of environmental factors vs. spatially autocorrelated, potentially neutral factors in determining the species composition of communities, we used redundancy analysis within a variance partitioning framework (Legendre et al. 2005, Jones et al. 2008). Specifically, we partitioned the variation in a species-by-site matrix among four additive components: (1) a pure environmental component; (2) a pure spatial component; (3) a combined spatial–environmental component (due to the correlated effects of environmental and spatial factors); and (4) an unexplained component.

We used two different types of species-by-site matrices in variance partitioning analyses. The first was a presence/absence matrix in which species were represented by a 1 if present in a community survey and a 0 if absent. The second consisted of the relative abundance of each species in each community survey.

The environmental component was represented by the natural logarithm of the 21 measured soil variables for each community survey site. We used the mean value of soil variables for sites from which more than one soil sample was collected. As the number of explanatory variables compared to the number of sites was already high, we did not attempt to examine polynomial or higher order combinations of environmental variables.

The variables used to represent spatial autocorrelation were principal components of neighbor matrices (PCNMs), generated using the program SpaceMaker 2.0 (Borcard and Legendre 2004). The PCNMs represent spatial structure at multiple spatial scales and were obtained by a principal components analysis of a truncated geographic distance matrix of pairwise distances between survey sites. The truncation distance represents the minimum scale at which spatial structure can be detected in the data. Following the recommendations of Borcard and Legendre (2004), we set the truncation distance at the minimum distance needed to connect all survey sites within a single network (82 km).

We used a forward selection approach (Blanchet et al. 2008), separately for the environmental and spatial variables, to determine which variables contributed significantly to variation in community composition. Only those variables that were significant (at $P < 0.05$ with 9999 permutations) were included in variance partitioning analyses. Forward selection was conducted separately for each variance partitioning analysis. We used the function forward.sel in the packfor package (Dray 2007) in the R statistical environment (R Development Core Team 2008) to conduct forward selection.

We used the function varpart in the R package vegan (Oksanen et al. 2008) to conduct variance partitioning analyses. We adjusted R^2 values for the number of sample sites and explanatory variables (Peres-Neto et al. 2006) to give an unbiased estimate of the proportion of variation explained by each component. We evaluated the significance of the pure environmental and pure spatial fractions using the functions rda and anova.cca in the vegan package with 9999 permutations (Oksanen et al. 2008).

Distance decay in community similarity.—The importance of spatially autocorrelated factors, such as dispersal, in structuring communities has also been frequently examined through analyses of distance decay, the decline in similarity in species composition of communities with geographic distance (Condit et al. 2002, Morlon et al. 2008). We used partial Mantel tests (cf. Phillips et al. 2003, Tuomisto et al. 2003, Tuomisto and Ruokolainen 2006) to evaluate the relationship between community similarity and geographic distance, while controlling for environmental variation. Specifically, we assessed the correlation between a pairwise community similarity matrix (actually evaluated as dissimilarity matrix) and a geographic distance matrix, while including a matrix representing environmental distance between communities as a covariate. While partial Mantel tests may not be appropriate for partitioning variation in community composition (Legendre et al. 2005, 2008), they can be useful in determining whether there is significant distance decay. We used the function mantel.partial in the R package vegan (Oksanen et al. 2008).

We analyzed distance decay using both the Jaccard and Bray-Curtis indices of community similarity. The indices were calculated using the program EstimateS version 8.00 (Colwell 2006). The Jaccard index uses

presence/absence information, while the Bray-Curtis index additionally incorporates abundance information. Both of these indices vary from 0 to 1, with 1 representing maximal community similarity. Partial Mantel tests require that all matrices be phrased in terms of distances between communities, so we converted similarity values to dissimilarity values by taking 1 minus a given similarity index. We constructed two different matrices of geographic distance between communities, one with straight-line geographic distance and one with the natural logarithm of geographic distances.

We used data from collected soil samples to measure the environmental distance between communities. We first conducted a principal component analysis (PCA) on all measured soil variables. We then calculated the Euclidean distance between communities for the first three principal component axes individually and for their two- and three-dimensional combinations (these are the only axes that explained >10% of variation in the soil data). Separate environmental distance matrices were constructed using each measure of Euclidean distance. For a given community similarity and geographic distance matrix, we conducted partial Mantel tests multiple times, using each possible environmental distance matrix as a covariate. This allowed us to control for any possible significant environmental variation in evaluating the distance–decay relationship.

The intercept and slope of the relationship between community similarity and geographic distance are often interpreted to reflect the level of dispersal limitation in the landscape (Nekola and White 1999, Chave and Leigh 2002, Morlon et al. 2008). We estimated these parameters using simple linear models relating community similarity to geographic distance (using both straight-line distances between communities and their logarithm).

We then assessed whether the distance-decay parameter estimates for the original and revised delimitations differed more than expected by chance. We compared the observed difference in parameter estimates between the two delimitations to that between two null sets of community similarity values. In the first set, we randomly selected whether each community pair was represented by the original or revised similarity value, while the alternate value was assigned to the same pair in the second set. Across both sets, we preserved the original geographic distances between each pair of communities. This serves to maintain the distance decay inherent in the data, while assessing how different the distance-decay parameters can be by chance. The proportion of 999 null replicates with a difference in parameter estimates greater than that observed in the real data gives a P value for this one-tailed test.

Species-level ecological analyses.—We used a modified version of the approach of Phillips et al. (2003) to determine whether individual species are neutral with respect to habitat or specialize on terra firme or floodplain. We first calculated θ, the relative abundance of a given species across the entire data set. This value is also the species' expected relative abundance in floodplain and terra firme if the species is neutral. For species that we suspected to be floodplain specialists (i.e., that had greater relative abundance in floodplain), we calculated the binomial probability, $\text{Bin}(Y_{\text{FP}} \mid N_{\text{FP}}, \theta)$, that we found Y_{FP} individuals of the species in floodplain habitat given N_{FP}, the total number of individuals sampled in floodplain habitat, and θ. This is equivalent to assaying whether the observed relative abundance of the species in floodplain is significantly different from that expected by chance or under neutrality, given our sampling effort. We performed a similar test for suspected terra firme specialists using $\text{Bin}(Y_{\text{TF}} \mid N_{\text{TF}}, \theta)$. If the P value for a given species for the selected binomial test was greater than 0.05, then the species was classified as neutral. Note that these one-tailed tests are based on first predicting the direction of a species habitat specialization.

If a species is only represented by a few individuals, then statistical power will be lacking to falsify the hypothesis that the species is neutral with respect to habitat specificity. We determined the minimum number of individuals needed to detect whether a hypothetical species is a habitat specialist if all sampled individuals were found within a single habitat type. We did not perform binomial tests on species with sample sizes below this threshold of detectability.

Results

Field sites and census results

We surveyed communities at 14 locations across Madre de Dios (Fig. 1), separated by a range of spatial distances (from 3 to 250 km apart). At each location, we attempted to survey both terra firme and floodplain forest. Floodplain forests can vary from swamp to successional to mature forests (Kalliola et al. 1991, Pitman et al. 1999); we attempted to survey communities only in mature floodplain forest, as judged by forest stature and the absence of early primary successional tree species (e.g., *Cecropia membrenacea*, *Ficus insipida*, *Cedrela odorata*). At three locations (Otorongo, Blanquillo, and Camungo), we could not access terra firme forest. There were thus 25 total community surveys, 14 in floodplain forest and 11 in terra firme forest. (See Supplement for species composition data, geographic coordinates, and soil data for each community survey.)

In several community surveys (floodplain at Blanquillo, CM2, and Maizal; terra firme at Boca Manu), the total number of sampled individuals was less than 80, while in other surveys, the sample size far exceeded 80 (131 ± 13 individuals [mean ± SE]; range, 46–282 individuals). Excluding low- or high-sample-size communities from our community-level analyses had little effect on the results presented here.

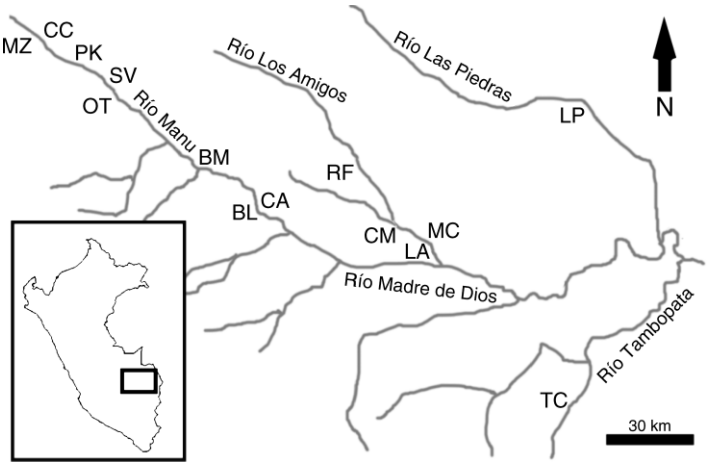


FIG. 1. Map of the study area in Madre de Dios, Peru. Locations of all sites where tree community surveys were conducted are given: MZ, Maizal; CC, Cocha Cashu Biological Station; PK, Pakitza; SV, Salvador; OT, Otorongo; BM, Boca Manu; BL, Blanquillo; CA, Camungo; RF, Refugio; CM, Centro de Monitoreo 2; MC, Centro de Monitoreo 3; LA, Los Amigos Research Center; LP, Las Piedras Biodiversity Station; TC, Tambopata Research Center.

