THE EFFECTS OF DISPERSAL ON MACROECOLOGICAL PATTERNS

by

Kyle Graham Dexter

Department of Biology
Duke University

Date:________________________
Approved:

____
Cliff Cunningham, Co-Supervisor

____
John Terborgh, Co-Supervisor

____
Paul Manos

____
Jim Clark

____
Mohamed Noor

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2008

ABSTRACT

THE EFFECTS OF DISPERSAL ON MACROECOLOGICAL PATTERNS

by

Kyle Graham Dexter

Department of Biology
Duke University

Date:_______________________
Approved:

____
Cliff Cunningham, Co-Supervisor

____
John Terborgh, Co-Supervisor

____
Paul Manos

____
Jim Clark

____
Mohamed Noor

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Ph.D. in the Department of
Biology in the Graduate School
of Duke University

2008

Abstract

Ecologists have long sought to uncover the mechanisms behind large-scale, macroecological patterns in the distribution and abundance of species. Macroecological patterns are often attributed to the dynamics of dispersal (e.g. dispersal limitation or widespread dispersal). However, few studies actually measure dispersal to determine if dispersal rates are commensurate with the observed macroecological patterns. In this dissertation, I use population genetic analyses across many species to obtain community-level estimates of dispersal rates for two different ecological systems: birds on islands and trees in tropical rainforests. These independent estimates of dispersal then allow me to determine if macroecological patterns in these two systems can be attributed to dispersal dynamics.

In chapter two, I explore the contrasting macroecological patterns of two groups of Lesser Antillean birds. The groups' differing macroecological patterns could be due to differences in dispersal, but other authors have advocated different mechanisms. Population genetic analyses show that the two groups do differ significantly in rates of inter-island dispersal, indicating that dispersal dynamics can explain their contrasting macroecological patterns.

In chapter three, I turn my attention to tropical tree communities. In contrast to studies of birds on islands, studies of trees in tropical rainforests may suffer from

misidentification of individuals in the field. Using a phylogenetic approach, I determine errors rates in identification, and then assess the effect of these errors on macroecological patterns and other ecological analyses of tropical tree communities. I find that error rates are substantial, but that they have little effect on macroecological patterns. In contrast, species-level ecological analyses can be dramatically affected by these errors.

In chapter four, I return to the influence of dispersal on macroecological patterns, this time in tropical tree communities. One notable macroecological pattern in Amazonian tree communities is a high correlation in the relative abundances of species shared across communities, which could indicate high rates of dispersal between communities. However, population genetic analyses show that dispersal is severely limited between communities. Thus, some factor besides dispersal, such as differences in competitive ability or susceptibility to disease, must be driving species to achieve similar relative abundances in geographically separated communities. In contrast, I show that dispersal limitation is the likely cause of another macroecological pattern frequently observed in tropical tree communities: the decline in the compositional similarity of communities with distance. However, this is not steady-state dispersal limitation in an equilibrium framework as is conventionally thought. Instead, the dispersal limitation appears to be historical in nature, which implies a heretofore unnoticed role for historical contingency in the assembly of Amazonian tree communities.

Dedication

To my family and friends, for their unconditional love and support through graduate school and my whole life.

Contents

List of Tables

List of Figures

Acknowledgements

Both of my advisors, Cliff Cunningham and John Terborgh, deserve deep thanks for their support of my scientific career. Before I had any experience working in tropical rainforests, John took a chance on me, hiring me to help him set up a long-term study on seed dispersal in Amazonian rainforests. Without that opportunity and the knowledge I gained in that year, this dissertation research would have been impossible. Since I began at Duke, Cliff has continually been there for me, for everything from last minute recommendations or edits to unsolicited, but welcome morale boosts.

My other committee members also deserve much thanks. Paul Manos has been like a third advisor to me, and I have always felt at home in his lab. Jim Clark has helped me to think critically and to see more clearly what the exciting frontiers in ecology are. Mohamed Noor's enthusiasm for a topic so distant from his own research has been very welcome.

This dissertation research involved extensive field work, which benefited from the generous assistance of many people. For assistance in Manu, I thank John Terborgh, Patricia Alvarez, Veronica Chavez, Carmen Chavez, Julian, Lucha, and Juan Curaca. For assistance in Los Amigos, I thank Nigel Pitman, Edwin Quispe, el Diablo, Jorge Chavez, John Janovec, Jason Wells, the Amazon Conservation Association, and the staff at CICRA. For assistance in northern Peru, I thank Devon Graham, Marcos Rios, Nallaret

Davila, Dilys Vela, Julio, Ruben, Proyecto Amazonas, and the staff at Madreselva and Sabalillo. For assistance in Las Piedras, I thank Emma Hume, Edwin Quispe, Juan Carlos Lara, Julio, and the staff of Tambopata Expeditions. For assistance in Tambopata, I thank Don Brightsmith, Karina Quinteros, Agustin Mishaha, and the staff at Tambopata Research Center. Two Peruvian undergraduates, Fabiola Parra Rondinel and Carlos Lazo Gaga, deserve special acknowledgement for putting in long hours in the field for weeks on end and providing good company in the process.

The Instituto Nacional para Los Recursos Naturales (INRENA) of Peru is deeply acknowledged for their permission to conduct all of this field research and to export DNA specimens to the United States.

At the Duke Herbarium, Connie Roberts and Sheri Henderson were both of great help. At Missouri Botanic Garden, Roy Gereau was a wonderful host. Terry Pennington, at Kew Botanic Garden, deserves a special thanks for making this dissertation possible, both through his revision of the genus *Inga* and his help in identifying my specimens.

For assistance and/or guidance in the laboratory, I thank Caroline Kim, Diana Massey, Zehao Shen, Bernie Ball, Sang-Hun Oh, Kathryn Perez, Jason McLachlan, Christy Henzler, and Cynthia Riginos. I also warmly thank Toby Pennington for being open and sharing with primer information and protocols for *Inga* genetics.

This research would not have been possible without funding from the following sources: the National Science Foundation, Duke University Graduate School, Duke

University Biology Department, Sigma Xi, Society for Systematic Biologists, Explorer's Club, American Philosophical Society, Organization for Tropical Studies, Consortium for Latin American and Caribbean Studies, and Amazon Conservation Association.

Lastly, I must thank my fellow graduate students and other friends in Durham for making graduate school such an enjoyable process. In particular, I thank Suzanne Joneson, Greg Bonito, Todd Twigg, Jen McCulley, Annette Baudisch, Kathryn Perez, members of the Duke Natural History Society, the bio basketball crew, all residents past and present of the Lakewood house, and various folks in the Nicholas School who have put up with me. Erin Tripp deserves special recognition as a faithful traveling companion who also taught me a lot.

1. Introduction

One of the principal tasks of ecology is to explain patterns in the distribution and abundance of species. This is no trivial task, as these patterns can be influenced by such diverse factors as the abiotic environment, biotic interactions, dispersal, and historical biogeography. This thesis focuses particularly on the role of dispersal in generating macroecological patterns (ecological patterns across multiple communities). Many fundamental macroecological patterns, such as species-area curves and the decline in community similarity with distance, have been interpreted in terms of dispersal (MacArthur and Wilson 1963, 1967, Nekola and White 1999, Condit et al. 2002). However, rarely are dispersal estimates made to determine if they are commensurate with the patterns observed. This is important, because alternative explanations will always exist for macroecological patterns.

In this thesis, I synthesize population genetic studies of many individual species to provide community-level estimates of dispersal rates. I then use these estimates to more properly assess macroecological patterns and determine if they are influenced by dispersal dynamics. I use this approach for two model systems in community ecology: birds on islands and trees in tropical rainforests.

Birds on islands have a long history of study in ecology and evolutionary biology (e.g. Lack 1947, MacArthur and Wilson 1963, 1967, Terborgh 1973, Diamond 1975, Diamond and Mayr 1976, Grant 1999, Ryan et al. 2007). One of the most frequently examined macroecological patterns for bird communities on islands is the species-area curve. Many authors (e.g. MacArthur and Wilson 1967, Diamond and Mayr 1976, Schoener 1976) have suggested that variation in the steepness of the species-area curve, or the value of the power-law exponent, is due to variation in dispersal, or immigration, rates. However, it has been pointed out that the steepness of the species-area curve can vary for a variety of reasons (Connor and McCoy 1979, Martin 1981). Likewise, the decline with distance, or distance-decay, in the similarity of bird communities on islands has been interpreted in terms of dispersal (Terborgh 1973, Fernandez-Palacios and Andersson 1993), but again, other factors may be responsible. Estimates of dispersal are needed to determine if variation in the form of these macroecological patterns is indeed due to variation in dispersal rates.

I examine these two macroecological patterns, species area curves and distance-decay in community similarity, for two sets of bird communities in the Lesser Antillean archipelago. Rainforest communities show a steeper species-area curve and greater distance-decay in community similarity between islands than dry forest communities. This contrast suggests that rainforest birds disperse between islands more often than dry forest species, but alternative explanations exist (Ricklefs and Cox 1978, Ricklefs and Bermingham 2004). Next, I take advantage of publicly available population genetic data from the majority of rainforest and dry forest species (Ricklefs and Bermingham 2001) to assess rates of inter-island dispersal. I use two approaches to estimate dispersal: 1) conventional analyses of isolation by distance based on M, the estimated number of migrants per generation between islands and 2) a novel method utilizing information from D_a, the genetic distance between island populations. Both approaches show that rainforest species disperse significantly less often between islands than dry forest species, thus validating a role for dispersal dynamics in shaping macroecological patterns for island bird communities.

Tropical tree communities have a more recent history as a model system in community ecology (Hubbell and Foster 1986, Condit et al. 2002, Losos et al. 2004). Studies of tree communities in rainforests and bird communities on islands differ in many ways. One important difference is in the quality of the data. Legions of bird watchers visit archipelagos, such as the Lesser Antilles, every year and identify all the birds they encounter based on well-established, diagnostic characters. Thus, one can be fairly confident about which bird species are found on which islands. In contrast, tropical tree communities are often visited only by the ecologists that study them, and these ecologists must identify a diverse set of tree species based on poorly-known characters of questionable diagnostic utility. Tropical ecologists may frequently be incorrect in their identifications of trees and this may affect the results and conclusions of their studies, including those of macroecological patterns. However, there is currently little data to provide information on this.

As part of my thesis work on dispersal and macroecological patterns, I conducted surveys of tropical tree communities at 25 sites across the southern Peruvian Amazon. This involved identifying nearly 4,000 individual trees. In order to verify the accuracy of my identifications, I use a phylogenetic approach and assess the monophyly of putative species. To construct the phylogeny, I obtained DNA sequences from a large subset (25%) of all the individuals surveyed. According to the genealogical species concept (Baum and Shaw 1995), all individuals of a given species should be more closely related to each other than to individuals of other species and should form a monophyletic group in a phylogeny. If individuals of a given species appear at divergent locations in a phylogeny or are nested with other species, this can indicate errors in identification. I tabulate such errors and show that 8.7% of my identifications were incorrect in some manner.

In order to determine how erroneous identifications could be affecting studies of tropical tree communities, I compare macroecological patterns generated using my original identifications to patterns generated using identifications that were corrected based on the phylogenetic analyses. I find surprisingly that despite the large error rate, most macroecological patterns in tropical tree communities are robust to identification error. For example, using either set of identifications, soil type and edaphic factors explain most of the variation in the species composition of tree communities, while there is significant distance-decay in community similarity not explained by soil type or edaphic factors. In contrast, species-level analyses can be drastically affected by incorrect identifications. The phylogenetic analyses reveal that some purportedly generalist species are actually comprised of two species that specialize on different environments. Other species are found to have broader habitat tolerances than originally thought after phylogenetic analyses indicate that they should be combined with species having a different environmental preference.

Once the identifications had been verified, I investigated the potential effect of dispersal on the form of macroecological patterns in tropical tree communities. Researchers frequently observe that, within a given habitat type (e.g. terra firme), the similarity of tropical tree communities in terms of species composition declines with geographic distance (Condit et al. 2002, Tuomisto et al. 2003, Morlon et al. 2008). This has generally been attributed to dispersal limitation in tropical rainforests. However, this pattern could also be due to cryptic environmental gradients within habitat types (Pitman et al. 2008). In other studies, researchers have found that the relative abundances of species shared between tropical tree communities are highly correlated across space, even between communities separated by as much as 1,500 km (Terborgh et al. 1996, Pitman et al. 2001, Vormisto et al. 2004, Macia and Svenning 2005). This could signify consistent ecological determinism of species abundances or could be indicative of high rates of dispersal between communities. Thus, the two patterns described here, if interpreted in terms of dispersal, make contradictory predictions about the level of dispersal limitation in rainforest landscapes. In order to determine which of the above patterns is due to dispersal dynamics, estimates of dispersal between tropical tree communities are needed.

I estimated dispersal rates for 20 tropical tree species between two river basins in southern Amazonian Peru. Within both of the dominant habitat types (terra firme and floodplain), species' abundances are highly correlated between the two basins. However, seed dispersal, as measured by gene flow for a chloroplast genetic marker, is highly limited between the two basins. In fact, most species show no significant migration at all between the two basins. This correlation in relative abundances across space persists across all the communities surveyed during my dissertation, and does not decline with distance. Thus, the macroecological pattern of a high correlation in the relative abundance of species across these communities is not due to high rates of dispersal. Other ecological forces, such as competition or differential susceptibility to disease, must be acting to bring species to the same relative abundances in different communities.

The observed dispersal limitation is commensurate with a pattern of distance-decay in the proportion of species shared between communities, which I also documented for these communities. I explored this pattern further to determine if it is due to a steady-state, equilibrium scenario of dispersal limitation (cf. Hubbell 2001) or to historically limited dispersal. I find evidence for the latter. Further population genetic analyses show that seed dispersal between the two river basins mentioned above has been more limited than anywhere else in the landscape. Also, most of the turnover in species composition in the landscape occurs between these two river basins. Thus, it appears that this could be a transition zone where two communities of different provenance are coming together. Geological and paleoecological data suggest that rainforests may have retreated from parts of the southern Peruvian Amazon in the Pliocene or Pleistocene (Espurt et al. 2007, Olivier 2007, Olivier et al. 2008). If the area is now being colonized from multiple source populations, this could explain the ecological and phylogeographic results that I observe. While further work is needed to confirm the validity of this scenario, it is highly suggestive. Patterns of distance-decay in community similarity in tropical tree communities may be due to historical processes of extinction and re-colonization, and not just to equilibrium scenarios of dispersal limitation. Furthermore, this implies that membership in tropical tree communities can depend, at least in part, on historically contingent factors.

I now place the results of my dissertation in the broader context of our understanding of ecology and of the forces that control the distribution and abundance of species. I have shown that dispersal, particularly limited dispersal, can limit where species are found and can profoundly affect some macroecological patterns. However, species environmental preferences are still more important than dispersal in determining where species are found, at least at the spatial scales examined here. In my study of Lesser Antillean birds, I first classified birds as being either rainforest or dry forest species. While dispersal limitation may be preventing rainforest birds from occupying some islands, they are not found in dry forest on any island. If I use a multiple regression approach to analyze the causes of community similarity, habitat type is by far the most important factor, with dispersal limitation falling out as secondary. This is also true in Amazonian tree communities. While dispersal limitation is important in determining macroecological patterns within terra firme or floodplain habitats, habitat type is still the strongest determinant of where species are found.

However, it can be argued that ecologists already understand that habitat type is important in determining the species composition of communities, and that is why I have focused here on the role of dispersal. Recent ecological theory, particularly neutral theory, has emphasized the role of dispersal in community assembly and has minimized the role of species-dependent ecological processes. Needless to say, the proposition that species differences are not important has stirred up great controversy. I show that dispersal limitation requires recognition as an important force structuring ecological communities, but my research offers little support for the idea that species are neutral or ecologically equivalent. Although dispersal may limit where a species is found, ecological interactions between that species and its environment, and perhaps with other species, determine that species abundance. While further research is needed to elucidate the mechanisms that limit species dispersal and the ecological forces that determine species abundances, this dissertation has provided an important step forward in clarifying the role of dispersal in shaping macroecological patterns.

2. The Influence of Dispersal on Macroecological Patterns in Lesser Antillean Birds

2.1 Introduction

Dispersal can have a profound effect on macroecological patterns such as species-area curves and the decline in similarity in species composition of communities with geographic distance (distance-decay) (Gaston and Blackburn, 2000). Theory predicts that increased dispersal limitation should increase the *z*-value, or power law exponent, of the species-area relationship because local extinctions, which are more probable in smaller areas, are not readily replaced through re-colonization (MacArthur and Wilson, 1963, 1967; Bell, 2001; Hubbell, 2001; Chave et al., 2002). Theory also suggests that increased dispersal limitation should increase distance-decay in community similarity because high dispersal is the primary means by which community similarity is maintained in the face of differential local extinctions and potential allopatric speciation (Bell, 2001; Hubbell, 2001; Chave and Leigh, 2002; Mouquet and Loreau, 2003; Morlon et al., 2008). Early investigations of the role of dispersal in shaping macroecological patterns often focused on island systems (e.g. MacArthur and Wilson, 1963, 1967; Terborgh, 1973; Diamond and Mayr, 1976), while most recent theoretical developments have concerned continuous environment, mainland systems (all citations post-2000 above). Nevertheless, island archipelagos continue to serve as model systems in macroecology (Rosenzweig, 1995; Gaston and Blackburn, 2000; Whittaker, 2007), and most of the recent theoretical predictions from mainland systems should apply to island systems as well.