Initial morphological species identification

Our initial morphological delimitations revealed 63 putative species. Our initial examination of herbarium specimens from southern Peru allowed us to assign taxonomic names to 43 species, while 20 species remained as unidentified morphospecies. Our procedure of reciprocal illumination for revising these initial morphological identifications is described in the following sections.

DNA sequences.—In total, we obtained DNA sequences for 946 *Inga* individuals (651 for the ITS nuclear region and 892 for the trnD-T chloroplast region) from all 63 putative species that were encountered during the course of this ecological study (see Supplement for GenBank accession numbers; the master alignment was deposited in TreeBase). This represents 24.2% of the 3912 surveyed individuals. The ITS region varied in length from 638 to 668 base pairs (bp), and the total aligned data set was 671 bp in length. We found no evidence that any of the ITS sequences represented pseudogenes (Alvarez and Wendel 2003). The trnD-T region varied in length from 1060 to 1100 bp once the ambiguous ends of sequences were trimmed, and the total aligned data set was 1167 bp in length. Alignment of all sequences was unambiguous.

Phylogenetic analyses.—The concatenated data set contained 191 unique sequences, including seven sequences found in samples of the outgroup *Zygia*. Including individuals that had identical ITS and trnD-T sequences but represented putatively distinct species boosted the data set for phylogenetic analysis to a total of 222 concatenated sequences.

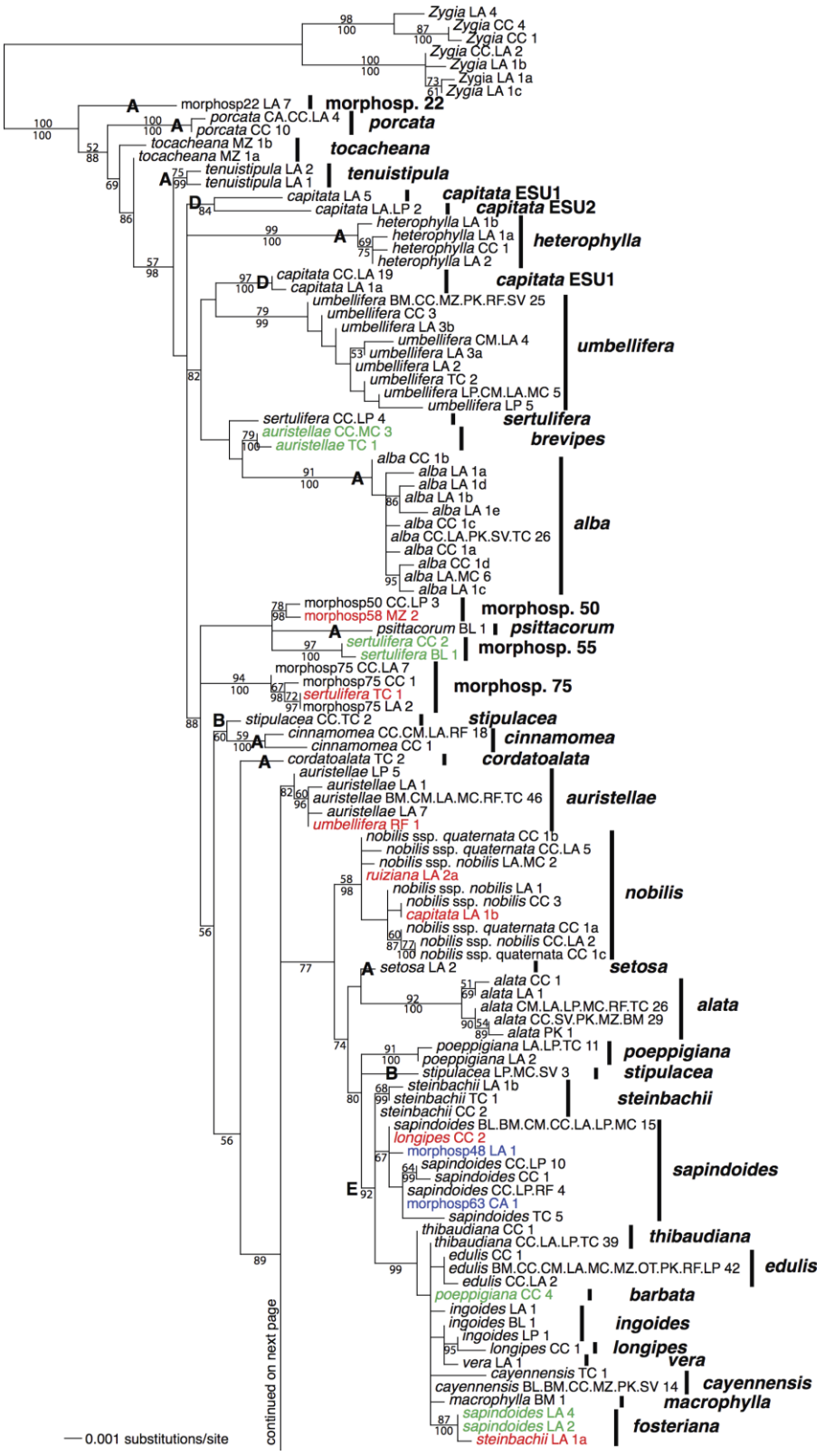
The maximum-likelihood tree for the concatenated data set is given in Fig. 2. The Bayesian phylogenetic analysis (not shown) gave a highly similar topology. The percentage of 1000 maximum-likelihood bootstrap replicates that supported each node (>50%) and the Bayesian posterior probabilities for each node (>0.5) are shown in Fig. 2.

The maximum-likelihood trees for the single-locus phylogenetic analyses, along with maximum-likelihood bootstrap support values and Bayesian posterior probabilities, are given in Appendices A and B.

Assessing identification errors and revising species delimitations.—Many species that were delimited based on vegetative morphology formed reciprocally monophyletic groups in the phylogeny (*Inga alba*, *cinnamomea*, *cordatoalata*, *heterophylla*, *porcata*, *psittacorum*, *setosa*, *suaveolans*, *tenuistipula*, and morphospecies 17, 22, and 54). These species did not require further assessment of morphology and were considered well-delimited species (marked with letter A in Fig. 2).

Mistakes in individual identifications.—Various species (*I. alata*, *chartacea*, *longipes*, *ruiziana*, *sapindoides*, *sertulifera*, *steinbachii*, *umbellifera*, and morphosp. 58) contained 1–3 sequenced individuals that were divergent from the majority of individuals sequenced for their species and that were nested within other species. In all of these cases, a review of vouchers showed that the divergent individuals had been identified incorrectly (highlighted in red in Fig. 2). Once these errors were corrected, additional species demonstrated reciprocal monophyly in the phylogeny (*I. alata*, *marginata*, *nobilis*, *ruiziana*, *umbellifera*, and morphospecies 50, 58, 71, and 75).

Incorrect lumping of distinct species.—Multiple species were found to comprise two divergent groups in the phylogeny. In the cases of *Inga auristellae*, *poeppigiana*, *sapindoides*, and *sertulifera*, our detailed reanalysis of voucher specimens uncovered morphological characters that distinguished the divergent groups (see Table 1). In each of these cases, one group corresponded better to herbarium specimens of the originally designated species. For the first three species, an extensive review of herbarium specimens demonstrated that the other group actually matched lesser-known, rarely collected species (*I. brevipes*, *barbata*, and *fosteriana*, respectively). In the case of *I. sertulifera*, the other group could not be matched up with any known species and was given the name *I.* morphospecies 55. These four cases represent