A recent simulation study demonstrated that, within the same archipelago, poorly dispersing taxa should have steeper species-area curves than taxa that disperse well, as predicted by theory (Hovestadt and Poethke, 2005). This result has also been found empirically in studies comparing volant versus non-volant animals (Wright, 1981; Lomolino, 1984). Diamond and Mayr (1976) classified birds according to their dispersal ability, and showed that putatively poorly dispersing birds have steeper species-area curves. Given that birds do not differ in any conspicuous way that relates to their dispersal abilities, as volant and non-volant animals do (sensu Wright, 1981; Lomolino, 1984), Diamond and Mayr (1976) instead grouped bird species into dispersal ability categories based on their distribution patterns (Diamond, 1975). However, this is circular because community-level macroecological patterns, such as species-area curves, stem directly from species-level distribution patterns. Other factors besides dispersal, such as habitat specialization or differential extinction rates, can also affect species' distribution patterns. Therefore, it could be erroneous to conclude that dispersal affects macroecological patterns when dispersal has been assessed solely based on distribution patterns. Many other studies have concluded that dispersal affects the form of species-area curves (e.g. Brown, 1971; Terborgh, 1973; Strong and Levin, 1975) without estimating dispersal in any rigorous manner (Connor and McCoy 1979). Likewise, the rate of distance-decay in community similarity has been interpreted in terms of dispersal when no attempts to estimate dispersal were made (e.g. Nekola and White, 1999; Condit et al., 2002). In order to validate a role for dispersal in shaping macroecological patterns, empirical studies are needed that estimate dispersal rates independently from the data used to elucidate the patterns.

There is a long history of macroecological studies of Lesser Antillean bird communities (Ricklefs and Cox, 1972; Terborgh, 1973; Ricklefs and Cox, 1978; Terborgh et al. 1978; Terborgh and Faaborg, 1980; Faaborg, 1982; Gotelli and Abele, 1982; Case et al., 1983; Faaborg, 1985; Ricklefs and Lovette, 1999; Ricklefs and Bermingham, 1999, 2001, 2004a), which makes these communities a fertile ground for examining how dispersal can affect macroecological patterns. Ricklefs and colleagues (Ricklefs and Cox, 1972; Ricklefs and Bermingham, 2004a) have emphasized the role of extinction, rather than dispersal, in generating contrasting species-area curves in Lesser Antillean birds. They show how putatively old colonists have a steeper species-area curve than recent colonists, because they colonized the archipelago longer ago and have had more time to go extinct on individual, usually smaller islands (for the purpose of simplification, I here refer to both the young species and widespread endemics of Ricklefs and Bermingham, 2004a as recent colonists and to the restricted endemics as old colonists). However, Ricklefs and Bermingham (2004a) used genetic distance estimates between Antillean and mainland populations to classify species as old versus recent colonists (Ricklefs and

Bermingham 2001), and these estimates could also be affected by migration rates (Cherry et al. 2002, Ricklefs and Bermingham 2004b). Putatively old colonists may actually be species with low dispersal rates, while putatively recent colonists may be species with high dispersal rates (Cherry et al. 2002, Ricklefs and Bermingham 2004b), and the differences in their species-area curves could instead be due to differences in dispersal ability.

In fact, previous researchers have predicted, although indirectly, that these different classes of birds have different rates of dispersal (Terborgh et al. 1978, Terborgh and Faaborg 1980). Terborgh and colleagues (Terborgh et al., 1978; Terborgh and Faaborg, 1980) actually predicted that rainforest species in the Lesser Antilles disperse less between islands than dry forest species, but their prediction is relevant because the putative old colonists of Ricklefs and colleagues are generally found in interior, rainforest habitats while the putative recent colonists are often found in peripheral, dry forest or scrub habitats (Ricklefs and Bermingham 2004a). The prediction of Terborgh and colleagues is based on the contrasting distribution patterns of rainforest and dry forest species; rainforest species are often absent from islands with suitable rainforest habitat while dry forest species are present nearly everywhere there is dry forest (Terborgh et al., 1978). However, I again emphasize that inferring dispersal rates solely from distribution patterns can be problematic. Rainforest species may be found on fewer islands, not because of reduced dispersal ability, but because of increased extinction probability as suggested by Ricklefs and colleagues. As the role of inter-island dispersal in shaping macroecological patterns in Lesser Antillean birds is still unresolved (Cherry et al. 2002, Ricklefs and Bermingham 2004b), a fresh examination of these patterns is warranted, particularly in integration with rigorous estimates of dispersal.

In this study, I examine macroecological patterns, specifically species-area curves and distance-decay in community similarity, for dry forest and rainforest Lesser Antillean bird communities. While many authors have constructed species-area curves for different classes of Antillean birds (Ricklefs and Cox, 1972; Terborgh, 1973; Terborgh and Faaborg, 1980; Faaborg, 1985; Ricklefs and Lovette, 1999; Ricklefs and Bermingham, 2004a), no study has explicitly contrasted these curves for dry forest and rainforest communities. Likewise, distance-decay in community similarity has been examined in Lesser Antillean bird communities (Terborgh, 1973), but patterns of distance-decay have not been compared for dry forest and rainforest communities. I find that rainforest communities have a steeper species-area curve and greater distance-decay in community similarity. Based on this, I predict, like Terborgh and colleagues (Terborgh et al., 1978; Terborgh and Faaborg 1980), that rainforest species should disperse between islands less frequently than dry forest species. I then take advantage of the remarkable phylogeographic data set that Ricklefs and Bermingham (2001) have assembled for Lesser Antillean birds (mitochondrial sequence data is available from multiple islands for half the dry forest and over half the rainforest species) to determine if differential inter-island dispersal can be responsible for the observed contrast in macroecological patterns. Importantly, the phylogeographic data is independent of species distribution data and thus provides a powerful test of the role of dispersal in shaping macroecological patterns.

2.2 Materials and Methods

2.2.1 Study System and Data Sources

The Lesser Antilles extend southeastwards from Puerto Rico and the Greater Antilles to the northern coast of Venezuela. The two principal habitats in the Lesser Antilles are lower elevation dry forest (sclerophyll scrub) and higher elevation rainforest (Terborgh et al., 1978; Faaborg, 1985). Faaborg (1985) and Ricklefs and Bermingham (2004a) give the island distributions and habitat preference for Lesser Antillean bird species; birds are classified as rainforest, dry forest, generalist, or miscellaneous species (the latter utilize rare habitats such as stream banks or mangroves). I constructed rainforest and dry forest specific species lists for all possible islands (n=13 for rainforest, n=16 for dry forest). Not all rainforest or dry forest species are confined exclusively to their designated habitat throughout the Lesser Antilles (Terborgh et al., 1978; Faaborg, 1985), but they do exhibit sufficient habitat preferences to not be classified as generalists (Faaborg, 1985; Ricklefs and Bermingham, 2004a). Generalist species and those with a miscellaneous habitat preference were excluded from habitat-specific species lists.

I obtained the Ricklefs and Bermingham (2001) mitochondrial DNA sequences (ATPase 6 and 8 genes; 842 bps in total) from multiple island populations for 8 of the 16 total dry forest species and 13 of the 24 total rainforest species (see Table 1 for species and Genbank accession numbers). Island populations are represented by anywhere from 1 to 14 sequenced individuals, although most island populations (69 of 91) are represented by only 2 individuals.

2.2.2 Distribution Analyses

While others have noted that rainforest species occupy fewer islands than dry forest species (Ricklefs and Cox, 1978; Terborgh et al., 1978), it was of interest to determine if this difference is statistically significant. I computed for each species the fraction of islands with suitable habitat that are occupied. I also counted how frequently a given species is absent from an island with suitable habitat when the species is present on a neighboring island. I compared dry forest and rainforest species for both of these measures using a Mann-Whitney U-Test.

2.2.3 Species-Area Curves

I obtained the area of dry forest and rainforest on each island from Ricklefs and Lovette (1999). Combining this information with habitat-specific species lists for islands allowed me to construct species-area curves separately for dry forest and rainforest communities (sensu Buckley, 1982). Two islands (St. Eustatius and Saba) have less rainforest than the least amount of dry forest present on any island (9 km^2). Because species-area curves are best compared when they cover a similar range of areas (Diamond and Mayr, 1976; Connor and McCoy, 1979), I conducted analyses both excluding and including these islands. I used an ANCOVA to determine if habitat had a significant effect on the power-law exponent, or z-value, of the species-area curve.

In order to determine if a particular ecological or phylogenetic group was responsible for the differences in species-area curves between dry forest and rainforest communities, I constructed habitat-specific species-area curves for different guilds and families of birds. Guilds were obtained from Faaborg (1985) and Ricklefs and Bermingham (2004b) and consisted of frugivores, flycatching insectivores, gleaning insectivores, and nectarivores. Birds were assigned to family following Ricklefs and Bermingham (2004b). I grouped Cardinalinae, Coerebinae, Emberizinae, Icterinae, Parulinae, and Thraupinae together as Fringillidae. In the family-level analyses, I constructed species-area curves for families with at least 5 species present in the Lesser Antilles, which limited analyses to the Columbidae, Fringillidae, Mimidae, Trochilidae, Tyrranidae, and Turdidae. I added one to all species counts before log transforming them in order to include islands with zero species in analyses (cf. McGuinness, 1984). I used ANCOVA, as above, to determine if habitat had a significant effect on the z-value of guild and family level species-area curves.

2.2.4 Distance-Decay Analyses

I calculated the similarity in species composition of rainforest and dry forest communities on different islands using the Sorensen similarity index. There is only 1 rainforest species on Barbados and St. Maarten, and these islands were therefore excluded from community similarity analyses. I assessed the effect of geographic distance on similarity of communities using a Mantel test, separately for dry forest and rainforest communities. I then used a permutation-based approach to determine if the rate of distance-decay (measured as the slope of the linear regression of community similarity on geographic distance) differs between these communities more than expected by chance. In order to generate a null expectation for the difference in rate of distance-decay, I first randomly shuffled the islands in geographic space 1000 times (preserving the geographic coordinates, but not the island names and constituent bird communities associated with each pair of coordinates). I then calculated the rate of distance-decay for these null dry forest and rainforest communities. The proportion of the 1000 replicates that has a difference in the rate of distance-decay between dry forest and rainforest communities greater than that found in the real data gives a p value for this 1-tailed test. This test, which determines whether dry forest and rainforest communities have significantly different slopes to the distance-decay relationship, is analogous to an ANCOVA, but the permutation approach deals with the fact that the pairwise data points are not independent. I also conducted distance-decay analyses using log transformed similarity values and log transformed geographic distance. This and all subsequent permutation analyses were conducted in the R Statistical Environment (R Core Development Team, 2007).

2.2.5 Estimating Dispersal Between Islands

I assessed the frequency of dispersal between islands for individual bird species using the mitochondrial DNA sequence data. I focus on measuring inter-island dispersal as opposed to dispersal between the mainland and the islands, because the former is likely more important in determining the species composition of individual islands. Nearly 3/4 of the Lesser Antillean avifauna is endemic (Terborgh et al., 1978) and most non-endemic species have a deeper divergence with their putative colonization source than between island populations (Ricklefs, 2000; Ricklefs and Bermingham, 2007).

One method of quantifying dispersal is to estimate M, the number of migrants per generation between populations (Slatkin and Voelm, 1991). I used Arlequin Version 2.000 (Schneider et al., 2000) to obtain an estimate of M between each pair of island populations for each species. Arlequin Version 2.000 calculates M from F_{ST}, the genetic differentiation between populations,using the following formula: $M = (1-F_{ST})/2F_{ST}$. Some M estimates were considered unrealistically high, particularly in cases where a single sequence is present in both populations (F_{ST} is estimated as 0 and M is calculated as infinity). I conducted analyses setting the maximum M value at two threshold levels, 20 and 5000. The former value was chosen as a seemingly reasonable cap while the latter value is higher than the highest obtained inter-island estimate of M (4253 migrants per generation) and thus only affects estimates of infinity. I then log transformed all M estimates (following Slatkin, 1993). For each species, I averaged the $log_{10}(M)$ values obtained across all island pairs (so that each species is represented only once in the analysis) and compared dry forest and rainforest species for average $log_{10}(M)$ using a Mann-Whitney U-Test.

Most species were not sampled from enough islands to allow for a formal isolation by distance analysis (Slatkin, 1993). I therefore grouped all pairwise comparisons for all species by habitat to determine if the general relationship between the logarithm of geographic distance and the logarithm of M differs for rainforest and dry forest species. I used a permutation approach to assess if the slope of the relationship between log_{10}(geographic distance) and $log_{10}(M)$ differs for species in the two habitats significantly more than expected by chance. I combined all pairwise estimates of $log_{10}(M)$ and log_{10}(geographic distance) for all species into a common pool. I then randomly drew, without replacement, 8 species and their respective $log_{10}(M)$ and log_{10}(geographic distance) values to represent null dry forest communities, while the remaining 13 species in the pool represented the null rainforest communities (the original data has 8 species in dry forest and 13 in rainforest). I repeated this 1000 times and determined the proportion of replicates that have a difference in slope between rainforest and dry forest species greater than that in the real data, which gives a p-value for this statistical test. As above, this test is analogous to an ANCOVA, but takes into account the issue that the pairwise estimates for each species are not independent. I conducted this analysis both excluding M estimates of infinity and setting the maximum M at 20 or 5000.

2.2.6 An Alternative Approach to Estimating Inter-Island Dispersal

If two island populations are reciprocally monophyletic with respect to each other, then the estimate of M, based on F_{ST}, will effectively be zero, regardless of the amount of genetic divergence between the two populations. However, the genetic divergence between island populations is related to the time since the last effective dispersal event between those islands, either directly or via another island. Analyses of M ignore the information about dispersal present in the distribution of inter-island divergences, which could inform comparisons of dispersal between rainforest and dry forest communities. In order to take advantage of this information, I devised an alternative approach to compare dispersal rates between dry forest and rainforest communities based on sequence divergence between island populations.

I obtained the Nei and Li (1979) genetic distance, D_A, between all pairs of island populations for each species using the Tamura and Nei distance formula in Arlequin Version 2.000 (Schneider et al., 2000). Although the genetic distance between two populations can be influenced by many factors, such as effective population size, time since divergence, and levels of migration since divergence, I make the simplifying assumption that the genetic distance, D_A, between two populations is proportional to the time since the last dispersal event between those two populations. I combined all inter-island estimates of D_A for all species for each habitat type, giving a separate distribution of D_A estimates for dry forest and rainforest communities. I assume that the rate of dispersal between islands is constant through time for all species within a given habitat and that the extinction rate is constant through time for all species across both habitats. Given these assumptions, the distribution of D_A estimates should be exponential with a rate parameter that approximates the rate of dispersal between islands for species in a given habitat.

I next assessed whether the distributions of D_A estimates are significantly different for dry forest and rainforest communities. Because each island population is used in multiple pairwise comparisons and not all data points are independent, I used a permutation-based approach to generate a null expectation for how different the D_A distributions could be by chance. I combined the D_A estimates for all species from dry forest and rainforest into one pool and then randomly drew species (and their associated D_A values) without replacement for the null communities (8 species for dry forest and 13 for rainforest). I then obtained the likelihood ratio that the D_A estimates from the two null communities belong to two different exponential distributions instead of one. I performed this permutation 1000 times and assembled a null distribution of likelihood ratios. I determined the proportion of the null distribution that has a likelihood ratio greater than that in the original data, which gives a p-value for this 1-tailed test. I repeated this analysis using only D_A estimates between neighbouring islands.

2.3 Results

2.3.1 Distribution Analyses

Rainforest species occupy a lower proportion of islands with suitable habitat than dry forest species (Fig. 1, Mann-Whitney U-Test, $p=0.0147$). Rainforest species are also more often absent from islands with suitable habitat when there are populations present on neighbouring islands (Fig. 1, Mann-Whitney U-Test, $p=0.0160$).

2.3.2 Species-Area Curves

Rainforest communities have a significantly higher power-law exponent, or *z*-value, to their species-area relationship than dry forest communities (Fig. 2, ANCOVA, interaction term, $F=23.7$, $p<0.0001$). Equivalent results were obtained when rainforest communities with less than $9 km^2$ of rainforest were included in the analysis (ANCOVA, interaction term, $F=6.2$, $p=0.021$). When analyses are restricted to specific ecological or phylogenetic groups, the contrast between rainforest and dry forest communities persists (Table 2). All guilds and families, except Mimidae, show a steeper species-area relationship in rainforest than in dry forest, and this difference is generally significant.

2.3.3 Distance-Decay Analyses

In both dry forest and rainforest communities, similarity in community composition is inversely correlated with geographic distance (Fig. 3, Mantel test, dry forest: slope=-4.5×10^{-4} (95% CI: -5.1×10^{-4}, -4.1×10^{-4}), $r^2=0.62$, $p<0.001$; rainforest: slope=-9.4×10^{-4} (95% CI: -1.12×10^{-3}, -7.6×10^{-4}), $r^2=0.50$, $p<0.001$). However, rainforest communities are less similar on average (Fig. 3) and decline more in similarity with distance (permutation test, p=0.01). Distance-decay analyses using log transformed similarity values and geographic distances gave equivalent results and are not presented here.

2.3.4 Assessment of Dispersal Between Islands

The average $\log_{10}(M)$ estimate was significantly lower for rainforest species than dry forest species whether the maximum M value was set at 5000 (Mann-Whitney U-Test, p=0.0018) or 20 (Mann-Whitney U-Test, p=0.0022). While habitat type does not seem to have a strong effect on the relationship between $\log_{10}$(geographic distance) and $\log_{10}(M)$ when the maximum M values are set to 5000 (permutation test, p=0.129) or 20 (permutation test, p=0.207), habitat type does show a significant effect when M estimates of infinity are excluded (Fig. 4, permutation test, p=0.013).

Dry forest and rainforest communities have significantly different distributions of inter-island genetic distances, D_A (permutation test, p=0.008), which is evident in Figure 5. The tail of the exponential distribution in rainforest communities includes much deeper divergences than in dry forest communities. The significant difference in distributions persists if analyses are restricted to populations on neighbouring islands (p=0.015).

2.4 Discussion

Differences in dispersal ability are often assumed to underlie contrasting macroecological patterns, but rarely is this tested. Here, I show how population genetic data can be used to assess dispersal ability and determine if dispersal differences can indeed be responsible for differences in macroecological patterns. Significantly, this population genetic data is independent from data, such as species lists or distribution maps, used to elucidate macroecological patterns.

In the Lesser Antilles, rainforest and dry forest bird communities show contrasting macroecological patterns that would be expected if rainforest species disperse between islands less often than dry forest species. Rainforest communities have a higher z-value to the species-area curve and greater decay in community similarity over geographic distance (Figs 2 and 3). Population genetic analyses confirm that rainforest species do indeed disperse between islands significantly less frequently than dry forest species (Figs 4 and 5). This illustrates that dispersal may be responsible for shaping macroecological patterns in the Lesser Antilles, although alternative explanations do exist. These alternatives are discussed below along with interpretations of the population genetic data.

2.4.1 Interpretation of Dispersal Analyses

The sample sizes used to estimate inter-island dispersal in this study are low, often only 2 individuals per island. These low sample sizes prevented effective implementation of recent likelihood and Bayesian methods that are ideally used to estimate dispersal or migration (Beerli and Felsenstein, 2001; Nielsen and Wakeley, 2001). Estimates of F_{ST} and D_A can be obtained with such low sample sizes, although they are prone to error. However, this error should not bias results differently for dry forest and rainforest species and should only serve to erase the statistical signal of any difference between them. Thus, it is all the more remarkable that the analyses do show statistically significant differences in dispersal ability between rainforest and dry forest species. Increased sample size is unlikely to change this pattern.

Analyses of M estimates, which are derived from F_{ST} estimates, show that dry forest birds have higher average rates of migration between communities than rainforest birds. This difference in migration rate is most evident in proximate communities, generally neighbouring islands, while at greater distances the difference between the two classes of birds lessens (Fig. 4). Dispersal between neighbouring islands is likely critical to influencing distribution patterns of Lesser Antillean birds, as dispersal across the archipelago is thought to occur in a stepping-stone fashion (Ricklefs and Bermingham, 2001; Ricklefs and Bermingham, 2004a). Thus, rainforest species being more limited in dispersal between neighbouring islands can help explain why they have patchy and limited distributions.

I next examined inter-island dispersal using an alternative approach based on the distribution of inter-island genetic distances, D_A, in each habitat. This analysis also supports the idea that rainforest species disperse between islands, including neighbouring ones, less frequently than dry forest species. The rate parameter from the exponential distribution of D_A (Fig. 5) approximates the rate of dispersal between islands for the two different communities, and rainforest communities have a significantly lower rate parameter. The deep divergences between some rainforest populations (Fig. 5) indicate that these populations have been resident on the islands for an extended period of time, which argues against the idea that rainforest species occupy fewer islands because they have higher extinction rates than dry forest species.

The finding that rainforest species disperse less between islands than dry forest species is in agreement with mainland studies, which have shown that rainforest species are less likely to disperse between habitat patches than dry forest or scrub species (Sodhi, et al. 2004; Laurance et al., 2004; Antongiovanni and Metzger, 2005). Also, if passive dispersal via hurricanes or other storms is an important dispersal mechanism in the Lesser Antilles, it might be expected that dry forest species, which occur in a peripheral, low elevation habitat, are more often carried between islands in this manner (Terborgh and Faaborg, 1980).

One issue that must be addressed in this study is that I am measuring effective dispersal, which is a combination of successful dispersal and successful establishment. It may be that rainforest species do disperse between islands as frequently as dry forest species, but that it is more difficult for individuals of rainforest species to successfully establish. This could be the case if competitive pressures in rainforest are greater or if there are endemic diseases or predators in the rainforest that prevent survival of new arrivals (Ricklefs and Bermingham, 2007). I cannot rule out these possibilities, but do note that establishment limitation should have the same effect as dispersal limitation on genetic measures of dispersal as well as on macroecological patterns.

2.4.2 Alternative Explanations for Contrasting Macroecological Patterns

The distribution of relative abundances of species, the spatial aggregation of individuals within species, and sampling effects can all affect the form of macroecological patterns (He and Legendre, 2002; Plotkin and Muller-Landau, 2002; Woodcock et al., 2006; Morlon et al., 2008). This study includes no data on the relative abundance of species or on the aggregation of individuals within species, and thus I cannot quantify the effect of these factors *per se* on the observed macroecological patterns. However, the influence of these factors is generally felt through their interaction with sampling, and sampling effects are unlikely to be very significant in this study. Hundreds of amateur bird watchers, plus many professionals, visit the Lesser Antilles every year, covering all habitat types on nearly every island. Rare species are especially sought out, and these repeated surveys have likely lead to the documentation of most, if not all, resident species in dry forest and rainforest on each island.

The spatial position of islands and distance to mainland colonization sources can affect the both slope of the species-area curve and the rate of distance-decay in communities (MacArthur and Wilson, 1967; Diamond and Mayr, 1976; Nekola and White, 1999). However, I have controlled for the effects of geography in this study by contrasting communities found in the same archipelago and on the same individual islands.

Environmental heterogeneity also has the potential to shape macroecological patterns. Larger areas may contain more species than small islands because they contain greater environmental heterogeneity, not because they are at different immigration-extinction equilibria (Terborgh and Faaborg, 1980; Rosenzweig, 1995). If environmental heterogeneity scales with area to a greater degree in rainforest than in dry forest, this could explain why rainforest communities have a steeper species-area curve. However, bird species in the Lesser Antilles are not generally limited in their distribution to specific sub-habitats within rainforest or dry forest and instead occur across the breadth of environmental variation found within either habitat (Ricklefs and Cox, 1978; Terborgh and Faaborg, 1980; Faaborg, 1985). Distance-decay may be caused, not by limited dispersal, but by environmental gradients and species responding in their distribution to these environmental gradients (Nekola and White, 1999). If there are important environmental gradients, within rainforest and dry forest, across the Lesser Antilles and environmental gradients in rainforest are steeper, then this could explain the steeper distance-decay in rainforest communities. However, several authors have emphasized that neither rainforest nor dry forest change greatly in their habitat characteristics across the archipelago (Ricklefs and Cox, 1978; Terborgh and Faaborg, 1980; Ricklefs and Lovette, 1999). Thus, previous studies of Lesser Antillean bird communities suggest that environmental heterogeneity may not be an important factor driving the macroecological patterns examined in this study.

As discussed above, Ricklefs and colleagues (Ricklefs and Cox, 1972, 1978; Ricklefs and Bermingham, 1999, 2004a) have explained contrasting macroecological patterns in Lesser Antillean birds based on differential extinction probabilities. This explanation is part of their taxon cycle hypothesis, which states that bird species initially colonize peripheral, dry forest habitats in the Lesser Antilles and then shift to interior, rainforest environments over time. This hypothesis is then the basis of their suggestion that rainforest species should have been present in the archipelago longer than dry forest species and have thus suffered more island extinctions. The low inter-island genetic divergence observed for dry forest species (Fig. 5) could be due to recent colonization, as predicted by the taxon cycle hypothesis, rather than frequent inter-island dispersal, as hypothesized in this study. Ideally, the relative influence of dispersal and colonization time on the population genetic patterns could be determined (sensu Nielsen and Wakeley, 2001), but as discussed, low sample sizes prevent this. Thus, I cannot definitively rule out the taxon cycle explanation for these contrasting macroecological patterns, although I have presented a viable, alternative explanation. Also, I note that the taxon cycle hypothesis predicts no recent inter-island movement for rainforest species (Ricklefs and Cox 1978, Ricklefs and Bermingham 2004a), while the genetic data show that many rainforest species have recently moved between islands (Fig. 5), just that they do so less frequently than dry forest species.

While operating on a different timescale than the taxon cycle, Pleistocene climate cycles may have also played a role in shaping distribution patterns of Lesser Antillean birds. During glaciation phases, sea levels were lower, exposing more land that was presumably covered in dry forest scrub (Pregill and Olson, 1981). Many of the islands that currently have a small area of dry forest may have had much larger areas of dry forest in the past, particularly during the last glacial maximum. Dry forest communities on these islands may not have responded yet, in terms of island population extinctions, to this drastic reduction in area, and the shallow species-area curve for dry forest communities could be a transient phenomenon. Furthermore, during glaciation phases the climate was likely drier, and rainforests could have contracted in size, perhaps disappearing from smaller islands (Pregill and Olson, 1981), which may be why rainforest species are often absent from islands that seem to have suitable habitat today. Thus, climate fluctuations in the Pleistocene providee an alternative explanation for the current macroecological patterns of Lesser Antillean birds. Further paleoecological research is needed to address this possibility (sensu Reis and Steadman, 1999).

2.4.3 Conclusions

Rainforest communities have a steeper species-area curve than dry forest communities and greater distance-decay in community similarity. These differences are consistent across all phylogenetic and ecological groups. Both conventional analyses of M, the number of migrants per generation between island populations, and alternative analyses of D_A, the genetic distance between island populations, demonstrate that rainforest species disperse between islands less frequently than dry forest species. Thus, greater dispersal limitation in rainforest communities is a plausible explanation for these contrasting macroecological patterns. However, historical factors, such as the taxon cycle and Pleistocene climate cycles may have also a played a role in establishing distribution patterns, and further research is needed to address these alternatives.

These results have important significance for conservation and management efforts in the region. If a rainforest species goes extinct on any given island through habitat loss or otherwise, we cannot expect populations on neighbouring islands to naturally replace the lost species within a reasonable amount of time. Thus, protecting rainforest species, ideally through habitat preservation, will be essential to maintaining natural bird populations in the Lesser Antilles.

Table 1: Bird Species Included in Study

Habitat preference and Genbank Accession Numbers for species included in this study.

Species	Habitat	Accession Numbers
Columbina passerina	Dry forest	AF132377-AF132382, AF485048-AF485063
Dendroica petechia	Dry forest	AY115296-AY115306
Elaenia martinica	Dry forest	AF1323838-AF132388, AY082535-AY082548
*Eulampis holosericeus**	Dry forest	AY082564-AY082572
Mimus gilvus	Dry forest	AF140960-AF140974
Saltator albicollis	Dry forest	AY115270-AY115278
Tiaris bicolor	Dry forest	AF132428-AF132433
Tyrannus dominicensis	Dry forest	AY115258-AY115268
Cinclocerthia ruficauda†	Rainforest	AF140899-AF140921
Contopus latirostris	Rainforest	AY115201-AY115207

Cichlherminia lherminieri	Rainforest	AY115158-AY115162
Cyanophaia bicolor	Rainforest	AY115163-AY115166
Dendroica plumbea	Rainforest	AF018226-AF018252
Eulampis jugularis	Rainforest	AY082552-AY082563
Geotrygon montana	Rainforest	AY082574-AY082583
Icterus dominicensis‡	Rainforest	AF109412-AF109423
Myadestes genibarbis	Rainforest	AY115150-AY115157
Myiarchus oberi	Rainforest	AY115173-AY115180
Troglodytes aedon	Rainforest	AY115227-AY115237
Tangara cucullata	Rainforest	AY115221-AY115224
Turdus fumigatus	Rainforest	AY115238-AY115243

* Same as *Sericotes holosericeus*

† Includes both *Cinclocerthia ruficauda* and *C. guttaralis*

‡ Includes all species in the *Icterus dominicensis* species complex: *I. dominicensis, I. laudabilis, I. bonana, and I. oberi*

Table 2: Summary of Species-Area Curve Analyses for Lesser Antillean Birds

Estimates of the power-law exponent of the species-area curve (z-value) for rainforest and dry forest communities for different categories of birds. The results are given of ANCOVA tests, which were used to determine if rainforest and dry forest communities have significantly different relationships between species number and area.

	Category	Dry forest *z*-value†	Rainforest *z*-value†	ANCOVA p-value
	All Birds	0.085	0.418	<0.0001*
Guild	Frugivores	0.090	0.267	0.0152
	Gleaning Insectivores	0.164	0.313	0.0679
	Flycatching Insectivores‡	0	0.310	<0.0001*
	Nectarivores‡	0	0.181	0.0019*
Family	Columbidae	0.011	0.091	0.0551
	Fringillidae	0.146	0.200	0.4610
	Mimidae	0.217	0.160	0.5713
	Trochilidae‡	0	0.181	0.0019*
	Turdidae	-0.011	0.241	0.0016*
	Tyrannidae‡	0	0.310	<0.0001*

* Significant at the 0.05 level after a Bonferroni correction for multiple comparisons

† Slope of the curve obtained from regression of $\log_{10}$(No. Species) on $\log_{10}$(area of habitat)

‡ The nectarivore guild is comprised entirely of Trochilidae, so these two comparisons represent the same analysis; The flycatching insectivore guild is comprised entirely of Tyrannidae, so these two comparisons represent the same analysis

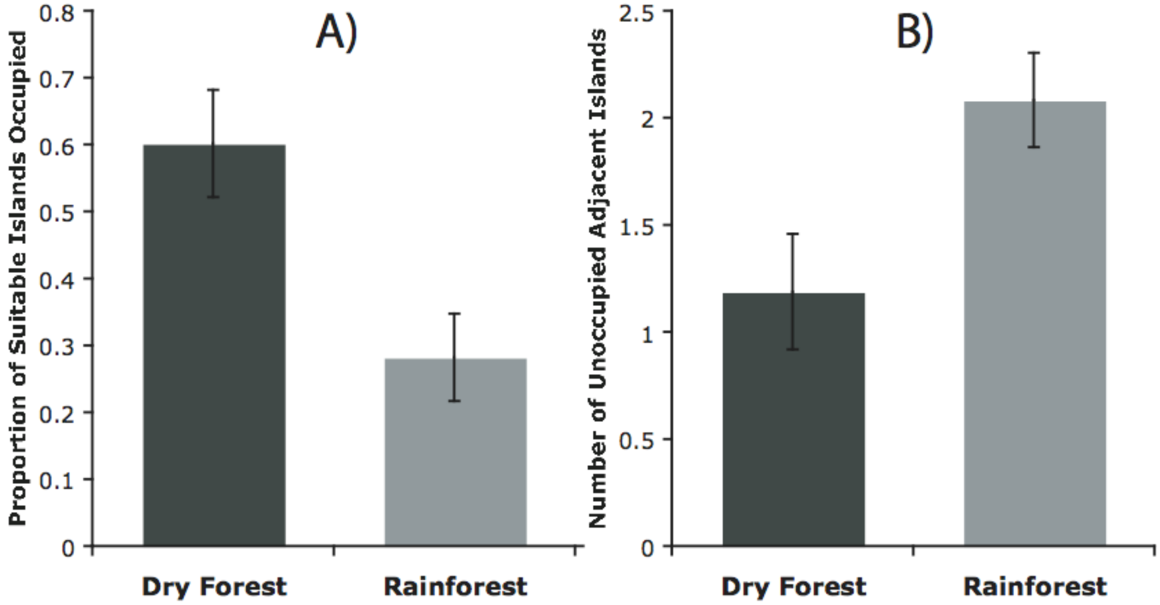


Figure 1: Distribution Patterns of Lesser Antillean Birds

A) The average proportion of islands with suitable habitat occupied by dry forest and rainforest species and **B)** the average number of unoccupied islands with suitable habitat that are adjacent to an island with a viable population. Error bars represent the standard error. Both patterns indicate that rainforest species have more limited distributions than dry forest species.