Zygia LA 4
Zygia CC 4
Zygia CC 1
Zygia CC.LA 2
Zygia LA 1b
Zygia LA 1a
Zygia LA 1c
morphosp22 LA 7
morphosp. 22
porcata CA.CC.LA 4
porcata CC 10
porcata
tocacheana MZ 1b
tocacheana MZ 1a
tocacheana
tenuistipula LA 2
tenuistipula LA 1
tenuistipula
capitata LA 5
capitata ESU1
capitata LA.LP 2
capitata ESU2
heterophylla LA 1b
heterophylla LA 1a
heterophylla CC 1
heterophylla LA 2
heterophylla
capitata CC.LA 19
capitata LA 1a
capitata ESU1
umbellifera BM.CC.MZ.PK.RF.SV 25
umbellifera CC 3
umbellifera LA 3b
umbellifera CM.LA 4
umbellifera LA 3a
umbellifera LA 2
umbellifera TC 2
umbellifera LP.CM.LA.MC 5
umbellifera LP 5
umbellifera
sertulifera CC.LP 4
auristellae CC.MC 3
auristellae TC 1
sertulifera
brevipes
alba CC 1b
alba LA 1a
alba LA 1d
alba LA 1b
alba LA 1e
alba CC 1c
alba CC.LA.PK.SV.TC 26
alba CC 1a
alba CC 1d
alba LA.MC 6
alba LA 1c
alba
morphosp50 CC.LP 3
morphosp58 MZ 2
morphosp. 50
psittacorum BL 1
psittacorum
sertulifera CC 2
sertulifera BL 1
morphosp. 55
morphosp75 CC.LA 7
morphosp75 CC 1
sertulifera TC 1
morphosp75 LA 2
morphosp. 75
stipulacea CC.TC 2
stipulacea
cinnamomea CC.CM.LA.RF 18
cinnamomea CC 1
cinnamomea
cordatoalata TC 2
cordatoalata
auristellae LP 5
auristellae LA 1
auristellae BM.CM.LA.MC.RF.TC 46
auristellae LA 7
umbellifera RF 1
auristellae
nobilis ssp. quaternata CC 1b
nobilis ssp. quaternata CC.LA 5
nobilis ssp. nobilis LA.MC 2
ruiziana LA 2a
nobilis ssp. nobilis LA 1
nobilis ssp. nobilis CC 3
capitata LA 1b
nobilis ssp. quaternata CC 1a
nobilis ssp. nobilis CC.LA 2
nobilis ssp. quaternata CC 1c
nobilis
setosa LA 2
setosa
alata CC 1
alata LA 1
alata CM.LA.LP.MC.RF.TC 26
alata CC.SV.PK.MZ.BM 29
alata PK 1
alata
poeppigiana LA.LP.TC 11
poeppigiana LA 2
poeppigiana
stipulacea LP.MC.SV 3
stipulacea
steinbachii LA 1b
steinbachii TC 1
steinbachii
steinbachii CC 2
sapindoides BL.BM.CM.CC.LA.LP.MC 15
longipes CC 2
morphosp48 LA 1
sapindoides CC.LP 10
sapindoides CC 1
sapindoides CC.LP.RF 4
morphosp63 CA 1
sapindoides TC 5
sapindoides
thibaudiana CC 1
thibaudiana CC.LA.LP.TC 39
thibaudiana
edulis CC 1
edulis BM.CC.CM.LA.MC.MZ.OT.PK.RF.LP 42
edulis CC.LA 2
edulis
poeppigiana CC 4
barbata
ingoides LA 1
ingoides BL 1
ingoides LP 1
ingoides
longipes CC 1
longipes
vera LA 1
vera
cayennensis TC 1
cayennensis BL.BM.CC.MZ.PK.SV 14
cayennensis
macrophylla BM 1
macrophylla
sapindoides LA 4
sapindoides LA 2
steinbachii LA 1a
fosteriana
continued on next page
0.001 substitutions/site

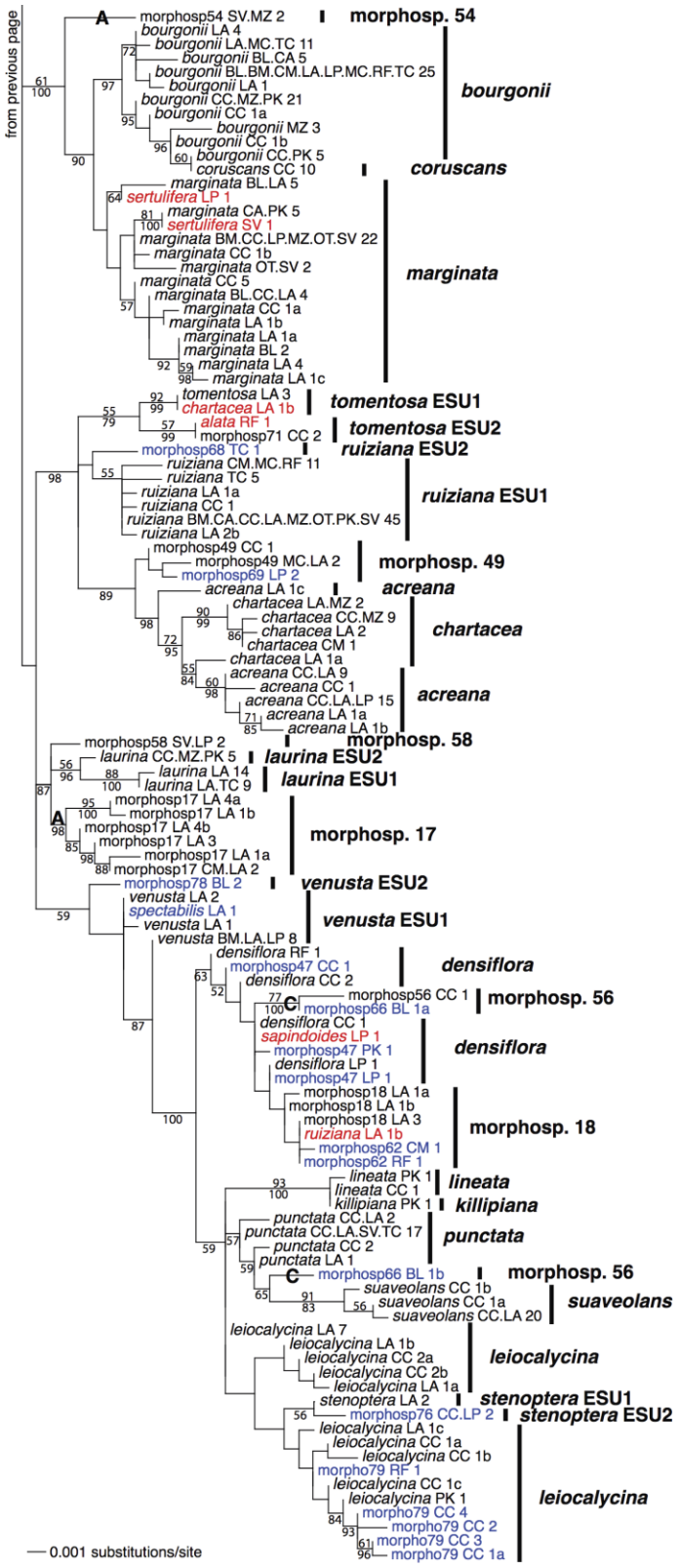


FIG. 2. Maximum-likelihood tree for *Inga* samples from Madre de Dios for the concatenated data set of nuclear internal transcribed spacer (ITS) and chloroplast trnD-T intergenic spacer sequences. The percentage of 1000 maximum-likelihood bootstrap replicates that support a given node is given above the branch preceding a node (only given if >50). The posterior probabilities for nodes from a Bayesian phylogenetic analysis are given below the branches (×100; only given if >0.5). The taxon labels are followed by two-letter codes that give the locations at which a given sequence was found (see Fig. 1) and the total number of individuals in the molecular sequence data set with that sequence. The last lowercase letter provides a unique identifier for alleles where necessary for comparison with Appendices A and B. The finalized species identities are given on the right-hand side of the tree. The three different categories of error are color-coded: red indicates mistakes in individual identification; green indicates incorrectly lumped species; blue indicates incorrectly split species. Large boldface letters are referred to in the text (A, species that were reciprocally monophyletic under the original delimitations; non-monophyletic species pending more information: B, *Inga stipulacea*; C, *I.* morphospecies [morphosp.] 56; D, *I. capitata*; E, a speciose clade with little genetic differentiation between species).

TABLE 1. Morphological characters that were used to delimit incorrectly lumped *Inga* species as well as characters that can be used to distinguish the newly segregated species or evolutionarily significant unit (ESU).

Original species	Segregated species	Characters used to define original species	Characters used to define segregated species
I. auristellae	*I. brevipes*	2–3 pairs leaflets proximal leaflets basal (short petiole) winged rachis flares distally short, orange pubescence on rachis narrow stipules	stipules persistent larger leaflets (3–7 × 6–15 cm vs. 2–5 × 5–10 cm) pubescence also on mid-rib of leaflets
I. poeppigiana	*I. barbata*	3–4 pairs leaflets relatively small leaflets (<13 cm long) winged rachis long hispid pubescence (>1.5 mm)	broader, oblanceolate leaflets shorter, narrower stipule (1 mm long, <0.5 mm wide) brochidodromous venation
I. sapindoides	*I. fosteriana*	≥3 pairs leaflets large leaflets (often >25 cm in length) winged rachis orange-red pubescence regular cup-shaped extra-floral nectary	denser, more tomentose pubescence usually ≥4 pairs leaflets larger, spadiform stipule (1.5–2 × 1–1.5 cm)
I. sertulifera	*I.* morphospecies 55	2 pairs leaflets (6–10 × 4–7 cm) rounded, elliptical leaflets glabrous narrow stipules	narrow, apressed wing on rachis reticulate tertiary venation
I. capitata	*I. capitata* ESU2	2–3 pairs coriaceous leaflets glabrous reticulate tertiary venation persistent stipules	narrower stipule (<3 mm wide) smaller, more elliptic leaflets (3–7 × 7–15 cm vs. 4–10 × 10–25 cm)
I. laurina	*I. laurina* ESU2	2 pairs leaflets (10–16 × 6–12 cm) rounded, elliptical leaflets glabrous sparse, tertiary venation	persistent, short stipules (<1 mm)

Note: The study was conducted in Madre de Dios, in southern Peru.

instances of incorrectly lumped species (highlighted in green in Fig. 2).

Inga stipulacea, morphospecies 56, and *capitata* were also polyphyletic in the phylogeny (B–D, respectively, in Fig. 2). Regarding *I. stipulacea*, the sampled individuals strongly resemble one another and are morphologically very distinct from any other species. The sampled individuals do in fact share a chloroplast allele, but comprise two divergent clades for the ITS locus. There is no geographic or morphological segregation of these two clades, and we are uncertain of the cause of their non-monophyly with regards to ITS. We therefore leave the species designation as is, and future research with further nuclear markers may in fact reveal that the species does form a cohesive monophyletic clade. Regarding *I.* morphospecies 56, it is difficult to say much. The species was only found three times, and in fact originally comprised two species (*I.* morphospecies 66 was lumped with this species; see *Materials and methods: Incorrect splitting of single species*). All three individuals share the same chloroplast allele, but one individual is divergent for ITS. For now, our conclusions regarding this species' status are very tentative, and further sampling is needed.

Inga capitata formed three groups in the phylogeny. All three groups are divergent from one another for the ITS marker, while one is divergent from the other two for the trnD-T marker (Appendices A and B). This latter group (*capitata* LA.LP_2) is also distinct morphologically (see Table 1). Because *capitata* LA.LP_2 is morphologically and genetically distinct from the others, it has been designated as a distinct evolutionarily significant unit (ESU 2). This leaves the other morphotype polyphyletic pending further information (ESU 1).

Inga laurina was found to comprise two well-supported sister groups in the phylogeny that, upon reexamination of vouchers, were found to differ slightly morphologically (see Table 1). We therefore split *I. laurina* into two separate ESUs. While other species also comprised sister groups in the phylogeny (e.g., *I. alata*), these groups were not strongly supported or distinguishable morphologically.

Incorrect splitting of single species.—Many species were paraphyletic or otherwise phylogenetically intermixed with other species (including multiple cases in which species shared alleles). In the cases of *Inga* morphospecies 18, 49, 56, *sapindoides*, *leiocalycina*, and *densiflora*, a broader review of herbarium specimens showed that other, originally delimited species did not possess sufficient segregating morphological characters to be distinguished as separate species. These represent cases in which we incorrectly split a single species into multiple species, and we corrected this by lumping the species together (highlighted in blue in Fig. 2). In most of the cases above, the newly defined species form monophyletic clades. However, in the latter two cases (*I. leiocalycina* and *I. densiflora*), the newly delimited species form a paraphyletic grade with respect to other

TABLE 2. Frequency of different identification and delimitation errors, as assessed through phylogenetic analyses, across an ecological study of trees in Amazonian Peru.

Category	With ESUs lumped			With ESUs treated as species		
	No. species	Total no. sequenced individuals	Total no. individuals in data set	No. species	Total no. sequenced individuals	Total no. individuals in data set
Mistakes in individual ID	10 (15.9%)	16 (1.7%)	NA†	10 (15.6%)	16 (1.7%)	NA†
Incorrectly lumped	4 (6.3%)	17 (1.8%)	77 (2.0%)	6 (9.4%)	24 (2.6%)	145 (3.7%)
Incorrectly split	12 (19.0%)	31 (3.3%)	83 (2.1%)	9 (14.1%)	32 (3.4%)	126 (3.2%)
Total errors	24 (38.1%)	64 (6.8%)	NA†	24 (37.5%)	72 (7.6%)	NA†
Total	63	946	3912	64	946	3912

Notes: Values are the number of species or individuals with each type of error. Evolutionarily significant units (ESUs) of a given species either were lumped together as one species or were treated as separate species-level entities.

† It is not possible to extrapolate the number of mistakes in individual identifications to the entire data set, and thus the total error rate cannot be calculated for the entire data set.

species, from which they are distinguished by multiple morphological characters. *Inga spectabilis* was nested phylogenetically within a paraphyletic *I. venusta*, and based on similar morphology of our vouchers, we lumped this species with *I. venusta*. As the putative *I. spectabilis* was represented by only one vegetative accession, it may be premature to take this as signifying that *I. spectabilis* is not a good species.

Inga nobilis is the only species that we originally delimited to the level of subspecies. Subspecies are conceptually similar to ESUs (they should be genetically and morphologically distinct). In the case of *I. nobilis*, the two subspecies were intermixed within a monophyletic group. Therefore, we lumped the two subspecies together as one species, without distinguishing them as separate subspecies or ESUs.

In other cases of potentially incorrect splitting, a review of voucher and herbarium specimens demonstrated that the originally described species clearly possessed multiple, segregating morphological characters. This was so for *Inga punctata*, *steinbachii*, and *tocacheana* (paraphyletic with respect to other species), for several pairs of species that were mixed phylogenetically (*I. bourgonii* and *coruscans*, *I. acreana* and *chartacea*, and *I. lineata* and *killipiana*), and in one large clade with little genetic divergence between any species (E in Fig. 2). In all of these cases, we maintained the original identifications.

Evolutionarily significant units.—In several pairs of species (*Inga ruiziana* and morphospecies 68, *I. tomentosa* and morphospecies 71, and *I. stenoptera* and morphospecies 76), the members of the pair fell out as sister to one another in the phylogeny. Upon extensive review of herbarium vouchers of the named species, it was determined that the unnamed morphospecies did not possess sufficient distinguishing characters to be separated as distinct species. Thus, these also represent cases of incorrectly split species (highlighted in blue in Fig. 2). However, the sister groups are distinct genetically and somewhat distinct morphologically. We therefore labeled these sister groups as distinct ESUs of the nominate species. A similar situation was found for *I. venusta* and morphospecies 78, although in this case the two ESUs form a paraphyletic grade.

Summary of errors and revisions.—Once the above errors were taken into account, we revised the identifications of all 946 sequenced individuals. We then applied these revisions to the entire ecological data set of 3912 individuals (see Supplement for revised species composition data). In cases of incorrect lumping, we used the morphological characters in Table 1 to determine the identity of unsequenced individuals. In cases of incorrect splitting, it was straightforward to assign unsequenced individuals of the previously segregated species to a single species. Mistakes in individual identifications were only detectable through DNA sequencing and could not be translated to the entire data set. The total number and proportion of different types of delimitation and identification errors are given in Table 2. The species abundance distribution (SAD) for the original and revised delimitations is given in Fig. 3, showing fewer rare species under the revised delimitations.

Ecological analyses

In presenting the results of ecological analyses, we focus, for the purposes of brevity, on contrasting the results using the original delimitations against the revised delimitations treating ESUs as distinct species. The results using the revised delimitations in which ESUs were lumped as single species were similar to the latter and are not presented here.

Partitioning variation in community composition.—Using presence/absence matrices of species composition, the results of variance partitioning analyses differed markedly between the original and revised species delimitations (Table 3); namely, there was an increase in the total variation in community composition explained. This was due to a large increase in the proportion of variation explained purely by spatial variables (PCNMs). When relative abundance matrices of species composition were used, the original and revised delimitations showed nearly identically results across all analyses.

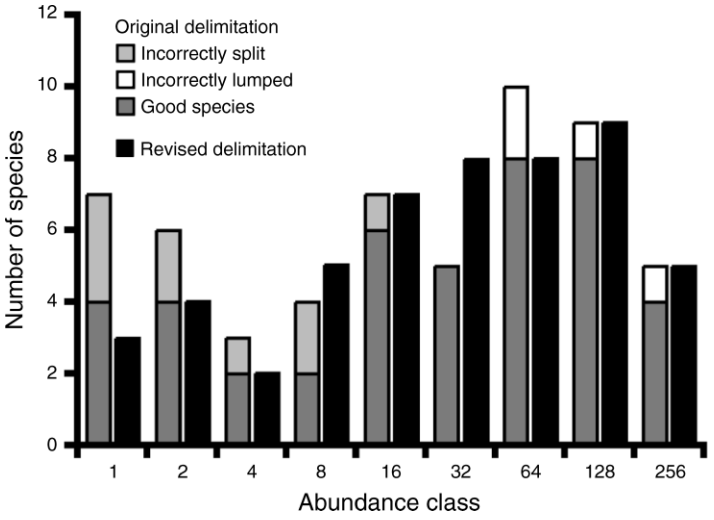


FIG. 3. The distribution of relative abundances of *Inga* species across all community surveys for the original and revised species delimitations. For the original delimitations, species that represented incorrect splitting or lumping are noted. Following convention, a $\log_2$ scale is used for the *x*-axis.

Distance decay in community similarity.—The results of distance-decay analyses also differed between the revised and original delimitations (Table 4; see Appendix C for analyses using log-transformed geographic distance). In nearly all cases (except floodplain analyses using the Bray-Curtis index), the revised delimitations showed a stronger correlation between geographic distance and community similarity. When analyses were conducted using the Jaccard index of community similarity, there were also marked differences in estimates of the slope parameter (Table 4, Fig. 4). This difference was significant when analyses were restricted to terra firme surveys (permutation test, $P = 0.012$) and marginally significant when analyses included all surveys (permutation test, $P = 0.081$). Taken together, these results indicate that the revised delimitations give greater support to dispersal limitation being an important force structuring these communities.

TABLE 3. Results of analyses to partition the variation in composition of *Inga* communities.

	Selected variables†		Variance explained (%)			
Species delimitation	Environmental	Spatial	Environment	Space	Environment/space	Unexplained
Presence/absence						
All sites						
Original	Ca, B, Na, NO_3^-, BS_K	1, 13	0.24***	0.02	0.07	0.67
Revised	Ca, B, Na, NO_3^-	1, 13, 2, 6	0.19***	0.12**	0.08	0.61
Terra firme						
Original	Mg, Zn, Mn	1, 5	0.05	0.01	0.2	0.74
Revised	Mg, P	1, 5, 3	0.00	0.12*	0.25	0.63
Floodplain						
Original	B	1, 2	0.00	0.02	0.16	0.82
Revised	B	1, 2	0.00	0.03	0.14	0.83
Relative abundances						
All sites						
Original	Ca, Cu, P	1, 13	0.28***	0.08***	0.07	0.57
Revised	Ca, Cu, P	1, 13	0.27***	0.08***	0.08	0.57
Terra firme						
Original	Mg, Zn	1, 5	0.01	0.11*	0.36	0.52
Revised	Mg, Zn	1, 5	0.02	0.12*	0.37	0.49
Floodplain						
Original	pH	1	0.00	0.00	0.13	0.87
Revised	pH	1	0.00	0.00	0.13	0.87

Note: Only the pure environmental and pure spatial fractions can be analyzed for significance, as the other two fractions are obtained via subtraction.