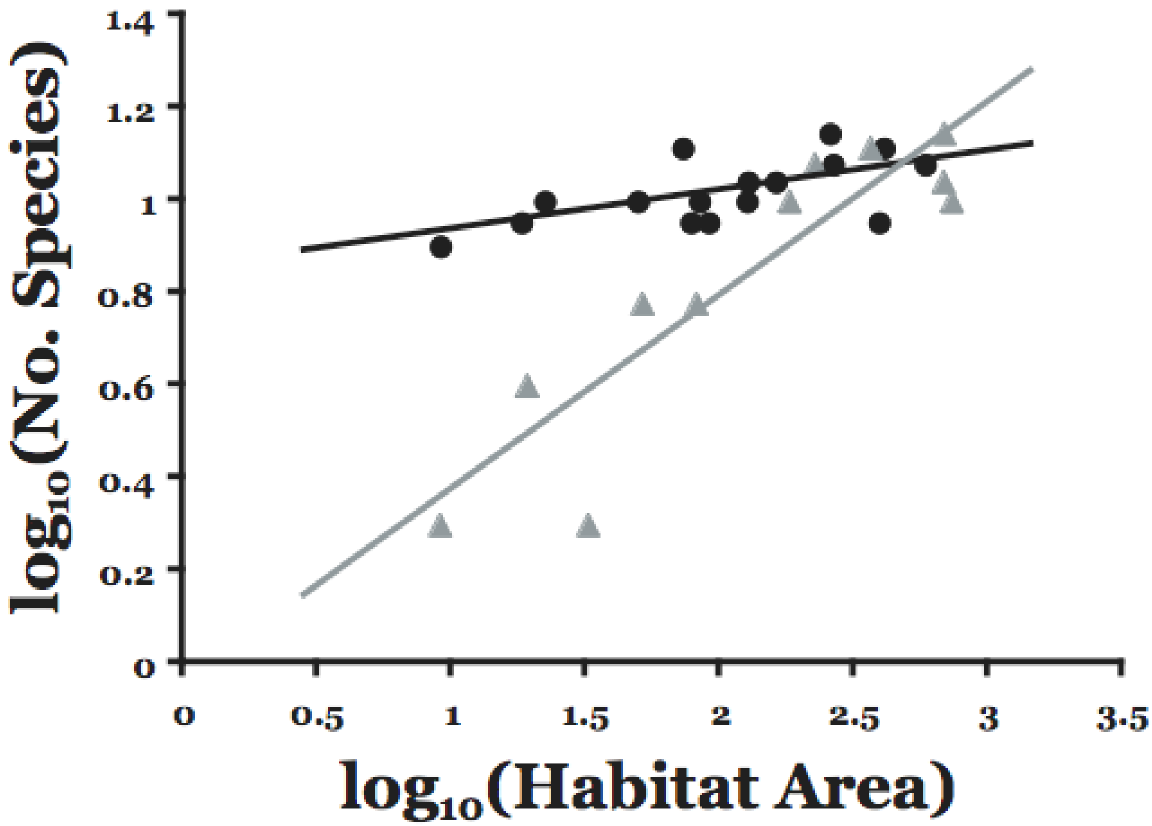


Figure 2: Species Area-Curves for Lesser Antillean Birds

The species-area relationship for dry forest communities (black circles) and rainforest communities (gray triangles) in the Lesser Antilles. The lines represent the best-fit least squares linear regression of $\log_{10}$ (no. species) on $\log_{10}$ (area of habitat). These species-area curves are significantly different (see text).

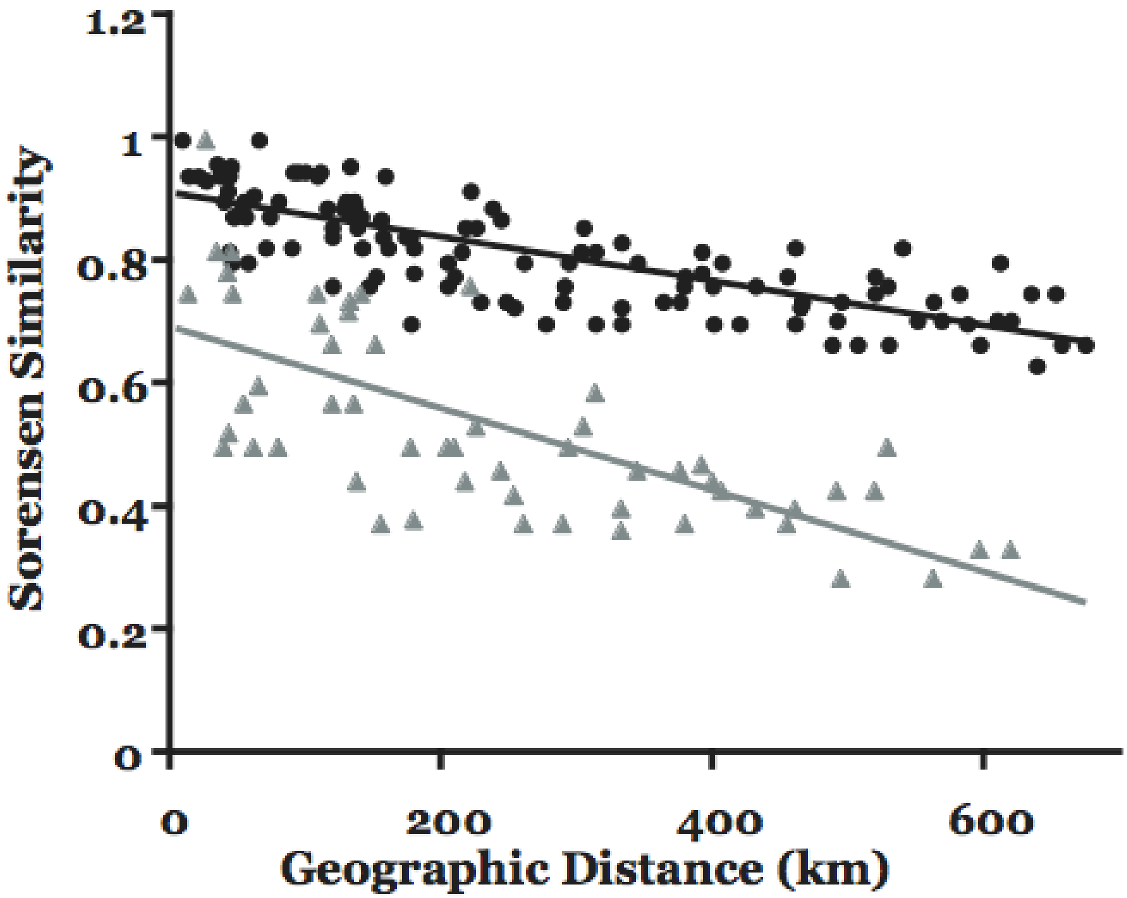


Figure 3: Distance-decay in Similarity of Lesser Antillean Bird Communities

Distance-decay in the similarity in species composition between dry forest (black circles) and rainforest communities (gray triangles). The lines represent the best-fit least squares linear regression of Sorensen similarity on geographic distance for each habitat. Significance of relationships was assessed using a Mantel test (see text for details).

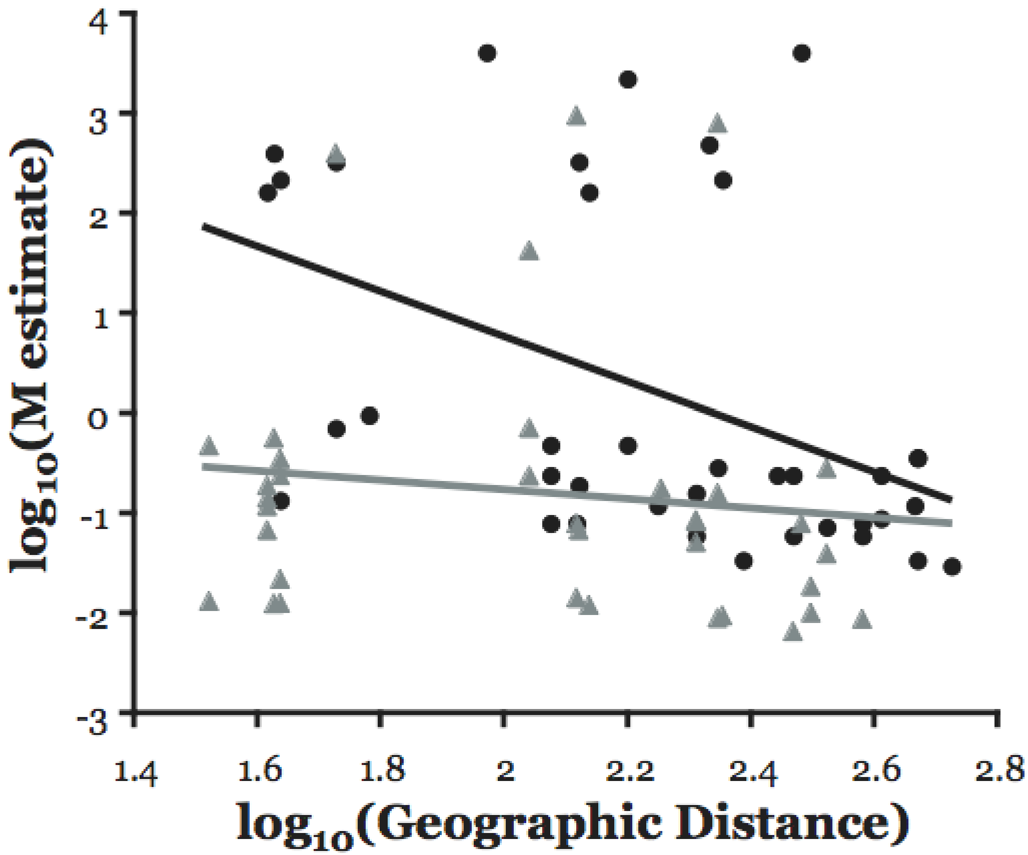


Figure 4: Isolation by Distance Patterns for Lesser Antillean Birds

Isolation by distance patterns for Lesser Antillean bird species in dry forest (black circles) and rainforest (gray triangles). The lines represent the best-fit least squares linear regression of the $\log_{10}(M)$, the number of migrants per generation, on the $\log_{10}$(geographic distance). Dry forest species have higher rates of inter-island dispersal between proximate communities, while the difference between dry forest and rainforest species lessens with distance (see text for analysis).

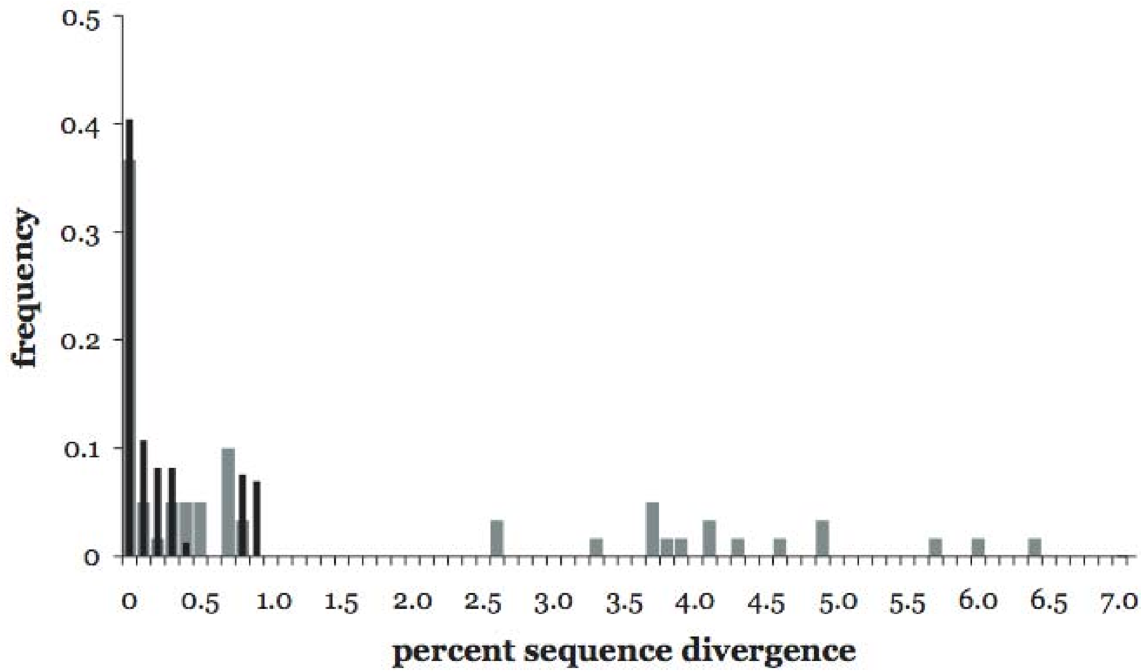


Figure 5: Inter-Island Divergences for Lesser Antillean Birds

Relative frequency of inter-island genetic distances, D_A, for dry forest (black bars, foreground) and rainforest species (grey bars, background). Nei and Li genetic distances were calculated using the Tamura and Nei distance formula in Arlequin Version 2.000 and are given as percent sequence divergence (Schneider et al. 2000). Rainforest species have a long right-hand tail of deep inter-island divergences that is lacking in the dry forest distribution.

3. Errors in Vegetative Identifications of Tropical Trees and Their Effects on Ecological Studies

3.1 Introduction

In any field-based ecological study, an ecologist must identify individuals to species. This is often difficult in the species rich tropics, with many undescribed species, and often subtle, morphological distinctions between described species. Even further, with plants, only a subset of the characters used to describe a species are available for any given individual. For example, tropical plant taxonomists primarily rely on fertile characters (i.e. the morphology of flowers, fruits, and their associated structures) to delimit species. Taxonomists generally consider vegetative characters like the morphology of leaves, twigs, bark, and wood to be highly variable and often unreliable. Yet, ecological surveys of tropical plants must count every individual, whether it is reproductive or not. For this reason, tropical woody plant ecologists have traditionally relied on vegetative characters to identify individual shrubs or trees encountered in the field. This disconnect between the characters generally used to describe species and those used to identify individuals in the field begs the question of how accurate are the species identifications made by tropical woody plant 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 identify and correct mistakes in species identification and delimitation. Then, we compare the results of ecological analyses using species identified with our method to results with species identified by vegetative characters alone.

The impact of potentially systematic errors in tropical tree species identification is important because much of 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, Hubbell 2001, Pitman et al. 2001, McGill 2003, Volkov et al. 2003). Other examples of important theory being assessed with data from tropical tree communities 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 identification mistakes are made by ecologists (Goldstein 1997, Vecchione et al. 2000, Bortolus 2008), including tropical tree ecologists (Sheil 1995, Condit 1998). Determining how these mistakes affect the results and conclusions of ecological studies is more difficult. Firstly, an independent authority or data source is needed to determine when ecologists have made an identification mistake. One useful method is to have a taxonomic expert review all of the identifications of an ecologist (c.f. Oliver and Beattie 1993, Derraik et al. 2002, Scott and Hallam 2002, Barratt et al. 2003) or to otherwise quantify error by using multiple observers (Archaux et al.

2006, Condit 1998). Alternatively, genetic data, ideally 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 each other 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, 2004b, 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 can be sufficient to delimit and identify species (Hebert et al. 2003, 2004b, Janzen et al. 2005, although see Hickerson et al. 2006, Knowles and Carstens 2007). There is less reason to expect a purely genetic approach to be as successful with plants. Plants may engage in inter-specific gene flow more often than animals (Chase et al. 2005, Cowan et al. 2006), and most of the genetic markers commonly used for plants – including the ones used in this study – evolve much more slowly than the mitochondrial markers commonly used for animals (e.g. cytochrome oxidase I) (Kress et al. 2005, Newmaster et al. 2006). In plants, closely related or hybridizing species may not be readily distinguishable using DNA sequence data alone, and morphological information may still be necessary to delimit and identify species.

As part of a conventional ecological study, we surveyed communities of the tropical tree genus Inga (Mimosoideae: Fabaceae) across 25 sites in lowland Amazonian Peru. We encountered nearly 4,000 individual trees belonging to 64 putative species. Individuals were preliminarily identified using vegetative morphological characters and with careful consultation of herbarium specimens. Of critical importance, the taxonomic expert in the genus (2nd author T.D.P.) aided in these identifications. We then constructed a phylogeny with nuclear and chloroplast DNA sequences for one quarter of the surveyed individuals (946 total), encompassing all 64 putative species. We use this phylogeny to propose hypotheses about possible mistakes in species delimitations and identifications. For example, if a putative species – delimited using vegetative characters – falls out as two separate, monophyletic groups in a phylogeny, we re-examine the morphology of the species to determine if the two groups possess consistent, distinguishing morphological characters. If such characters are found, then we consider the original delimitation as a single species incorrect. Importantly, we use two genetic markers, one nuclear and one chloroplast, to construct the phylogeny. While two markers is still a limited number, comparing the two can reveal cases of hybridization and incomplete lineage-sorting that cause errors in genealogical species delimitation (Hickerson et al. 2006, Knowles and Carstens 2007).

This process of reciprocal illumination between DNA sequence data and vegetative morphology identified many mistakes in species identification based on vegetative characters alone. We then compared ecological analyses relying on the preliminary versus the revised species identifications to determine the effect that these identification errors may have on results and conclusions. We have focused on ecological analyses that are conventionally used to test neutral ecological theory. We first examined the distribution of relative abundances and assessed if it can be statistically distinguished from a log-normal distribution (McGill et al. 2006). A log-normal distribution of relative abundances is often interpreted as evidence against neutral theory (e.g. McGill 2003), while an excess of rare species, which results in the zero-sum multinomial distribution, is considered as evidence for neutral theory (Hubbell 2001, Volkov et al.; 2003). If there are systematic errors in identification such that rare species are over or under-represented, this could affect the shape of the distribution of relative abundances and conclusions drawn there from.

We also conduct analyses that assess if variation in the species composition of communities is better explained by environmental factors or spatially autocorrelated factors, such as dispersal limitation (Duivenvoorden et al. 2002, Tuomisto and Ruokolainen 2006). If environmental factors predominate, this is evidence against neutral dynamics. However, if spatially autocorrelated factors predominate, this indicates that dispersal limitation, a potentially neutral force, may be more important in determining community composition (Hubbell 2001, Tuomisto et al. 2003). We conduct these analyses both within and across the two dominant habitat types in the region. Within habitat types, we also examine patterns of distance-decay in community similarity (Condit et al. 2002, Morlon et al. 2008). If systematic errors in identification result in incorrect lumping or splitting of geographically segregated species or of species with different habitat preferences, then this could readily impact the results and conclusions of these analyses.

In addition to the community-level approaches above, we also perform species-level analyses of habitat specialization to test neutral theory (John et al. 2007, Phillips et al. 2003). If individual species show preference for specific habitats, then this is evidence that they are not neutral. As above, incorrect lumping or splitting of habitat specialists could readily affect the results of these analyses.

3.2 Materials and Methods

3.2.1 Field Sites and Methods

This study is focused on distinguishing species at a low taxonomic level, 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. We visited 14 sites across Madre de Dios (Fig. 6), separated by a range of spatial distances (from 3 to 250 km apart). At each site, we surveyed communities in the two dominant habitat types found in Madre de Dios, terra firme (upland) forest and floodplain (bottomland) forest. Floodplain forests can vary from swamp forests to successional forests 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 certain early primary successional tree species (e.g. *Ficus insipida*, *Cedrela odorata*, *Cecropia membrenacea*). At three sites, we were unable to survey terra firme forest because it was inaccessible. There were thus 25 total community surveys, 14 in floodplain forest and 11 in terra firme forest.

In community surveys, we censused all *Inga* individuals in a 50 x 50m plot. If there were less than 80 individuals in a plot, we sampled additional individuals in transects until that number was reached. Transects were run in straight lines from the plot and restricted to the same habitat type as the plot (mature floodplain or terra firme forest). In several cases (Blanquillo, CM2, and Maizal floodplain, Boca Manu terra firme), time limitations prevented us from acquiring a sample size of at least 80 individuals. In several communities, additional 25 x 25m plots and/or transects were surveyed and overall sample size of individuals was much greater (particularly at Cocha Cashu and Los Amigos in both floodplain and terra firme). The total number of *Inga* individuals included in community surveys varied from 46 to 282 (mean = 131 +/- 13 S.E.). Excluding low or high sample size communities from our community-level analyses had little effect on the results presented here. Within plots or transects, all *Inga* individuals that reached breast height (1.3m) were measured (for diameter at breast height, DBH, and absolute height), identified, and collected for a genetic specimen and in most cases, an herbarium voucher.

All surveyed individuals were initially identified to morphospecies based on sterile, or vegetative, characters. *Inga* species are not highly variable in trunk and bark characters, and we therefore relied mostly on leaf characters to delimit species. This has the added advantage that herbarium vouchers can later be compared with each other and with specimens already present in herbaria. *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 in various herbaria to determine their taxonomic identity. We were not able to put definitive names on all morphospecies. 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).

For each community, we obtained a soil sample from the 50 x 50m plot by bulking five soil cores taken at random points throughout the plot. For communities where additional samples 25 x 25m plots were surveyed, we collected additional soil samples. All soil samples were sent to the Clemson University Agricultural Services Laboratory for analyses of pH, concentration of multiple base cations (K, Ca, Mg, Zn, Mn, Cu, B, Na), nitrate nitrogen (NO3), phosphorous, cation exchange capacity, and base saturation. Samples were also sent to the North Carolina State Soil Services Laboratory for analysis of particle size distribution (percent sand, silt, and clay).

3.2.2 Laboratory Methods

We selected three sets of individuals for verification of morphological identification via DNA sequencing. First, we selected approximately 100 individuals at random from both dominant habitat types at our two principal field sites, Cocha Cashu Biological Station and Los Amigos Research Center (a total of 442 randomly selected individuals). Second, we sequenced at least two individuals (except for singletons for which we used the only available specimen) of all putative species and unidentified morphospecies that were not covered by the first selection (i.e. species not found at either of the two principal field sites). Third, for most widespread species (*I. alata, alba, auristellae, bourgonii, cayennensis, chartacea, cinnamomea, edulis, marginata, poeppigiana, ruiziana, sapindoides, thibaudiana, umbellifera*), we sequenced individuals from additional sites that spanned the breadth of their distribution within the study area. All individuals selected for DNA sequencing were also vouchered so that any potential mis-identifications or cryptic species could later be assessed morphologically. We also collected and sequenced individuals of the genus *Zygia* (Mimosoideae: Fabaceae), the likely sister genus to *Inga* (Jobson and Luckow 2007), both for use as an outgroup in phylogenetic analyses and to confirm the monophyly of our *Inga* species.

We used a modified CTAB protocol (Doyle and Doyle 1990) to extract DNA from silica gel-dried leaf specimens from all individuals selected for DNA sequencing. We used 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 actually spans multiple intergenic spacers). We selected the ITS region based on its previously demonstrated variability in the genus (Richardson et al. 2001, K.G.D. unpublished results) 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 (RT 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: 1 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 Inc., Valencia, California), 2.5 µL Taq Buffer, 0.5 µL dNTP mix (10 mM concentration for each nucleotide), 1.25 µL primer 1 (10 μM concentration), 1.25 μL primer 2 (10 μM 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). Sequences were assembled using Sequencher v4.5 (Gene Codes Corp., Ann Arbor, Michigan) and manually aligned in MacClade v4.06 (Maddison and Maddison 2003). All sequences used in this study have been deposited in Genbank and the master alignment is deposited in TreeBASE.

3.2.3 Phylogenetic Analyses

We first ran Modeltest v3.7 (Posada and Crandall 1998), and used the AIC criterion to select the best model of sequence evolution separately for the ITS and trnD-T regions. Modeltest gave the GTR + I + G model (general time reversible model with a proportion of sites invariant and a gamma distribution for variable sites) for both regions. Because the same model of evolution fit both genetic regions and preliminary analyses indicated no strong conflict between the loci, we concatenated the regions for further phylogenetic analyses. Once the sequences were concatenated, we used COLLAPSE v1.2 (available from http://darwin.uvigo.es) to reduce the dataset to unique sequences. Thus, individuals with identical ITS and trnD-T sequence were represented by only one placeholder in phylogenetic analyses (note that if individuals differed for either the ITS or trnD-T region they would not be considered to have the same unique sequence under this framework). 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 re-ran Modeltest on the concatenated and collapsed dataset, which confirmed that the GTR + G+ I model remained the best model of sequence evolution. We conducted a maximum likelihood bootstrap analysis with 100 bootstrap replicates in Garli v0.95 (Zwickl 2006). We used the GTR + G + I model, but we allowed Garli to estimate the parameters for base frequencies, rates, proportion of invariant sites, and the gamma distribution for each run. We constructed a 50% majority-rule consensus tree of the 100 bootstrap replicates in PAUP beta v10.4 (Swofford 2002), with the option to show all compatible groupings selected. We optimized branch lengths on this majority-rule consensus tree using the model of sequence evolution specified by Modeltest.

We also conducted a Bayesian phylogenetic analysis of the concatenated dataset, using Mr. Bayes v3.1.2 (Ronquist and Huelsenbeck 2003). We used the GTR + I + G model, but we partitioned the data such that the two genetic regions (ITS and trnD-T) could take on different values for the parameters in the model. We ran two independent runs of four Markov chains with the default heating scheme (Temp=0.2, Swapfreq=1, Nswaps=1), for 10 million generations. Trees were sampled every 1000 generations. The average standard deviation of split frequencies reached 0.055 after 5 million generations and did not drop appreciably after this. We discarded trees from the first 5 million generations of each run as the burn-in. We constructed a majority-rule consensus tree in PAUP beta v10.4 (Swofford 2002) using the remaining trees from both runs. The proportion of trees that possess a given node represents the posterior probability for that node.

In addition to the aforementioned preliminary analyses for each marker, we conducted phylogenetic analyses separately for the ITS and trnD-T regions using the final collapsed dataset. These maximum likelihood and Bayesian analyses followed the same protocols as for the concatenated dataset.

3.2.4 Assessment of Identification Errors and Revising Species Delimitations

There are many methods currently available that can use phylogenetic information to assess errors in species delimitations and identifications (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. 2007). However, we found these methods unsatisfactory for our purposes for several reasons. First, many methods (e.g. Nielsen and Matz 2006, Knowles and Carstens 2007) work best with a limited number of species, while we are working within a diverse system of over 50 species. 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, K.G.D. unpublished data), including among closely related species, which limits the usefulness of allopatry as a criterion for species status.

Lastly, our system may be unsuitable for any species delimitation approach that relies solely on phylogenetic information. *Inga* is a young (2-10 million yrs old; Richardson et al. 2001, Lavin 2006), yet incredibly diverse genus (~300 spp.; Pennington 1997). Given the rapid rate of speciation that this must entail and the slow rate of evolution of the molecular markers we have used, genetic non-monophyly is to be expected both due to incomplete lineage sorting and insufficient time having passed for species to become reciprocally monophyletic. Furthermore, hybridization could obscure patterns of monophyly of species. While one study found little hybridization between *Inga* species in Costa Rican montane forests (Koptur 1984), this has not been assessed in the more speciose lowland Amazon. Given the rapid speciation of *Inga* and the potential for hybridization, morphological information is still critical to successfully delimiting species within the genus.

We used a two-step process of reciprocal illumination to assess errors in the preliminary morphology-based delimitations and identifications. Potential errors were first identified through examination of the generated molecular phylogenies. We then reviewed the morphology of the specimens involved as well as relevant herbarium material to confirm that an actual error had been made. We attempt to classify errors into three categories, which are explained below.

1) Incorrect lumping of two distinct species. If a putative species falls out as two or more monophyletic groups in the phylogeny, this could indicate that multiple species were incorrectly lumped together as one species. In these cases, we re-examined the morphology of the vouchers for those individuals to determine if there were any consistently segregating morphological characters between the phylogenetically divergent groups. If any such characters were found, we considered the two groups to be two separate species. If no such characters were found, we retained the original species delimitation, and concluded that lineage sorting or hybridization could be responsible for the non-monophyly of the species.

2) 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 reexamination of vouchers found no consistent 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 consider them as separate species that may have either speciated recently or historically engaged in limited hybridization.

When the phylogeny placed an unidentified morphospecies as sister to a known species in the phylogeny, we re-examined the morphology of these individuals and of herbarium records of the known species. While reciprocal monophyly of the two putative species indicates that they are probably evolving independently, a review of herbarium records may indicate that the unidentified morphospecies falls within the morphological limits of the known species. In these cases, we labeled the putative species as different evolutionarily significant units (ESUs, Mortiz 1994) of the known species. When we tabulated error rates, we did so twice, both treating ESUs as single species and as separate species.

3) Direct mistakes in individual identifications. The first two categories of error comprise errors in species delimitation, while the last category consists of direct mistakes in identification. This can be revealed when one or more accessions of a given species are placed phylogenetically within another species. In these cases, we consulted herbarium vouchers of both species and determined which species the ambiguous individuals matched better morphologically. If the ambiguous individual better matched the morphology of the species within which it was phylogenetically grouped, then we determined that we had made an error in the original identification of the individual. If the individual's morphology better matched the species to which it was originally assigned, the phylogenetic result could be due to hybridization, incomplete lineage sorting, or some error in the lab (e.g. tube switching).

Once the above errors were taken into account, we were able to revise the identifications of all sequenced individuals. We then had to translate these revisions to the entire ecological dataset of 3,912 individuals. In cases where we split a putatively single species in two, we reviewed the herbarium vouchers of all individuals of the original species. These individuals were then assigned to one of the newly segregated species based on the morphological characters that were found to distinguish them. In cases where we lumped previously segregated species, it was straightforward to assign all individuals of the segregated species to a single species. Incorrect individual identifications could only be determined for sequenced individuals and could not be translated to the entire dataset.

3.2.5 Community-level 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 single species; and 3) revised delimitations treating ESUs as distinct species.

We first examined the distribution of relative abundances of species across the entire dataset. We then used a Kolmogorov test to determine if the distribution differed significantly from a log-normal distribution.

We next used Partial Mantel tests to determine the relative influence of environmental factors and spatially autocorrelated factors on tree community composition. Partial Mantel tests can determine the partial correlation coefficient between a dependent matrix and an independent matrix while controlling for any correlation between the dependent matrix and another independent matrix (analogous to a multiple regression). We constructed three distance matrices representing pairwise distances between communities: 1) species compositional distance (dependent matrix); 2) environmental distance (independent matrix); and 3) geographic distance (independent matrix). We used the Isolation by Distance Web Service v3.15 (Jensen et al. 2005) to conduct the Partial Mantel tests.

The species compositional distance between any two communities can be measured by taking 1 minus the value of a given similarity index calculated between the two communities (most indices vary from 0 to 1, with 1 representing two communities that are maximally similar). We used two indices, the Jaccard Similarity Index, which relies only on presence/absence data and the Bray-Curtis Similarity Index, which incorporates abundance information. Similarity indices were calculated using EstimateS v8.00 (Colwell 2006). Partial Mantel tests were conducted using both compositional distance matrices.

The environmental distance matrix was calculated as follows. We first conducted a principal component analysis in the R Statistical Environment (available at http://www.r-project.org/) of all measured soil variables (22 total) in order to reduce the dimensionality of the data. Principal component axis 1 (PCA1) explains 51% of the variation in the soils data, while the other axes individually explain at most 15% of the variation. PCA1 is correlated with most of the measured soil variables including pH, base cation concentrations, cation exchange capacity, base saturation, phosphorous concentration, and particle size distribution. Many of these variables have been implicated in influencing plant distributions (Tuomisto et al. 2003, Phillips et al. 2003), and we chose this axis to represent environmental variation in Madre de Dios. We calculated the Euclidean distance between each community pair for their values of PCA1, which gives a matrix of environmental dissimilarity, or distance, for all communities. If a community was represented by multiple soil samples, we used the average value of PCA1 for those soil samples in analyses.

We used two sets of geographic distance matrices, one comprised of straight-line geographic distances between communities and one using the natural logarithm of geographic distances.

We conducted partial mantel tests both using the entire dataset and restricting analyses to floodplain or terra firme communities. It may be that species demonstrate preference for a specific habitat type, but that within habitats, spatially autocorrelated processes, such as dispersal limitation, assume greater importance (Zillio and Condit 2007, Jabot et al. 2008). Within habitat types, we also focused on patterns of distance-decay, or the decline with distance, in community similarity (Chave and Leigh 2002, Morlon et al. 2008). The intercept and slope of the relationship between geographic distance and community similarity are often interpreted to reflect the level of dispersal limitation in the landscape (Nekola and White 1999, Morlon et al. 2008). We estimated these parameters using least-squares linear regression. We conducted distance-decay analyses for both the Jaccard and Bray-Curtis Similarity Indices.

3.2.6 Species-level Ecological Analyses

We used a modified version of the approach of Phillips et al. (2003) to determine if individual species showed a significant preference for terra firme or floodplain habitat. We first calculated the relative abundance of a given species across the entire dataset. This value is also the species' expected relative abundance in floodplain and terra firme if the species is a generalist, or neutral with respect to habitat. For species that we suspected to be floodplain specialists, we calculated the binomial probability, $\mathrm{Bin}(Y_{FP} | N_{FP}, \theta)$, 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, the relative abundance of the species across the entire dataset. This is equivalent to assaying if the observed relative abundance of the species in floodplain is significantly different from that expected by chance for a generalist, given our sampling effort. If this probability is less than 0.05, then the data is inconsistent with the species being a habitat generalist, and we classified the species as a floodplain specialist. We performed a similar test for suspected terra firme specialists to determine if they significantly prefer terra firme using $\mathrm{Bin}(Y_{TF} | N_{TF}, \theta)$. If the p-value for a given species was greater than 0.05, the species was classified as a habitat generalist. Note that these one-tailed tests are based on first predicting the direction of a species habitat specialization.

3.3 Results

3.3.1 Phylogenetic Analyses

In total, we obtained DNA sequences from 946 individuals (651 for the ITS nuclear region and 892 for the trnD-T chloroplast region) and from all 64 putative *Inga* species that were found during the course of this ecological study. This represents 24.2% of the 3912 surveyed individuals. The ITS region varied in length from 638 to 668 base pairs and the total aligned dataset was 671 base pairs in length. The trnD-T region varied in length from 1060 to 1110 base pairs and the total aligned dataset was 1167 base pairs in length. The concatenated dataset contained 191 unique sequences, including 7 sequences found in samples of the outgroup *Zygia*. Including individuals that had identical ITS and trnD-T sequences but represented distinct species boosted the dataset for phylogenetic analysis to a total of 222 concatenated sequences.

The consensus tree from the maximum likelihood bootstrap analysis is given in Figure 7. The Bayesian analysis gave a similar topology, and the proportion of Bayesian trees supporting a given node (i.e. the posterior probability for that node) is noted in Figure 7. In general there is low bootstrap and posterior probability support for most nodes in the tree, although many species do fall out as monophyletic groups. When all members of a given species have the same sequence (e.g. *morphospecies 22*), there is not a support value associated with the species, because only one sequence for that species was considered in the analysis. Nevertheless, we can consider these species strongly supported assuming that the sequence is reciprocally monophyletic with respect to other species.

The phylogenetic trees for the single locus phylogenetic analyses, along with maximum likelihood bootstrap support values and Bayesian posterior probabilities, are given in Figures 8 and 9.

3.3.2 Assessment of Identification Errors and Revising Species Delimitations

Many species that were delimited based on vegetative morphology formed reciprocally monophyletic groups in the phylogeny (*Inga alata, alba, cinnamomea, cordatoalata, heterophylla, porcata, psittacorum, setosa, suaveolans, tenuistipula,* and *morphospp. 17, 22,* and *54*). These species (marked with letter A in Fig. 7) did not require further assessment of morphology and were considered well-delimited species. Many other species (*Inga auristellae, densiflora, fosteriana, marginata, nobilis, sapindoides, tomentosa* and *morphospp. 18, 50,* and *75*) contained 1-2 individuals that were mistakenly identified as belonging to other species (these fall into the third category of identification error: direct mistakes in identification, which are highlighted in red in Fig. 7). Once these errors were corrected for, additional species demonstrated reciprocal monophyly in the phylogeny: *Inga umbellifera* and *morphospp. 50, 58,* and *75* (also marked with letter A in Fig. 7).