† These are the variables that were chosen via forward selection for each variance partitioning analysis. They are given in the order selected. Environmental variables represent soil nutrient concentrations (in the case of named chemicals), pH, or the percentage base saturation of nutrients (i.e., BS_K). Spatial variables represent spatial autocorrelation via principal components of neighbor matrices (PCNMs) (see *Materials and methods: Partitioning variation in community composition* for explanation).

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

TABLE 4. Summary of distance-decay analyses for *Inga* communities in Madre de Dios by survey sites included and similarity index.

Species delimitation	Slope	Intercept	Partial Mantel correlation
All			
Jaccard			
Original	-2.62×10^{-4}	0.363	0.14*
Revised	-3.88×10^{-4}	0.388	0.32***
Bray-Curtis			
Original	-1.96×10^{-4}	0.295	0.10
Revised	-3.36×10^{-4}	0.311	0.17**
Terra firme			
Jaccard			
Original	-4.38×10^{-4}	0.508	0.28*
Revised	-1.08×10^{-3}	0.543	0.57**
Bray-Curtis			
Original	-8.05×10^{-4}	0.481	0.41**
Revised	-9.77×10^{-4}	0.481	0.47**
Floodplain			
Jaccard			
Original	-1.11×10^{-3}	0.504	0.40***
Revised	-1.05×10^{-3}	0.508	0.41***
Bray-Curtis			
Original	-7.52×10^{-4}	0.461	0.27*
Revised	-6.66×10^{-4}	0.473	0.23*

Notes: The slope and intercept of the relationship between community similarity and geographic distance were estimated using a general linear model while the strength and significance of the relationship were evaluated using partial Mantel tests.
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Nevertheless, both the original and revised delimitations did produce similar overall results. Both consistently showed significant distance decay across and within habitat types and for both community similarity indices (Table 4, Fig. 4). This result was consistent no matter which environmental distance matrix was used as a covariate in the analyses. The results shown in Table 4 are those in which we constructed the environmental distance matrix using the Euclidean distance between communities along the first principal component axis of all soil variables. The first axis explained 51% of the variation in the soils data, while all other axes individually explained at most 15% of the variation. This environmental distance matrix showed the strongest relationship with community similarity matrices, and distance-decay analyses using alternative environmental distance matrices as covariates showed the same or even stronger distance decay (results not shown).

Species-level ecological analyses.—The original and revised delimitations often differed in how species were classified with respect to habitat specialization (Appendices D and E). For example, in three of the six cases in which species or ESUs were split based on the reciprocal illumination procedure, the newly segregated species was classified differently than the species with which it was originally lumped (*I. laurina*, *poeppigiana*, and *sertulifera*). In three of nine cases in which a species or ESU was lumped with another species, the originally segregated species was classified differently than the species with which it is now lumped (*I. leiocalycina*, *nobilis*, and *venusta*). Using the revised species delimitations, there were fewer rare species in general and therefore fewer cases with too few individuals to detect habitat specialization (Table 5). For species with sufficient sample size to perform the binomial test, both original and revised delimitations showed the large majority of species to be habitat specialists (Table 5; 74% for the original delimitations vs. 76% for the revised).

DISCUSSION

This study represents the first large-scale assessment of the accuracy of vegetative-morphology-based delimitation and identification of tropical tree species. We constructed a DNA sequence phylogeny for one-quarter

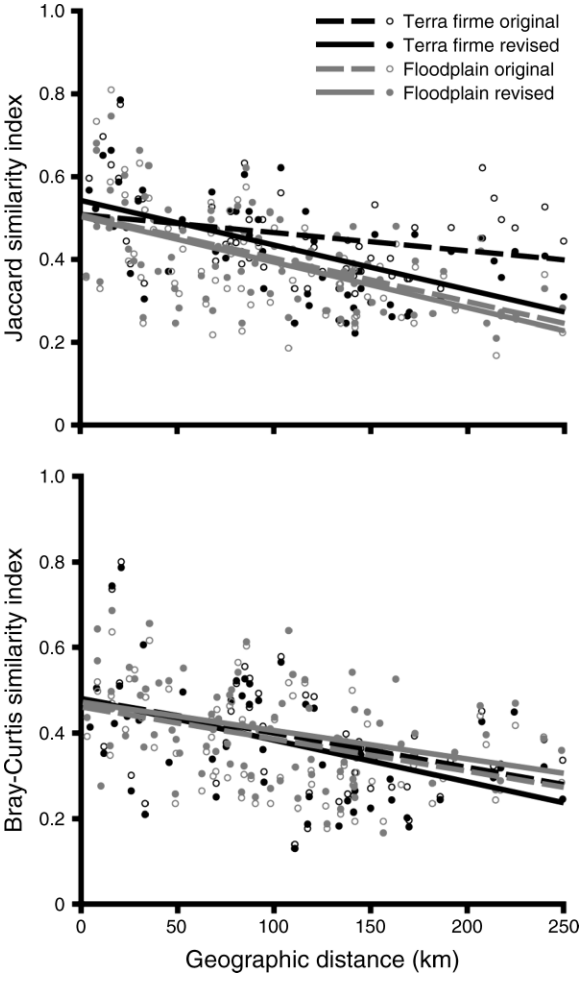


FIG. 4. Decline in (upper panel) Jaccard similarity index and (lower panel) Bray-Curtis similarity index with geographic distance between communities in Madre de Dios for floodplain and terra firme *Inga* communities, showing original and revised species delimitations. Best-fit lines were obtained using a general linear model.

TABLE 5. Summary of species-level ecological analyses.

Classification	Original delimitations	Revised delimitations
Floodplain specialist	18	20
Terra firme specialist	16	18
Neutral	12	12
Too rare to perform test†	18	11
Total	64	61

Notes: The total number of species (with evolutionarily significant units treated as distinct species) for each category according to the results of binomial tests for habitat specialization is given (see *Materials and methods: Species-level ecological analyses*).

† Sample sizes for these species were too low to successfully implement the binomial test (see *Materials and methods: Species-level ecological analyses* for explanation).

of nearly 4000 individuals across 63 putative species that were encountered in our conventional ecological study of tropical trees. Through a procedure of reciprocal illumination, we revised the morphological identifications using the generated phylogenies. Our revised assessments revealed only 55 *Inga* species, with 6.8–7.6% of stems having been misidentified in some manner (Table 2).

Identification errors led to systematic underestimation of the effect of geography

The morphological identifications systematically underestimated the potential strength of dispersal limitation in our system. For example, species that were incorrectly lumped tended to be geographically segregated, while there was no geographic signal to species that were incorrectly split. Morphological identifications therefore overestimated the proportion of species shared across space (Fig. 4) and underestimated the effect of spatial autocorrelation on community composition (Tables 3 and 4). The effect of these errors was largely due to incorrect lumping of terra firme species, as analyses of floodplain communities did not differ greatly between the original and revised delimitations (Table 3). In addition to better reflecting geography, the revised identifications were able to explain a greater total proportion of the variation in community composition.

The slope of the distance–decay relationship may depend upon the aggregation of common species (Morlon et al. 2008), which can be driven by dispersal-related processes. By incorrectly lumping geographically disjunct species in terra firme, we underestimated their aggregation and obtained a slope to the distance–decay relationship that was artificially shallow (Table 4, Fig. 4). We thus underestimated the strength of dispersal limitation in terra firme communities. Once identification errors were corrected, the slope converged to the value found for floodplain communities (Table 4, Fig. 4). It is interesting that both habitats show the same pattern, although the exact biological significance of this is uncertain. Overall though, the results of distance-decay analyses were not greatly affected by identification errors. Most slope and intercept values do not differ significantly between the original and revised identifications.

Both environmental and geographic variables are important to community assembly

When data from all community surveys are included, we find that environmental factors, namely the soil variables we measured, are more important than potentially neutral factors, such as dispersal limitation, in determining the species composition of *Inga* communities. As described below, the effects of environmental factors are due almost entirely to differences between terra firme and floodplain habitats.

Despite the importance of environmental factors (Table 3, all surveys), spatially autocorrelated, potentially neutral factors do significantly affect composition, just less so than environmental factors (Table 3, all surveys). This is particularly evident when analyses are restricted to community surveys within habitat types (Table 3, terra firme and floodplain surveys). The fraction of variation explained by the purely environmental component drops to a nonsignificant, nearly zero value. Instead, variation in community composition is explained by purely spatial factors (in terra firme) and the correlated effects of environmental and spatial factors (in both terra firme and floodplain). A sampling scheme that removes spatial autocorrelation from environmental variables (cf. Gilbert and Lechowicz 2004) would be needed to completely tease apart the effects of these factors.