The sequences for *Inga auristellae* formed two different monophyletic groups in the phylogeny (B in Fig. 7). Upon reexamination of the vouchers associated with these specimens, morphological characters were found that distinguished the two groups (see Table 3 for distinguishing characters). One group better matched herbarium accessions of *I. auristellae*, while the other was found to match the lesser known species *I. brevipes*. Therefore, the group split off *I. auristellae* was assigned the name *I. brevipes*. The sequences for *I. sertulifera* also fell out in two monophyletic groups (C in Fig. 7) that could later be distinguished morphologically (see Table 3). One of these groups matched herbarium accessions of *I. sertulifera* well, while the other group could not be matched up with any known species and was given the name *Inga morphosp. 55*.

Two species, *Inga marginata* and *punctata*, were paraphyletic with respect to other species (D in Fig. 7). This is not unexpected if these species recently diverged from or gave rise to other species. Because these species form cohesive morphological entities, distinct from their close relatives, the original identifications were determined to be valid. Several pairs of species, *I. bourgonii* and *coruscans*, *I. acreana* and *chartacea*, and *I. lineata* and *killipiana*, either share sequences or are otherwise phylogenetically mixed within monophyletic clades (E in Fig. 7). Because these species retain several segregating morphological characters, we retained the original identifications and attributed the phylogenetic result to recent speciation or hybridization.

In several other pairs of species, *Inga ruiziana* and *morphosp. 68*, *I. tomentosa* and *morphosp. 71*, and *I. stenoptera* and *morphosp. 76*, the members of the pair fell out as sister to each other in the phylogeny (F in Fig. 7). Upon extensive review of herbarium vouchers of the named species, it was determined that the morphospecies did not possess sufficient distinguishing characters to be separated as distinct species. The two sister groups were therefore labeled as different evolutionarily significant units (ESUs). *Inga laurina* was found to comprise two sister groups that upon reexamination, were found to differ slightly morphologically (see Table 3). 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 distinguishable morphologically and were not separated into distinct ESUs.

In multiple cases where species were mixed together phylogenetically, morphological review of herbarium vouchers showed that the species should be lumped together as single species (G in Fig. 7). This was the case for *Inga leiocalycina* and *morphosp. 79*, *I. morphospp. 18* and *62*, *I. morphospp. 49* and *69*, and *I. densiflora* and *morphosp. 47*. In the latter two cases, the newly delimited species form a paraphyletic grade with respect to other species, but as discussed above, paraphyly is not unexpected given recent speciation. *Inga spectabilis* and *morphosp. 78* were both nested phylogenetically within *I. venusta*, and based on similar morphology, we lumped these species with *I. venusta*. As the putative *I. spectabilis* was represented by only one vegetative accession, it would 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 evolutionarily significant units as defined above (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 ESU equivalents.

There is little genetic divergence between species in a large clade (H in Fig. 7) that includes *Inga cayennensis, edulis, ingoides, longipes, poeppigiana*, *sapindoides, steinbachii, stipulacea* (or part thereof, see below), *thibaudiana*, *vera*, and *morphospp. 48* and *63*. Several species (*cayennensis, edulis, ingoides*, and *thibaudiana*) while not displaying reciprocal monophyly, at least from genetically cohesive groups within this clade. There are also multiple species in this clade that were only collected once. Two of these singletons (*I. morphosp. 48* and *63*) were nested within *I. sapindoides*, and were lumped with that species upon review of herbarium vouchers. The other three singletons (*I. longipes, macrophylla*, and *vera*) were sufficiently morphologically distinct to be maintained as unique entities. Several specimens of *I. sapindoides* are divergent from the main clade of the species. Upon review of herbarium vouchers it was determined that this divergent group was distinct morphologically and actually represented the little known species *I. fosteriana* (see Table 3 for distinguishing characters). Likewise, a morphologically and genetically divergent clade of *I. poeppigiana* was found to represent *I. barbata* (see Table 3). Both *I. longipes* and *steinbachii* also fell out in two divergent locations within the clade, but this was determined to be due to initial misidentifications.

3.3.2.1 Problematic Species

In several species (*I. capitata, stipulacea, tocacheana*, and *morphosp. 56*), high intraspecific divergence at the ITS marker, but not the trnD-T marker (see Figs 8 & 9), drove the species to fall out as divergent groups in the combined phylogeny (Fig. 7). In order to determine if this could be due to amplification of ITS pseudogenes (Alvarez and Wendel 2003), we assayed the divergent ITS sequences for base composition (% GC), number of substitutions in the 5.8S region, and folding energy. The divergent sequences did not differ for these metrics from the rest of our ITS dataset (K.G.D., unpublished data), which suggests that they do not represent pseudogenes. They may instead represent incomplete lineage sorting, which should be more common for the ITS marker than the trnD-T marker, given its larger effective population size.

Inga capitata appears as three divergent groups in the phylogeny (J in Fig. 7). All three of these groups are divergent from each other for the ITS marker, while one is divergent from the other two for the trnD-T marker. The group that is divergent at the trnD-T marker (*capitata* LA.LP_2) is also distinct morphologically, and we have therefore labeled it as a separate ESU (see Table 3 for distinguishing characters).

Inga morphospp. 56 and *66* (K in Fig. 7) are represented by only 3 individuals in the analysis (only 3 individuals were found in the field). All three individual share the same trnD-T allele (Fig. 9) and are very similar morphologically. For this reason, we have lumped the species together. However, one individual of the former *I. morphosp. 66* is divergent for the ITS marker (Fig. 8).

The two surveyed individuals of *Inga tocacheana* (M in Fig. 2) have divergent ITS alleles, but identical trnD-T alleles (Figs 8 and 9). These two individuals are very similar morphologically and distinct from all other species, and we have retained the original species delimitation. Likewise, *I. stipulacea* appears in two very divergent locations in the phylogeny (N in Figure 7) due solely to high within-species divergence at the ITS marker (Fig. 8). As with *I. tocacheana*, all *I. stipulacea* accessions are very similar morphologically and distinct from all other species, and we retain the original species delimitation. For all of the above-described problematic species, further sequencing effort and morphological examination is needed to determine the exact limits of species and ESUs.

3.3.2.2 Summary of Error Assessment and Revising Species Delimitations

The total number and proportion of different types of delimitation and identification errors are given in Table 4. We used Table 3 and the protocol outlined in Materials and Methods to revise the species delimitations of all individuals, both sequenced and unsequenced, in the ecological dataset.

3.3.3 Community-Level Ecological Analyses

We conducted all ecological analyses with the original species delimitations and with the revised delimitations, both ignoring ESUs and treating ESUs as distinct species.

For the purposes of brevity, we focus on contrasting the results using the original delimitations with revised delimitations treating ESUs as distinct species. The results in which ESU distinctions are ignored are generally intermediate between these two scenarios and are not presented here.

Using the original delimitations, there was a large number of rare species (Fig. 10a) and the distribution of relative abundances differed significantly from a log-normal distribution (Kolmogorov test, D=0.129, p=0.010). Using the revised delimitations, there were fewer rare species (Fig. 10b), and the distribution of relative abundances did not differ significantly from a log-normal distribution (Kolmogorov test, D=0.080, p=0.150).

The results of the Partial Mantel tests are given in Tables 5, 6, and 7. These tests were used to evaluate the correlation of environmental and geographic distance with community similarity (for purposes of Mantel tests, correlation was actually assessed in terms of dissimilarity, 1 – given similarity index). When all communities are included in the analysis (Table 5), the correlation between geographic distance and community similarity increased substantially with the revised delimitations, particularly for the Jaccard Index. There was little difference in Mantel test results between the original and revised species delimitations when analyses were restricted to floodplain plots (Table 6). In contrast, when analyses were restricted to terra firme plots (Table 7), the correlation between geographic distance and community similarity increased considerably using the revised delimitations. For the Bray-Curtis Index, the correlation between environmental distance and community similarity also increased (Table 7).

While the revised delimitations did alter some aspects of the Mantel test results, other aspects remained the same. For both indices and both delimitations, community similarity showed a greater correlation with environmental distance than geographic distance when all communities were considered in the analysis. However, when analyses were restricted to one habitat type, community similarity was more highly correlated with geographic distance. Also, under both delimitations, community similarity as measured by the Jaccard Index showed greater correlation with geographic distance and lower correlation with environmental distance than similarity as measured by the Bray-Curtis Index.

Using the original identifications, the slope of the distance-decay relationship for the Jaccard Index differed between floodplain and terra firme communities (Fig. 11, Table 8), but using the revised delimitations, the slopes converged on the same value. For the Bray-Curtis Index, the revised species delimitations did not substantially change the distance-decay relationship in either floodplain or terra firme (Fig. 12, Table 8). The confidence intervals for the slope (Table 8) and intercept (results not shown) of the distance-decay relationship overlap broadly for all analyses, and all analyses agree in showing a significant decline with distance in the similarity of communities.

3.3.3 Species-Level Ecological Analyses

Results of species-level ecological analyses often differed considerably between the original and revised species delimitations (Tables 9 and 10). In 3 of the 6 cases where species or ESUs were split based on the reciprocal illumination procedure, the newly segregated species was found to have a different habitat preference than the species with which it was originally lumped (e.g. floodplain versus terra firme or floodplain versus generalist). In 7 out of 10 cases where a species was lumped with another species based on the reciprocal illumination procedure, the originally segregated species had a different habitat preference than the species with which it is now lumped. Using either species delimitation, the majority of species were found to be habitat specialists, but the proportion of specialists was higher under the revised delimitations (62% versus 53% under the original delimitations).

3.4 Discussion

This study represents the first large-scale assessment of the accuracy of vegetative-morphology based delimitation and identification of tropical tree species. We sequenced and constructed a phylogeny for one quarter of nearly 4,000 individuals across 64 putative species that were encountered in our conventional ecological study of tropical trees. This type of assessment is important, because much of our current understanding of ecology comes from studies of tropical tree communities. We found that we did make delimitation and identification mistakes; nearly 8% of stems were misidentified in some manner (Table 4).

These errors had a pronounced effect on community-level ecological analyses that are used to assess neutral theory (Tables 5-8, Figs 10-12). For example, under the original species delimitations, the distribution of relative abundances differed significantly from a log-normal distribution, while under the revised delimitations, it did not. A log-normal distribution of relative abundances has been interpreted as evidence against neutral theory (McGill 2003, McGill et al. 2006). The results of community similarity analyses also changed when the revised delimitations were used, although, in this case, the new results leant more support to neutral theory. Community similarity showed a higher correlation with geographic distance between communities, and such correlations have been interpreted as evidence for the importance of neutral processes, namely dispersal limitation, in structuring communities (e.g. Hubbell 2001, Tuomisto et al. 2003).

Incorrect identifications also affected species-level assessments of neutral theory. Many species that were split or lumped based on the revised delimitations had contrasting habitat preference classifications (Tables 9 and 10), and overall, a lower proportion of species were found to be neutral with respect to habitat preference using the revised delimitations.

3.4.1 Using DNA sequence data to aid in the delimitation of *Inga* species

Our phylogenetic analyses and subsequent review of morphological characteristics (the reciprocal illumination process) led us to combine twelve species with previously known species and to split four species (Table 4). In total, the revised species delimitations encompass 55 species and 60 evolutionarily significant units (ESUs), while the original dataset comprised 63 species and 64 ESUs. Most of the species that were lumped with other species were rare, often only collected once or twice. Given so few specimens, it was often difficult to determine whether the species contained segregating morphological characters that distinguished them from the species within which they were phylogenetically placed. If our review of herbarium specimens suggested that these individuals could be considered morphological variants of the species within which they were phylogenetically placed, they were lumped. Species that were split based on the reciprocal illumination process were generally collected more frequently, which allowed us to readily determine if there were consistent morphological characters that distinguished the two genetically distinct groups.

We classified the various errors in delimitation and identification into three categories: incorrect lumping, incorrect splitting, and straight misidentification (Table 4). The first two categories comprise errors in species delimitation while the last category, which involves actual mistakes in identification, includes only 1.7% of all stems. This is higher than the 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). The methods used on BCI accounted for misidentification, but not species delimitation, errors.

We assessed errors in species delimitation in addition to misidentifications and found a total error rate of 6.8%, or 7.8% if ESUs are treated as distinct entities (Table 4). 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). While our results only represent one study, we can state preliminarily that tropical tree ecology studies do not seem to be subject to exceptionally high rates of error in delimitation and identification despite the higher diversity and perhaps greater potential for taxonomic confusion.

Our process of reciprocal illumination utilized information from DNA sequences and vegetative morphology. While we recognize that ecological information can be very useful for delimiting species (sensu Raxworthy et al. 2007, Rissler and Apodaca 2007), we avoided this because we were interested testing the relative importance of neutral versus ecological factors in structuring communities.

3.4.2 Effects of Errors on Ecological Analyses

Systematic errors in species delimitation and identification have the potential to strongly influence the results and conclusions of ecological analyses. For example, our reciprocal illumination process suggested that we systematically and incorrectly distinguished rare morphological variants of common species as distinct species. When these rare variants were lumped with more common species, it significantly changed the form of the distribution of relative abundances (Fig. 10), which in turn affected our interpretation of that distribution.

Systematic errors also have great potential to affect analyses of community similarity. In many of the cases where we split a single species into multiple species or ESUs based on the reciprocal illumination process, the newly segregated taxa were geographically disjunct. This led us to underestimate the correlation between geographic distance and community similarity (Table 5). Our community similarity analyses restricted to single habitat types suggest that these systematic errors were largely confined to terra firme communities (contrast Tables 6 and 7).

This systematic error in lumping of geographically segregated species also lead to a substantial change in the distance-decay relationship for terra firme communities. Once these errors were corrected, the slope of the distance-decay relationship for terra firme communities converged to the value found for floodplain communities (Fig. 11, Table 8). This is interesting, although the exact biological significance of this pattern is uncertain. Overall though, distance-decay analyses were not strongly affected by systematic identification and delimitation errors. The confidence intervals for the parameters of the distance-decay relationship (slope and intercept) overlapped broadly for the original and revised species delimitations (Table 8), and all analyses showed significant distance-decay in community similarity.

Systematic errors can also affect species-level analyses. Incorrect lumping of species with different habitat preferences would lead to underestimation of the proportion of species that are habitat specialists. Incorrect splitting of species with different habitat preferences would have the opposite effect. We found evidence for both types of errors in our analyses (Tables 9 and 10). For example, *Inga laurina*, which was originally classified as a generalist, was found to comprise two ESUs with contrasting habitat preferences. An opposite example is found in *I. morphosp. 79*. This species was designated as preferring floodplain habitat, but was lumped with *I. leiocalycina*, which prefers terra firme habitat. However, these two types of errors balanced each other out. The difference between the original and revised delimitations in the proportion of species that were found to be habitat specialists is due to a different type of systematic error. As discussed above, we systematically and incorrectly distinguished rare morphological variants of known species as distinct species. Due to low sample size, statistical power was lacking to definitely assign these rare species a habitat preference and they were classified as generalists. However, they were lumped with more common species that had larger samples sizes and generally showed a distinct habitat preference, thus reducing the total proportion of habitat generalists in the revised dataset.

3.4.3 General Conclusions of Ecological Analyses

When communities in floodplain and terra firme were considered together, similarity in species composition was more related to environmental similarity than to geographic distance between communities (Table 5). However, when analyses are restricted to a single habitat type, community similarity was more related to geographic distance (Tables 6 and 7). These patterns hold across both similarity indices and using either the revised or original species delimitations. All analyses also demonstrate significant distance-decay in community similarity within habitat types. Thus, habitat type may be the most important factor in determining the composition of these communities, but analyses within habitat types reveal that neutral processes, such as dispersal limitation, may also play a significant role in community assembly.

Both delimitations also agree that a community similarity index that incorporates species abundance information (the Bray-Curtis Similarity Index) shows a higher correlation with environmental variables and lower correlation with geographic distance than an index that utilizes only presence/absence information (the Jaccard Similarity Index). This could indicate that spatially autocorrelated processes, namely dispersal limitation, may be responsible for influencing community composition, but that species' abundances are more determined by environmental conditions.

Overall, our conclusions are in line with the results of other studies of tropical tree communities (e.g. Phillips et al. 2003, Tuomisto et al. 2003, John et al. 2007, Norden et al. 2007, Queenborough et al. 2007) and plant ecological studies in general (Tilman 1994, Gurevitch et al. 2006).

3.4.4 Implications for Species Discovery

Through the process of this ecological research and subsequent verification of identifications, we have documented 55 Inga species in the department of Madre de Dios, Peru. A monograph of the genus published in 1997 (Pennington 1997) lists 40 species for Madre de Dios. We found 26 of those 40 species. Eighteen of the species we found are new records for Madre de Dios, and 4 of those are new records for the country of Peru. Additionally, we found 10 morphospecies that we have been unable match with any named species even after surveying specimens across multiple herbaria; most of these morphospecies are genetically distinct from any of the named species that we have sequenced (Fig. 7). All of these species are potentially new to science. However, botanical convention generally dictates that a species must be represented by fertile vouchers to be officially described. Only one of our morphospecies was found in a fertile state, and this species will be described (Dexter and Pennington, unpublished manuscript). The rest of the morphospecies will remain in a state of limbo until they are found in a fertile state, at which time they may be assigned to previously described species or described as novel.