When including all possible measures of environmental variation as covariates, we consistently found significant distance decay in community similarity, as calculated by either the Jaccard index or the Bray-Curtis index (Table 4). This suggests that dispersal limitation, a largely neutral process, is responsible for the observed decline with distance in community similarity, although distance decay could be due instead to unmeasured environmental gradients (Nekola and White 1999, Legendre et al. 2005, 2008). However, we have measured any potentially significant soil variables, and climate varies little across our 150 × 200 km study region (Killeen et al. 2007). This leaves neutral factors as the most likely cause of any potential correlations between compositional and geographic distance in our system.

Effect of identification errors on species-level analyses

Ecological analyses at the level of individual species were often dramatically impacted by identification errors (Appendices D and E). For example, *Inga laurina*, which was originally classified as a generalist, was found to comprise two ESUs with contrasting habitat preferences. In contrast, *Inga* morphospecies 79 was designated as preferring floodplain habitat, but was lumped with *I. leiocalycina*, which prefers terra firme habitat. If either

PLATE 1. Long-term forest dynamics plate at Nouragues Research Station, French Guiana. This plot is typical of those implemented to study the population dynamics and ecology of tropical trees. The majority of trees in these plots are never observed in a fertile state. Photo credit: Elodie Courtois.

of these types of errors (incorrect lumping or splitting of species with contrasting habitat preferences) were overrepresented, we could have misestimated the proportion of species that are habitat specialists. However, the two types of errors largely balanced out, and both delimitations found a high proportion of species to be habitat specialists.

Effect of identification errors on interpretation of neutral theory

Systematic errors in splitting species led us to overestimate the number of rare species, which in turn altered the species abundance distribution, or SAD (Fig. 3). The SAD is of great relevance in ecology (e.g., Preston 1962), and determining the probability distribution that best fits the SAD has previously been used to test neutral theory (Hubbell 2001, McGill 2003, Volkov et al. 2003). These tests can be particularly sensitive to the number of rare species, and our original and revised delimitations do give different results for some of these tests (K. G. Dexter, *unpublished data*). However, analyses of SADs actually have little ability to adequately evaluate neutral theory (McGill et al. 2006).

Instead, we evaluated neutral theory following the approach of partitioning the variation in community composition between ecological and neutral factors (Tuomisto et al. 2003, Legendre et al. 2009). The results from this approach were robust to misidentifications. Across both the original and revised delimitations, our analyses show that ecological factors, namely the specialization of species on different soil environments, are paramount in determining the composition of *Inga* communities. Secondary to this, neutral factors, such as dispersal limitation, may also influence community composition and cause a significant decline with distance in compositional similarity, particularly within habitat types. This supports the idea that neutral factors may be most important to determining community structure within homogeneous environments (Zillio and Condit 2007, Jabot et al. 2008). Our results are in line with the results of other studies of tropical tree communities (e.g., Condit et al. 2002, Phillips et al. 2003, Tuomisto et al. 2003, John et al. 2007, Norden et al. 2007, Queenborough et al. 2007, Morlon et al. 2008) and plant ecological studies in general (Tilman 1994, Gurevitch et al. 2006).

Comparison with other estimates of error rates

We classified identification errors into three categories: mistakes in individual identification, incorrect lumping of species, and incorrect splitting of species (Table 2). The first category included 1.7% of all stems. This is higher than the individual misidentification rate calculated on Barro Colorado Island (BCI) for the entire tree flora (0.85%), but similar to the rate on BCI for the diverse tree genus *Protium* (1.6%; Condit 1998).

We assessed errors in species delimitation (incorrect lumping or splitting) in addition to individual misidentifications and found a total error rate of 6.8–7.6% (Table 2). These total error rates are on par with those in temperate plant ecology studies (5.6–10.5%, Archaux et al. 2006; 7.4%, Scott and Hallam 2002). This suggests

that tropical tree ecology studies do not seem to be subject to higher rates of error in delimitation and identification despite the higher diversity and perhaps greater potential for taxonomic confusion.

Phylogenetic analyses and species delimitations

We used a method of reciprocal illumination between morphological characters and gene genealogies to delimit species. Our phylogeny represents the most complete sampling to date of a diverse, tropical tree genus in any one geographical area. Although there is low bootstrap support and posterior probability values for many nodes in the tree (Fig. 2), we obtained better resolution than in previous phylogenetic studies of *Inga* (Richardson et al. 2001, Coley et al. 2005). This may be due to the greater length, in base pairs, of our chloroplast marker, our approach of concatenating the ITS and trnD-T sequences, or differential taxa selection.

Our reciprocal illumination procedure substantially improved our initial, morphology-only identifications (7.6% of stem identifications were changed). Also, in the final delimitations, 32 of 50 species (with more than one sequenced individual) formed monophyletic groups in the phylogeny, while under the original delimitations only 12 of 50 species were monophyletic.

Nevertheless, under the revised delimitations, many species remained paraphyletic (eight species), polyphyletic (four species), or even shared sequences with other species (six species). These species did not differ from monophyletic species in ecological abundance, habitat preference, or any other evident factors, and these species form cohesive morphological entities according to vegetative characters. Thus, our results could indicate that vegetative morphology-based taxonomy has an even higher magnitude of error than we have stated here (e.g., due to cryptic species). What is more likely is that the phylogenetic results are due to incomplete lineage sorting, which is highly probable in *Inga*. However, further sampling of both individuals and genetic regions is needed to definitively determine the causes of conflict between vegetative morphology and the phylogenetic results.

In using morphology and gene genealogies, we eschewed previously published methods (e.g., Davis and Nixon 1992, Wiens and Penkrot 2002, Nielsen and Matz 2006, Pons et al. 2006, Hart and Sunday 2007, Knowles and Carstens 2007, Rach et al. 2008) that could potentially use sequence data to evaluate the accuracy of morphological species delimitation for several reasons. First, many methods work best with a limited number of species (e.g., Nielsen and Matz 2006, Knowles and Carstens 2007), while the genus *Inga* contains >300 species (Pennington 1997). Second, other methods (e.g., Davis and Nixon 1992, Wiens and Penkrot 2002) rely on geographic distribution information, particularly on allopatry, in assessing species delimitations. Limited knowledge of species distributions in the Amazon prevents us from making definitive statements about whether putative species are sympatric or allopatric. Furthermore, there is incredibly high sympatry among *Inga* species (>20 can be found on one soil type in one location; Supplement; Valencia et al. 2004), including among closely related species. This limits the usefulness of allopatry as a criterion for species delimitation.

Finally, any method that relies solely on genetic information to delimit species would not likely be successful in our system. For example, Hart and Sunday (2007) proposed using statistical parsimony methods (Clement et al. 2000) with a 95% connection limit to delimit species with DNA sequence data. If groups of sequences fall out as separate networks in the analysis, then they are presumed to represent distinct species. We applied this method to our data set, and all of our *Inga* sequences fell out in one network (K. G. Dexter, *unpublished data*), which would indicate, by their method, that we have only sampled one *Inga* species. Unless traditional morphological, taxonomic methods are terribly wrong, we have in fact sampled a much greater number of *Inga* species.

The statistical parsimony method (Clement et al. 2000), and other methods that rely solely on genetic information, may fail in *Inga* for several reasons. *Inga* is a rapidly radiating genus, which has attained a diversity of >300 species in 2–10 million years (Richardson et al. 2001, Lavin 2006). Given this rapid rate of speciation and the slow rate of evolution of the molecular markers we have used, genetic non-monophyly of species is to be expected due to incomplete lineage sorting (Avise and Ball 1990). Hybridization may also play a role in obscuring patterns of monophyly, although an assessment of hybridization between seven sympatric *Inga* species in Costa Rican montane forests found no evidence for interspecific fertility (Koptur 1984).

Ecological information (e.g., habitat preference) can also be useful in delimiting species (cf. Raxworthy et al. 2007, Rissler and Apodaca 2007). However, we have avoided using ecological information in delimiting species as we are interested in testing neutral theory and determining whether species are effectively neutral.

Implications for DNA barcoding of tropical trees

DNA barcoding has been heralded as an approach that will allow us to document and classify the great plant diversity of tropical regions in a timely manner (Kress et al. 2005, Cowan et al. 2006, Lahaye et al. 2008). We have used a phylogenetic approach in combination with morphology to revise species delimitation. In the previous section, we argued that without morphological information, even phylogeny-based approaches to DNA barcoding (Pons et al. 2006, Knowles and Carstens 2007) would likely fail with our data, as many of our species are non-monophyletic and certain clades show profound phylogenetic mixing of morphologically distinct species (e.g., clade E in Fig. 2; these probably represent species that have radiated too

recently to be distinguished with our existing sequence data).

Genetic-distance-based approaches have also been suggested for DNA barcoding (e.g., Hebert et al. 2003, Lahaye et al. 2008), but these show even less promise than tree-based approaches. In multiple cases in our data, identical sequences are shared across very morphologically distinct species, and no matter what genetic distance threshold is set, these species will not be resolved as distinct. In other words, our data show no evidence for a DNA barcoding gap (Meyer and Paulay 2005) between intraspecific and interspecific divergences.

The ITS region has been advocated for use as a DNA barcoding marker in plants (Kress et al. 2005). The trnD-T intergenic spacer has not, but it is over twice as long, in base pairs, as other chloroplast intergenic spacers advocated for barcoding (e.g., trnH-psbA) and should therefore contain as many or more substitutions. Our data show that neither of these markers will be entirely successful in the DNA barcoding of diverse tropical genera such as *Inga*, and it is these diverse genera that form the bulk of tropical plant diversity. However, it must be noted that we are referring to DNA barcoding sensu stricto, at the species level. If one wishes DNA barcoding to be successful in identifying species to clades or higher taxonomic levels than species, then there may be more room for optimism.

Conclusions

We have shown that systematic errors in the identification of tropical trees can affect the results of ecological studies of tropical tree communities. These errors principally consisted of incorrectly classifying rare, morphological variants of common species as distinct species and incorrectly lumping geographically segregated, morphologically similar species as single species. It is in these two areas that tropical tree ecologists encounter the most difficulty in making species identification and delimitation decisions. We have demonstrated that DNA sequence data can be useful to improve identification accuracy in these challenging situations. In studies in which great emphasis is placed on the results of analyses of single species, particularly when that species is presumed to occur across multiple sites, we advocate using DNA sequence data to confirm the common identity of sampled individuals. However, in community-level ecological analyses, particularly those that incorporate relative abundance data, the results obtained are fairly robust to misidentifications. In these situations, it may not be necessary to conduct the massive sequencing efforts presented here.

Our approach has the additional benefit of contributing significantly to biodiversity documentation. Adding DNA sequence data to our conventional ecological study revealed four species that would not have been documented otherwise (incorrectly lumped species). Furthermore, our study uncovered 18 named species not previously known from the study region (Madre de Dios, Peru) and discovered 10 species potentially new to science. We agree with Janzen et al. (2005) and Caesar et al. (2006) that the combination of traditional morphology-based biodiversity inventories with large-scale DNA sequence data generation offers the most fruitful approach to documenting biodiversity in a timely manner in threatened, tropical environments. Tropical ecologists, who often make extensive collections in remote, rarely visited locations, are well poised to take up this approach and contribute substantially and importantly to knowledge of species' distributions and the discovery of new species.

Acknowledgments

We thank Paul Manos, Toby Pennington, John Terborgh, Cam Webb, and an anonymous reviewer for reading and improving earlier drafts of this manuscript. This work was funded by an NSF Doctoral Dissertation Improvement Grant to C. W. Cunningham, K. G. Dexter, and John Terborgh and grants to K. G. Dexter from the Duke University Graduate School, the Duke University Biology Department, the Lewis and Clark Fund of the American Philosophical Society, the Exploration Fund of the Explorer's Club, Sigma Xi, the Society of Systematic Biologists, the Caribbean and Latin American Studies Consortium, the Amazon Conservation Association, and the Organization for Tropical Studies. K. G. Dexter was supported by an NSF pre-doctoral fellowship during the time this work was completed. The authors also express gratitude to the Instituto Nacional para Recursos Naturales (INRENA) of Peru for permission to conduct fieldwork in Peru and to Fabiola Para, Carlos Lazo, Marcos Rios, Nallaret Davila, Dilys Vela, Edwin Quispe, Juan Carlos Lara, and Agustin Mishaha for assistance in the field.

Literature Cited

Alvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29:417–434.

Archaux, F., F. Gosselin, L. Berges, and R. Chevalier. 2006. Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. Journal of Vegetation Science 17:299–306.

Avise, J. C., and R. M. Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Pages 45–67 *in* D. Futuyama and J. Antonovics, editors. Oxford surveys in evolutionary biology. Oxford University Press, Oxford, UK.

Barratt, B. I. P., J. G. B. Derraik, C. G. Rufaut, A. J. Goodman, and K. J. M. Dickinson. 2003. Morphospecies as a substitute for Coleoptera species identification, and the value of experience in improving accuracy. Journal of the Royal Society of New Zealand 33:583–590.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pages 289–303 *in* P. C. Hoch and A. G. Stephenson, editors. Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis, Missouri, USA.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22:148–155.

Blanchet, F. G., P. Legendre, and D. Borcard. 2008. Forward selection of explanatory variables. Ecology 89:2623–2632.

Borcard, D., and P. Legendre. 2004. SpaceMaker2. ⟨www.bio.umontreal.ca/casgrain/en/labo/spacemaker.html⟩

Bortolus, A. 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. Ambio 37:114–118.

Brown, B., R. M. Emberson, and A. M. Paterson. 1999. Mitochondrial COI and II provide useful markers for Wiseana (Lepidoptera: Hepialidae) species identification. Bulletin of Entomological Research 89:287–293.

Burger, W. C. 1975. The species concept in *Quercus*. Taxon 24: 45–50.

Caesar, R. M., M. Sorensson, and A. I. Cognato. 2006. Integrating DNA data and traditional taxonomy to streamline biodiversity assessment: an example from edaphic beetles in the Klamath ecoregion, California, USA. Diversity and Distributions 12:483–489.

Chase, M. W., N. Salamin, M. Wilkinson, J. M. Dunwell, R. P. Kesanakurthi, N. Haidar, and V. Savolainen. 2005. Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society B 360:1889–1895.

Chave, J., and E. G. Leigh. 2002. A spatially explicit neutral model of beta-diversity in tropical forests. Theoretical Population Biology 62:153–168.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9:1657–1659.

Coley, P. D., et al. 2005. Divergent defensive strategies of young leaves in two species of *Inga*. Ecology 86:2633–2643.

Colwell, R. K. 2006. EstimateS: statistical estimation of species richness and shared species from samples. University of Connecticut, Storrs, Connecticut, USA.

Condit, R. 1998. Tropical forest census plots: methods and results from Barro Colorado Island, Panama and comparison with other plots. Springer-Verlag, Berlin, Germany.

Condit, R., et al. 2002. Beta-diversity in tropical forest trees. Science 295:666–669.

Connell, J. H. 1971. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. Pages 289–312 *in* P. J. den Boer and G. R. Gradwell, editors. Dynamics of numbers in populations. Center for Agricultural Publication and Documentation, Wageningen, The Netherlands.

Connell, J. H. 1978. Diversity in tropical rainforests and coral reefs: high diversity of trees and corals is maintained only in a non-equilibrium state. Science 199:1302–1310.

Cowan, R. S., M. W. Chase, W. J. Kress, and V. Savolainen. 2006. 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55:611–616.

Davis, J. I., and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Systematic Biology 41:421–435.

Demesure, B., N. Sodzi, and R. J. Petit. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Molecular Ecology 4:129–131.

Denslow, J. S. 1987. Tropical rainforest gaps and tree species diversity. Annual Review of Ecology and Systematics 18: 431–451.

Derraik, J. G. B., G. P. Closs, K. J. M. Dickinson, P. Sirvid, B. I. P. Barratt, and B. H. Patrick. 2002. Arthropod morphospecies versus taxonomic species: a case study with Araneae, Coleoptera, and Lepidoptera. Conservation Biology 16:1015–1023.

Doyle, J. J., and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13–15.

Dray, S. 2007. Packfor: an R package for forward selection. ⟨http://www.bio.umontreal.ca/legendre/⟩

Duivenvoorden, J. F., J. C. Svenning, and S. J. Wright. 2002. Ecology—beta diversity in tropical forests. Science 295:636–637.

Gilbert, B., and M. J. Lechowicz. 2004. Neutrality, niches, and dispersal in a temperate forest understory. Proceedings of the National Academy of Sciences USA 101:7651–7656.

Goldstein, P. Z. 1997. How many things are there? A reply to Oliver and Beattie, Beattie and Oliver, Oliver and Beattie, and Oliver and Beattie. Conservation Biology 11:571–574.

Gurevitch, J., S. M. Scheiner, and G. A. Fox. 2006. The ecology of plants. Sinnauer, Sunderland, Massachusetts, USA.

Hart, M. W., and J. Sunday. 2007. Things fall apart: biological species form unconnected parsimony networks. Biology Letters 3:509–512.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B 270:313–321.

Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of birds through DNA barcodes. PLoS Biology 2:1657–1663.

Hickerson, M. J., C. P. Meyer, and C. Moritz. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Systematic Biology 55:729–739.

Hubbell, S. P. 1979. Tree dispersion, abundance, and diversity in a tropical dry forest. Science 203:1299–1309.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.

Hubbell, S. P., R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, and S. L. de Lao. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. Science 283:554–557.

Jabot, F., R. S. Etienne, and J. Chave. 2008. Reconciling neutral community models and environmental filtering: theory and an empirical test. Oikos 117:1308–1320.

Janzen, D. H. 1970. Herbivores and the number of tree species in tropical forests. American Naturalist 104:501–528.

Janzen, D. H., M. Hajibabaei, J. M. Burns, W. Hallwachs, E. Remigio, and P. D. N. Hebert. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Philosophical Transactions of the Royal Society B 360:1835–1845.

Jobson, R. W., and M. Luckow. 2007. Phylogenetic study of the genus Piptadenia (Mimosoideae: Leguminosae) using plastid trnL-F and trnK/matK sequence data. Systematic Botany 32:569–575.

John, R., J. W. Dalling, K. E. Harms, J. B. Yavitt, R. F. Stallard, M. Mirabello, S. P. Hubbell, R. Valencia, H. Navarrete, M. Vallejo, and R. B. Foster. 2007. Soil nutrients influence spatial distributions of tropical tree species. Proceedings of the National Academy of Sciences USA 104:864–869.

Jones, M. M., H. Tuomisto, D. Borcard, P. Legendre, D. B. Clark, and P. C. Olivas. 2008. Explaining variation in tropical plant community composition: influence of environmental and spatial data quality. Oecologia 155:593–604.