Ecologists, particularly tropical ecologists, frequently come across individual organisms for which they are not able to definitively assign a species name. While many of these individuals probably belong to already-described species, many of them may also represent new species. In this study, we conducted sufficient genetic analyses and taxonomic work in herbaria to be fairly confident that we have uncovered new tree species. Given this groundwork as a prerequisite, ecologists stand to contribute substantially to species discovery. In a time when taxonomic funding and expertise are on the decline, researchers other than taxonomists and systematists may need to play an active role in species discovery if we are to document existing biodiversity before it is gone. This is particularly critical in tropical ecosystems, which face grave threats and where most undiscovered diversity probably lies (e.g. Hebert et al. 2004a, Janzen et al. 2005).

3.4.5 Conclusions

We have shown that tropical tree ecologists frequently make mistakes in species delimitation and identification of individual trees. However, these mistakes are not sufficient to overturn our current understanding of the forces controlling tropical tree community composition. Both deterministic ecological factors and stochastic, "neutral" factors (e.g. dispersal limitation) seem to play a significant role in structuring tree communities, although ecological factors predominate. Nevertheless, our results do show that ecologists must take care in detailed studies of individual species to ensure that the species under study is in fact the same species across all sampled individuals and study sites. Carelessness in this regard could lead to erroneous conclusions about habitat specializations of individual species or any other potentially studied topics.

Our genetic analyses have shown that many closely related species are mixed together phylogenetically and even sometimes share alleles. Any method to delimit species that relies solely on our genetic data would not be able to distinguish these close relatives. We use the same genetic markers (chloroplast intergenic spacers and ITS) as most plant biologists, which suggests that these markers will not be adequate, on their own, to delimit species in diverse tropical genera. We instead used a reciprocal illumination procedure that incorporates genetic and morphological information. 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 the remaining biodiversity on the planet. Ecologists, whom often make extensive collections in remote, rarely visited locations, are well poised to take up this approach and contribute substantially and importantly to the discovery of new species.

Table 3: Morphological Characters of Split Species

The following species were split into two species or ESUs based on the reciprocal illumination procedure. Morphological characters are given that originally united the two species as well as characters that distinguish the newly segregated species.

Characters used to define original species	**Distinguishing characters of newly segregated species or ESU**
auristellae	*brevipes*
2-3 pairs leaflets	Stipules persistent
Proximal leaflets basal (short petiole)	Larger leaflets (3-7x6-15cm vs. 2-5x5-10cm)
Winged rachis flares distally	
Short, orange pubescence on rachis	Pubescence also on mid-rib of leaflets
Narrow stipules	
poeppigiana	*barbata*
3-4 pairs leaflets	Broader, oblanceolate leaflets
Relatively small leaflets (<13cm in length)	Shorter, narrower stipule (1mm long, <0.5mm wide)
Winged rachis	Brochidodromous venation
Long hispid pubescence (>1.5mm)	
sapindoides	*fosteriana*
≥3 pairs leaflets	Denser, more tomentose pubescence

Large leaflets (often >25cm in length) Winged rachis Orange-red pubescence Regular cup-shaped extra-floral nectary	Usually ≥4 pairs leaflets Larger, spadiform stipule (1.5-2x1-1.5cm)
sertulifera	*morphosp. 55*
2 pairs leaflets (6-10x4-7cm) Rounded, elliptical leaflets Glabrous Narrow stipules	Narrow, apressed wing on rachis Reticulate tertiary venation
capitata	*capitata ESU2*
2-3 prs coriaceous leaflets glabrous reticulate tertiary venation persistent stipules	narrower stipule (<3mm wide) smaller, more elliptic leaflets (3-7x7-15cm vs. 4-10x10-25cm)
laurina	*laurina ESU2*
2 pairs leaflets (10-16x6-12cm) Rounded, elliptical leaflets Glabrous Sparse, tertiary venation	Persistent, short stipules (<1mm)

Table 4: Error Rates in Species Delimitation and Identification

Frequency of different delimitation and identification errors across an ecological study of trees in Amazonian Peru, as assessed through a reciprocal illumination procedure utilizing phylogenetic and morphological data. The number of species or individuals with a given error as well as the percentage of the total (in parentheses) is reported.

Category	With ESU distinctions ignored		With ESUs treated as distinct species	
	No. Species	Total No. Sequenced Individuals	No. Species	Total No. Sequenced Individuals
Incorrectly Lumped	4 (6.3%)	17 (1.8%)	6 (9.4%)	24 (2.6%)
Incorrectly Split	12 (19.0%)	31 (3.3%)	10 (15.6%)	34 (3.6%)
Bad Identification	10 (15.9%)	16 (1.7%)	10 (15.6%)	16 (1.7%)
Total with Errors	24 (38.1%)	64 (6.8%)	24 (37.5%)	74 (7.8%)
Total	63	946	64	946

Table 5: Results of Partial Mantel Tests for All Communities

Results of partial mantel tests (for all communities) to determine the correlation between a community dissimilarity matrix (1-similarity index) and environmental and geographic distance matrices.

Similarity Index	Independent Variable*	Species Delimitation	Mantel Correlation Coefficient	p-value
Jaccard	Environment	Original	0.6828	<0.001
		Revised	0.6573	<0.001
	Geography	Original	0.1379	<0.001
		Revised	0.3189	<0.001
Bray-Curtis	Environment	Original	0.7303	<0.001
		Revised	0.6973	<0.001
	Geography	Original	0.0978	0.060
		Revised	0.1693	0.009

*Environment = Euclidean distance between two communities for value of axis 1 of a principal component analysis that summarizes soil environment variation (PCA 1 explains 51% of the variation in the soils data); Geography = strait-line geographic distance between communities.

Table 6: Results of Partial Mantel Tests for Floodplain Communities

The results of partial mantel tests for floodplain communities to determine the correlation between a community dissimilarity matrix (1-similarity index) and environmental and geographic distance matrices.

Similarity Index	Explanatory Variable*	Species Delimitation	Mantel Correlation Coefficient	p-value
Jaccard	Environment	Original	0.2468	0.046
		Revised	0.2469	0.046
	Geography	Original	0.4017	<0.001
		Revised	0.4089	<0.001
Bray-Curtis	Environment	Original	0.2203	0.030
		Revised	0.2144	0.048
	Geography	Original	0.2694	0.076
		Revised	0.2275	0.066

*Environment = Euclidean distance between two communities for value of axis 1 of a principal component analysis that summarizes soil environment variation (PCA 1 explains 51% of the variation in the soils data); Geography = strait-line geographic distance between communities.

Table 7: Results of Partial Mantel Tests for Terra Firme Communities

The results of partial mantel tests for terra firme communities to determine the correlation between a community dissimilarity matrix (1-similarity index) and environmental and geographic distance matrices.

Similarity Index	Explanatory Variable*	Species Delimitation	Mantel Correlation Coefficient	p-value
Jaccard	Environment	Original	0.3515	0.004
		Revised	0.3453	0.012
	Geography	Original	0.2790	0.044
		Revised	0.5717	0.002
Bray-Curtis	Environment	Original	0.3262	0.025
		Revised	0.3718	0.015
	Geography	Original	0.4090	0.008
		Revised	0.4731	0.004

*Environment = Euclidean distance between two communities for value of axis 1 of a principal component analysis that summarizes soil environment variation (PCA 1 explains 51% of the variation in the soils data); Geography = strait-line geographic distance between communities.

Table 8: Slope and Confidence Intervals from Distance-Decay Analyses

Estimates of the rate, or slope, of distance-decay (how community similarity changes with geographic distance) for the two principal habitat types in Madre de Dios, for two different similarity indices, and for the original and revised species delimitations.

Habitat	Similarity Index	Species Delimitation	Slope	Lower 95% CI	Upper 95% CI
Floodplain	Jaccard	Original	-0.00203	-0.00239	-0.00167
		Revised	-0.00191	-0.00225	-0.00158
	Bray-Curtis	Original	-0.00182	-0.00214	-0.00147
		Revised	-0.00180	-0.00216	-0.00145
Terra firme	Jaccard	Original	-0.00153	-0.00193	-0.00112
		Revised	-0.00192	-0.00235	-0.00148
	Bray-Curtis	Original	-0.00196	-0.00245	-0.00147
		Revised	-0.00209	-0.00260	-0.00158

Table 9: Species Habitat Preferences Under the Original Species Delimitations

Habitat specialization of *Inga* species in Madre de Dios as per the original morphology-based species delimitations.

Species	Rel. Abundance Floodplain	Rel. Abundance Terra Firme	p-value	Classification Result
acreana	0.127	0.022	<0.001	floodplain
alata	0.024	0.028	0.249	generalist
alba	0	0.043	<0.001	terra firme
auristellae	0.006	0.151	<0.001	terra firme
bourgonii	0.111	0.023	<0.001	floodplain
capitata	0.061	0.072	0.158	generalist
cayennensis	0.002	0.021	<0.001	terra firme
chartacea	0.013	0.015	0.292	generalist
cinnamomea	0.029	0.002	<0.001	floodplain
cordatoalata	0	0.009	0.003	terra firme
coruscans	0.018	0	<0.001	floodplain
densiflora	0.009	0.014	0.118	generalist
edulis	0.034	0.030	0.301	generalist
heterophylla	0	0.005	0.015	terra firme
ingoides	0.004	0.001	0.054	generalist
killipiana	0	0.001	0.081	generalist

laurina	0.042	0.029	0.069	generalist
leiocalycina	0.003	0.062	<0.001	terra firme
lineata	0	0.001	0.081	generalist
longipes	0	0.001	0.069	generalist
macrophylla	0.001	0	0.099	generalist
marginata	0.038	0.002	<0.001	floodplain
nobilis ssp. nobilis	0.027	0.013	0.021	floodplain
nobilis ssp. quaternata	0.081	0.005	<0.001	floodplain
poeppigiana	0.016	0.029	0.025	terra firme
porcata	0.011	0	0.002	floodplain
psittacorum	0.001	0	0.099	generalist
punctata	0.020	0.006	0.007	floodplain
ruiziana	0.080	0.001	<0.001	floodplain
sapindoides	0.071	0.013	<0.001	floodplain
sertulifera	0.010	0.019	0.039	terra firme
setosa	0.007	0	0.009	floodplain
spectabilis	0.001	0	0.092	generalist
steinbachii	0.014	0.002	0.003	floodplain
stenoptera	0.001	0	0.092	generalist

stipulacea	0.005	0.003	0.234	generalist
suaveolans	0.045	0.004	<0.001	floodplain
tenuistipula	0.004	0.006	0.232	generalist
thibaudiana	0.001	0.094	<0.001	terra firme
tocacheana	0	0.001	0.069	generalist
tomentosa	0.027	0.001	<0.001	floodplain
umbellifera	0.006	0.053	<0.001	terra firme
venusta	0.007	0.032	<0.001	terra firme
vera	0.001	0	0.099	generalist
morphosp. 17	0.003	0.056	<0.001	terra firme
morphosp. 18	0	0.003	0.033	terra firme
morphosp. 22	0	0.013	<0.001	terra firme
morphosp. 47	0.001	0.002	0.089	generalist
morphosp. 48	0.001	0	0.099	generalist
morphosp. 49	0.001	0	0.099	generalist
morphosp. 50	0.002	0	0.064	generalist
morphosp. 54	0.008	0.006	0.225	generalist
morphosp. 56	0	0.001	0.081	generalist
morphosp. 58	0.002	0.003	0.195	generalist
morphosp. 62	0	0.001	0.069	generalist
morphosp. 63	0.001	0	0.099	generalist

morphosp. 66	0.001	0	0.092	generalist
morphosp. 68	0.001	0	0.099	generalist
morphosp. 69	0.004	0	0.036	floodplain
morphosp. 71	0	0.014	<0.001	terra firme
morphosp. 75	0.006	0.089	<0.001	terra firme
morphosp. 76	0.001	0	0.099	generalist
morphosp. 78	0.004	0	0.036	floodplain
morphosp. 79	0.013	0	0.001	floodplain

*This is the p-value for a binomial test that the species significantly prefers the habitat in which it has a greater relative abundance (see text for details).

Table 10: Species Habitat Preferences Under the Revised Species Delimitations

Habitat specialization of *Inga* species in Madre de Dios as per the species delimitations revised based on the reciprocal illumination procedure.

Species	Rel. Abundance Floodplain	Rel. Abundance Terra Firme	p-value	Classification Result
acreana	0.127	0.022	<0.001	floodplain
alata	0.023	0.028	0.228	generalist
alba	0	0.043	<0.001	terra firme
auristellae	0.006	0.139	<0.001	terra firme
barbata	0.015	0.003	0.005	floodplain
bourgonii	0.111	0.023	<0.001	floodplain
brevipes	0	0.013	<0.001	terra firme
capitata ESU1	0.051	0.057	0.243	generalist
capitata ESU2	0.010	0.014	0.144	generalist
cayennensis	0.002	0.021	<0.001	terra firme
chartacea	0.013	0.015	0.260	generalist
cinnamomea	0.029	0.002	<0.001	floodplain
cordatoalata	0	0.009	0.003	terra firme
coruscans	0.018	0	<0.001	floodplain

densiflora	0.009	0.017	0.061	generalist
edulis	0.034	0.030	0.302	generalist
fosteriana	0.003	0	0.043	floodplain
heterophylla	0	0.005	0.015	terra firme
ingoides	0.004	0.001	0.054	generalist
killipiana	0	0.001	0.081	generalist
laurina ESU1	0.042	0.024	0.015	floodplain
laurina ESU2	0	0.006	0.009	terra firme
leiocalycina	0.015	0.063	<0.001	terra firme
lineata	0	0.001	0.081	generalist
longipes	0	0.001	0.070	generalist
macrophylla	0.001	0	0.100	generalist
marginata	0.040	0.002	<0.001	floodplain
nobilis	0.110	0.018	<0.001	floodplain
poeppigiana	0.001	0.026	<0.001	terra firme
porcata	0.011	0	0.002	floodplain
psittacorum	0.001	0	0.100	generalist
punctata	0.020	0.006	0.007	floodplain
ruiziana ESU1	0.079	0.001	<0.001	floodplain

ruiziana ESU2	0.001	0	0.100	generalist
sapindoides	0.070	0.012	<0.001	floodplain
sertulifera	0.002	0.019	<0.001	terra firme
setosa	0.007	0	0.009	floodplain
steinbachii	0.014	0.002	0.004	floodplain
stenoptera ESU1	0.001	0	0.092	generalist
stenoptera ESU2	0.001	0	0.100	generalist
stipulacea	0.005	0.003	0.235	generalist
suaveolans	0.045	0.004	<0.001	floodplain
tenuistipula	0.004	0.006	0.232	generalist
thibaudiana	0.001	0.094	<0.001	terra firme
tocacheana	0	0.001	0.070	generalist
tomentosa ESU1	0.028	0.001	<0.001	floodplain
tomentosa ESU2	0	0.014	<0.001	terra firme
umbellifera	0.006	0.052	<0.001	terra firme
venusta	0.012	0.032	0.001	terra firme

vera	0.001	0	0.100	generalist
morphosp. 17	0.003	0.056	<0.001	terra firme
morphosp. 18	0	0.004	0.019	terra firme
morphosp. 22	0	0.013	<0.001	terra firme
morphosp. 49	0.004	0	0.029	floodplain
morphosp. 50	0.002	0.002	0.315	generalist
morphosp. 54	0.008	0.006	0.225	generalist
morphosp. 55	0.006	0	0.014	floodplain
morphosp. 56	0.001	0	0.092	generalist
morphosp. 58	0.002	0.002	0.315	generalist
morphosp. 75	0.006	0.089	<0.001	terra firme

*This is the p-value for a binomial test that the species significantly prefers the habitat in which it has a greater relative abundance (see text for details).