Kalliola, R., J. Salo, M. Puhakka, and M. Rajasilta. 1991. New site formation and colonizing vegetation in primary succession on the western Amazon floodplains. Journal of Ecology 79:877–901.

Killeen, T. J., M. Douglas, T. Consiglio, P. M. Jorgensen, and J. Mejia. 2007. Dry spots and wet spots in the Andean hotspot. Journal of Biogeography 34:1357–1373.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. Systematic Biology 56:887–895.

Knowlton, N., E. Weil, L. A. Weigt, and H. M. Guzman. 1992. Sibling species in *Montastraea annularis*, coral bleaching, and the coral climate record. Science 255:330–333.

Koptur, S. 1984. Outcrossing and pollinator limitation of fruit set: breeding systems of neotropical *Inga* trees (Fabaceae: Mimosoideae). Evolution 38:1130–1143.

Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen. 2005. Use of DNA barcodes to identify

flowering plants. Proceedings of the National Academy of Sciences USA 102:8369–8374.
Lahaye, R., M. Van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough, and V. Savolainen. 2008. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences USA 105:2923–2928.
Lavin, M. 2006. Floristic and geographic stability of discontinuous seasonally dry tropical forests explains patterns of plant phylogeny and endemism. Pages 433–447 *in* R. T. Pennington, J. A. Ratter, and G. P. Lewis, editors. Neotropical savannas and seasonally dry forests: plant biodiversity, biogeographic patterns, and conservation. CRC Press, Boca Raton, Florida, USA.
Legendre, P., D. Borcard, and P. R. Peres-Neto. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. Ecological Monographs 75: 435–450.
Legendre, P., D. Borcard, and P. R. Peres-Neto. 2008. Analyzing or explaining beta diversity? Comment. Ecology 89:3238–3244.
Legendre, P., X. Mi, H. Ren, K. Ma, M. Yu, I.-F. Sun, and F. He. 2009. Partitioning beta diversity in a subtropical broadleaved forest of China. Ecology 90:663–674.
Maddison, D. R., and W. P. Maddison. 2003. MacClade. Sinauer, Sunderland, Massachusetts, USA.
McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. Nature 422:881–885.
McGill, B. J., B. A. Maurer, and M. D. Weiser. 2006. Empirical evaluation of neutral theory. Ecology 87:1411–1423.
Meyer, C. P., and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biology 3:2229–2238.
Moritz, C. 1994. Defining evolutionarily significant units for conservation. Trends in Ecology and Evolution 9:373–375.
Morlon, H., G. Chuyong, R. Condit, S. Hubbell, D. Kenfack, D. Thomas, R. Valencia, and J. L. Green. 2008. A general framework for the distance-decay of similarity in ecological communities. Ecology Letters 11:1–14.
Nanney, D. L. 1982. Genes and phenes in Tetrahymena. BioScience 32:783–788.
Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26:867–878.
Newmaster, S. G., A. J. Fazekas, and S. Ragupathy. 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Canadian Journal of Botany 84: 335–341.
Nielsen, R., and M. Matz. 2006. Statistical approaches for DNA barcoding. Systematic Biology 55:162–169.
Norden, N., J. Chave, A. Caubere, P. Chatelet, N. Ferroni, P. M. Forget, and C. Thebaud. 2007. Is temporal variation of seedling communities determined by environment or by seed arrival? A test in a neotropical forest. Journal of Ecology 95: 507–516.
Oh, S. H., and D. Potter. 2003. Phylogenetic utility of the second intron of LEAFY in Neillia and Stephanandra (Rosaceae) and implications for the origin of Stephanandra. Molecular Phylogenetics and Evolution 29:203–215.
Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, G. L. Simpson, T. Solymos, M. H. H. Stevens, and H. Wagner. 2008. Vegan: community ecology package. ⟨http://cc.oulu.fi/~jarioksa/softhelp/vegan.html⟩
Oliver, I., and A. J. Beattie. 1993. A possible methods for the rapid assessment of biodiversity. Conservation Biology 7: 562–568.
Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. Science 276:734–740.
Pennington, T. D. 1997. The genus *Inga*: botany. Royal Botanic Gardens, Kew, UK.
Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87:2614–2625.
Phillips, O. L., P. N. Vargas, A. L. Monteagudo, A. P. Cruz, M. E. C. Zans, W. G. Sanchez, M. Yli-Halla, and S. Rose. 2003. Habitat association among Amazonian tree species: a landscape-scale approach. Journal of Ecology 91:757–775.
Pitman, N. C. A., J. Terborgh, M. R. Silman, and P. Nuñez V. 1999. Tree species distributions in an upper Amazonian forest. Ecology 80:2651–2661.
Pitman, N. C. A., J. W. Terborgh, M. R. Silman, P. Nuñez V., D. A. Neill, C. E. Ceron, W. A. Palacios, and M. Aulestia. 2001. Dominance and distribution of tree species in upper Amazonian terra firme forests. Ecology 82:2101–2117.
Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55:595–609.
Posada, D. 2006. Collapse: Describing haplotypes from sequence alignment. ⟨http://darwin.uvigo.es/software/collapse.html⟩
Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.
Preston, F. W. 1962. Canonical distribution of commonness and rarity. Ecology 43:185–215.
Queenborough, S. A., D. Burslem, N. C. Garwood, and R. Valencia. 2007. Habitat niche partitioning by 16 species of Myristicaceae in Amazonian Ecuador. Plant Ecology 192: 193–207.
R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rach, J., R. DeSalle, I. N. Sarkar, B. Schierwater, and H. Hadrys. 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proceedings of the Royal Society B 275:237–247.
Raxworthy, C. J., C. M. Ingram, N. Rabibisoa, and R. G. Pearson. 2007. Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. Systematic Biology 56:907–923.
Richardson, J. E., R. T. Pennington, T. D. Pennington, and P. M. Hollingsworth. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science 293:2242–2245.
Rissler, L. J., and J. J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). Systematic Biology 56: 924–942.
Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Scott, W. A., and C. J. Hallam. 2002. Assessing species misidentification rates through quality assurance of vegetation monitoring. Plant Ecology 165:101–115.
Sheil, D. 1995. A critique of permanent plot methods and analysis with examples from Budongo Forest, Uganda. Forest Ecology and Management 77:11–34.
Stanford, A. M., R. Harden, and C. R. Parks. 2000. Phylogeny and biogeography of Juglans (Juglandaceae) based on matK and ITS sequence data. American Journal of Botany 87:872–882.
Stuart, B. L., and U. Fritz. 2008. Historical DNA from museum type specimens clarifies diversity of Asian leaf turtles (*Cyclemys*). Biological Journal of the Linnean Society 94: 131–141.
Swofford, D. L. 2002. PAUP. Sinauer, Sunderland, Massachusetts, USA.

Tilman, D. 1994. Competition and biodiversity in spatially structured habitats. Ecology 75:2–16.
Tuomisto, H., and K. Ruokolainen. 2006. Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. Ecology 87:2697–2708.
Tuomisto, H., K. Ruokolainen, and M. Yli-Halla. 2003. Dispersal, environment, and floristic variation of western Amazonian forests. Science 299:241–244.
Valencia, R., R. B. Foster, G. Villa, R. Condit, J. C. Svenning, C. Hernandez, K. Romoleroux, E. Losos, E. Magard, and H. Balslev. 2004. Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. Journal of Ecology 92:214–229.
Vecchione, M., W. F. Mickevich, K. Fauchald, B. B. Collette, A. B. Williams, T. A. Munroe, and R. E. Young. 2000. Importance of assessing taxonomic adequacy in determining fishing effects on marine biodiversity. ICES Journal of Marine Science 57:677–681.
Volkov, I., J. R. Banavar, S. P. Hubbell, and A. Maritan. 2003. Neutral theory and relative species abundance in ecology. Nature 424:1035–1037.
Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. American Naturalist 156:145–155.
White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315–322 *in* M. Innis, D. Gelfand, J. Sninsky, and T. White, editors. PCR protocols: a guide to methods and applications. Academic Press, San Diego, California, USA.
Wiens, J. J., and T. A. Penkrot. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Systematic Biology 51:69–91.
Wills, C., R. Condit, R. B. Foster, and S. P. Hubbell. 1997. Strong density- and diversity-related effects help to maintain tree species diversity in a neotropical forest. Proceedings of the National Academy of Sciences USA 94:1252–1257.
Zillio, T., and R. Condit. 2007. The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. Oikos 116:931–940.
Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological datasets under the maximum likelihood criterion. University of Texas, Austin, Texas, USA.

APPENDIX A

Maximum-likelihood tree for the nuclear internal transcribed spacer (ITS) locus for all sequences used in concatenated analysis (*Ecological Archives* M080-009-A1).

APPENDIX B

Maximum-likelihood tree for the chloroplast trnD-T locus for all sequences used in concatenated analysis (*Ecological Archives* M080-009-A2).

APPENDIX C

Summary of distance-decay analyses for *Inga* communities in Madre de Dios using log(geographic distance) (*Ecological Archives* M080-009-A3).

APPENDIX D

Results of habitat-specialization analyses of *Inga* species in Madre de Dios as per the original morphology-based species delimitations (*Ecological Archives* M080-009-A4).

APPENDIX E

Results of habitat-specialization analyses of *Inga* species in Madre de Dios as per the species delimitations revised based on the reciprocal illumination procedure (*Ecological Archives* M080-009-A5).

SUPPLEMENT

Raw data used in our analyses (*Ecological Archives* M080-009-S1).