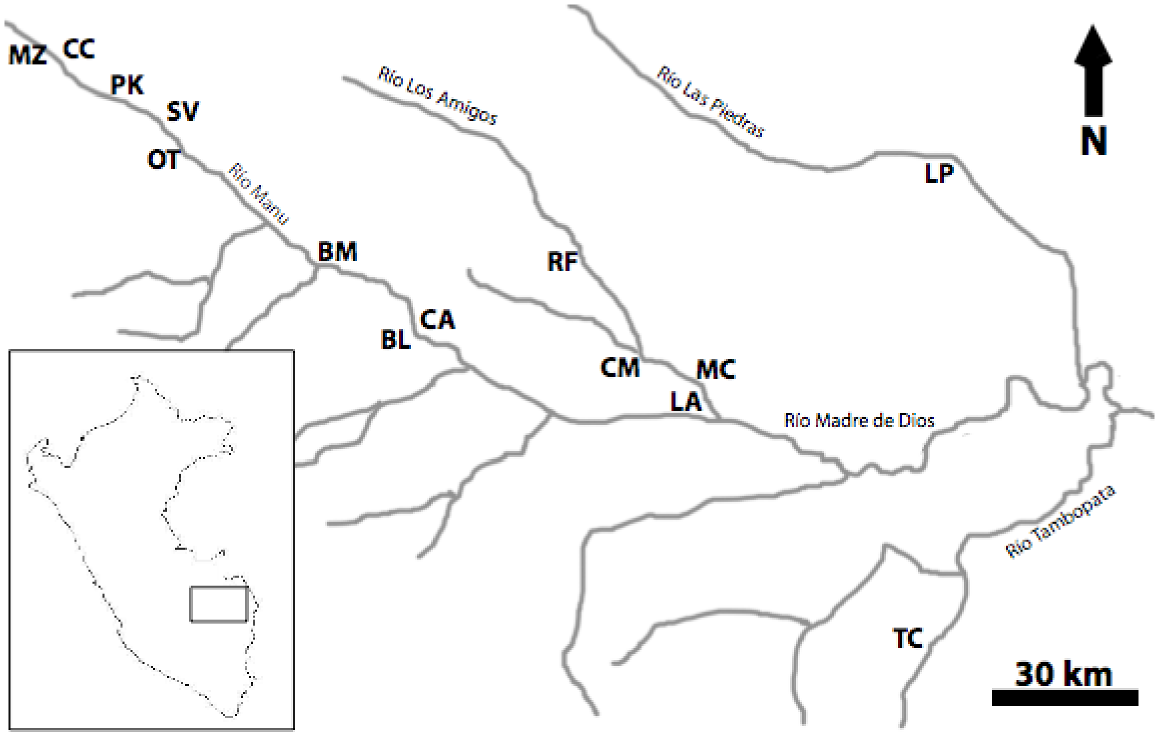
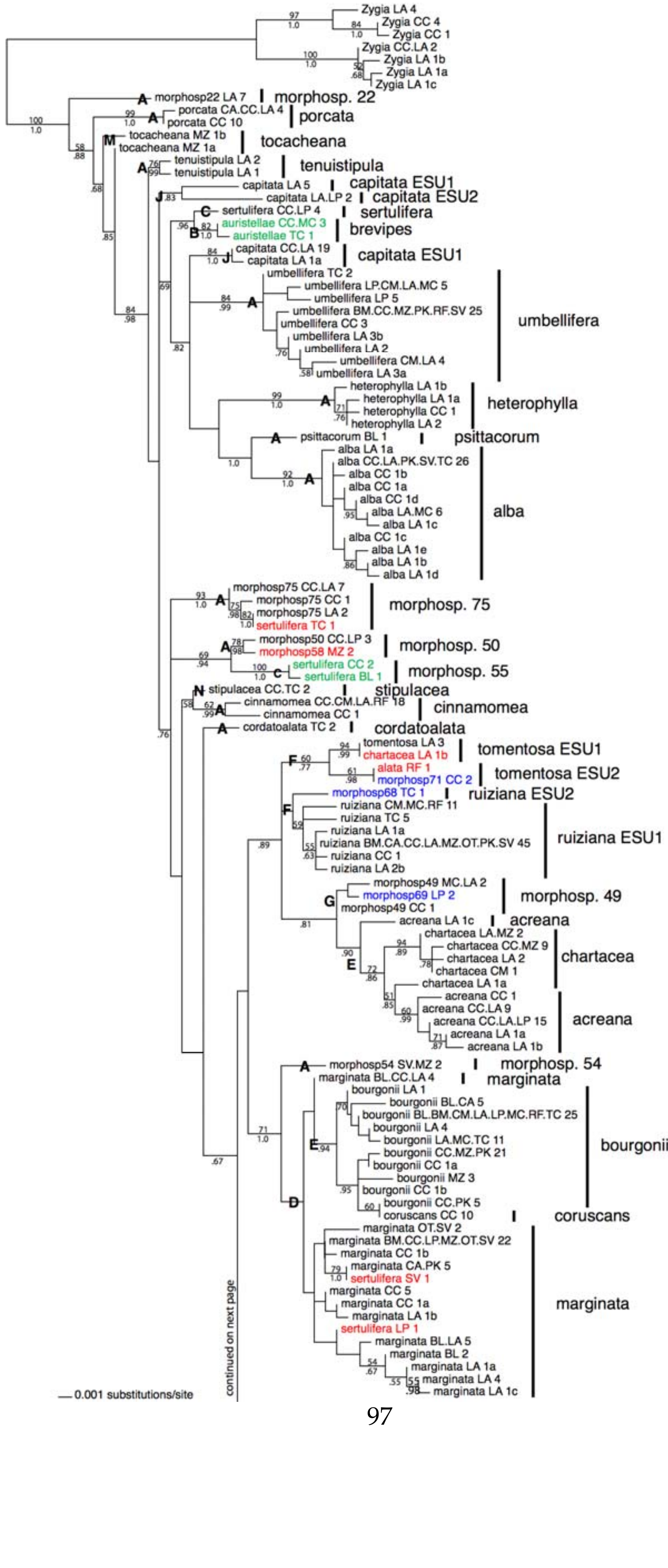
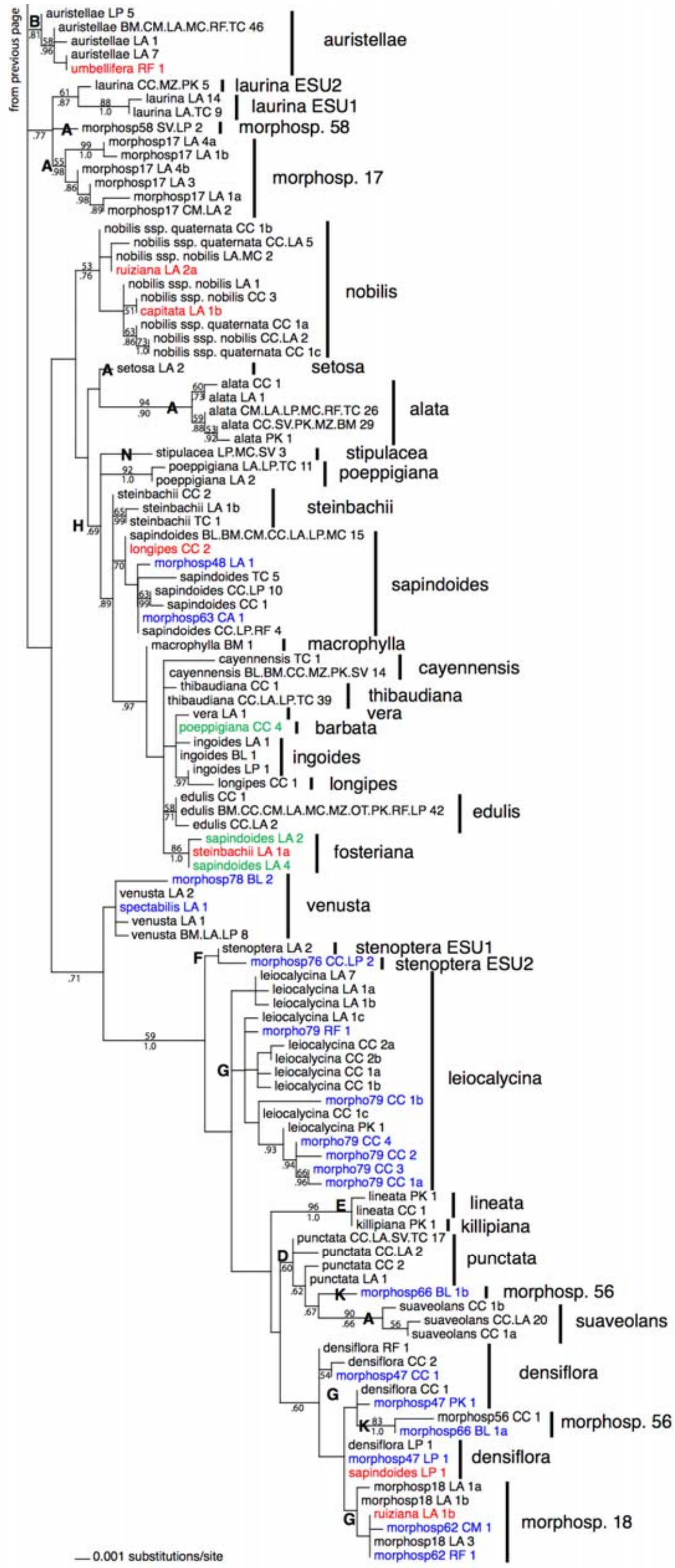


Figure 6: Map of Study Area

Map of the study area in Madre de Dios, Peru. The location 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.



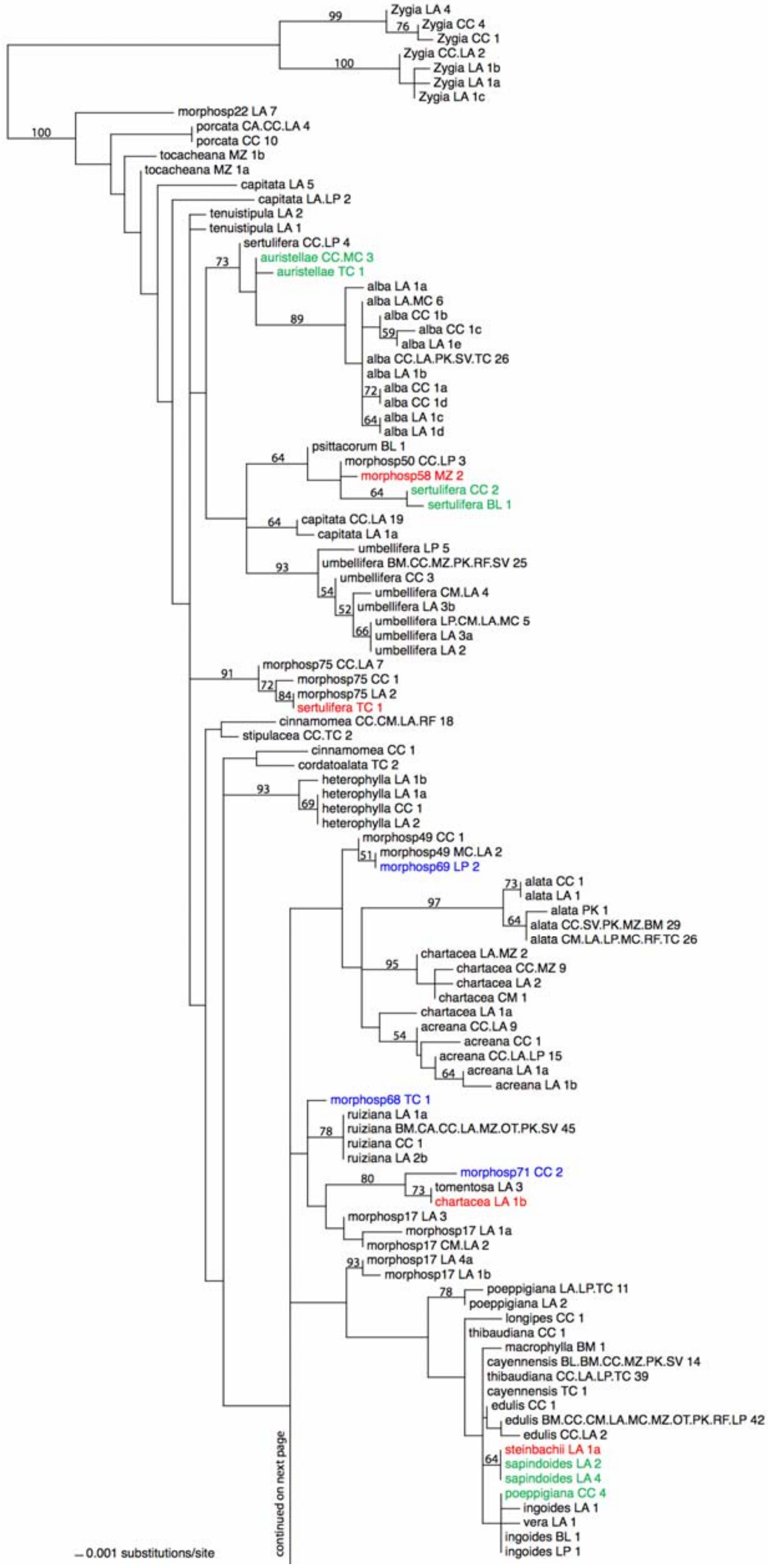
morphosp. 22
porcata
tocacheana
tenuistipula
capitata ESU1
capitata ESU2
sertulifera
brevipes
capitata ESU1
umbellifera
heterophylla
psittacorum
alba
morphosp. 75
morphosp. 50
morphosp. 55
stipulacea
cinnamomea
cordatoalata
tomentosa ESU1
tomentosa ESU2
ruiziana ESU2
ruiziana ESU1
morphosp. 49
acreana
chartacea
acreana
morphosp. 54
marginata
bourgonii
coruscans
marginata
continued on next page
0.001 substitutions/site



from previous page
auristellae
laurina ESU2
laurina ESU1
morphosp. 58
morphosp. 17
nobilis
setosa
alata
stipulacea
poeppigiana
steinbachii
sapindoides
macrophylla
cayennensis
thibaudiana
vera
barbata
ingoides
longipes
edulis
fosteriana
venusta
stenoptera ESU1
stenoptera ESU2
leiocalycina
lineata
killipiana
punctata
morphosp. 56
suaveolans
densiflora
morphosp. 56
densiflora
morphosp. 18
0.001 substitutions/site

Figure 7: Phylogenetic Tree for Combined ITS and trnD-T Dataset

Majority-Rule Consensus of 100 maximum likelihood bootstrap replicates for all unique *Inga* sequences in concatenated dataset (nuclear ITS and chloroplast trnD-T intergenic spacer). The numbers above the branches indicate the number of 100 bootstrap replicates that supported a given node (only given if >50). The numbers below the branches are the posterior probabilities for those nodes (only given if greater than 0.5) for a Bayesian phylogenetic analysis. The taxon labels are followed by two letter codes, which stand for the locations at which that sequence was found (see Fig. 6) and a number, which indicates the number of total individuals in the dataset with that sequence. The last lower case letter provides a unique identifier where necessary for comparison with Figures 8 and 9. The finalized species identities are given on the right-hand side. If a taxon's name was changed based on the results of phylogenetic analyses, it is colored: green indicates species that were incorrectly lumped with another species; blue indicates species that were incorrectly split from another species; red indicates straight mistakes in identification. The large bold letters demarcate clades that are referred to in the text.


99
Zygia LA 4
76
Zygia CC 4
Zygia CC 1
Zygia CC.LA 2
100
Zygia LA 1b
Zygia LA 1a
Zygia LA 1c
morphosp22 LA 7
100
porcata CA.CC.LA 4
porcata CC 10
tocacheana MZ 1b
tocacheana MZ 1a
capitata LA 5
capitata LA.LP 2
tenuistipula LA 2
tenuistipula LA 1
sertulifera CC.LP 4
73
auristellae CC.MC 3
auristellae TC 1
alba LA 1a
alba LA.MC 6
89
alba CC 1b
59
alba CC 1c
alba LA 1e
alba CC.LA.PK.SV.TC 26
alba LA 1b
72
alba CC 1a
alba CC 1d
64
alba LA 1c
alba LA 1d
psittacorum BL 1
64
morphosp50 CC.LP 3
morphosp58 MZ 2
64
sertulifera CC 2
sertulifera BL 1
64
capitata CC.LA 19
capitata LA 1a
umbellifera LP 5
93
umbellifera BM.CC.MZ.PK.RF.SV 25
umbellifera CC 3
54
umbellifera CM.LA 4
52
umbellifera LA 3b
66
umbellifera LP.CM.LA.MC 5
umbellifera LA 3a
umbellifera LA 2
91
morphosp75 CC.LA 7
72
morphosp75 CC 1
84
morphosp75 LA 2
sertulifera TC 1
cinnamomea CC.CM.LA.RF 18
stipulacea CC.TC 2
cinnamomea CC 1
cordatoalata TC 2
heterophylla LA 1b
93
69
heterophylla LA 1a
heterophylla CC 1
heterophylla LA 2
morphosp49 CC 1
51
morphosp49 MC.LA 2
morphosp69 LP 2
73
alata CC 1
97
alata LA 1
alata PK 1
64
alata CC.SV.PK.MZ.BM 29
alata CM.LA.LP.MC.RF.TC 26
chartacea LA.MZ 2
95
chartacea CC.MZ 9
chartacea LA 2
chartacea CM 1
chartacea LA 1a
54
acreana CC.LA 9
acreana CC 1
acreana CC.LA.LP 15
64
acreana LA 1a
acreana LA 1b
morphosp68 TC 1
ruiziana LA 1a
78
ruiziana BM.CA.CC.LA.MZ.OT.PK.SV 45
ruiziana CC 1
ruiziana LA 2b
morphosp71 CC 2
80
73
tomentosa LA 3
chartacea LA 1b
morphosp17 LA 3
morphosp17 LA 1a
morphosp17 CM.LA 2
93
morphosp17 LA 4a
morphosp17 LA 1b
78
poeppigiana LA.LP.TC 11
poeppigiana LA 2
longipes CC 1
thibaudiana CC 1
macrophylla BM 1
cayennensis BL.BM.CC.MZ.PK.SV 14
thibaudiana CC.LA.LP.TC 39
cayennensis TC 1
edulis CC 1
edulis BM.CC.CM.LA.MC.MZ.OT.PK.RF.LP 42
edulis CC.LA 2
steinbachii LA 1a
64
sapindoides LA 2
sapindoides LA 4
poeppigiana CC 4
ingoides LA 1
vera LA 1
ingoides BL 1
ingoides LP 1
continued on next page
0.001 substitutions/site

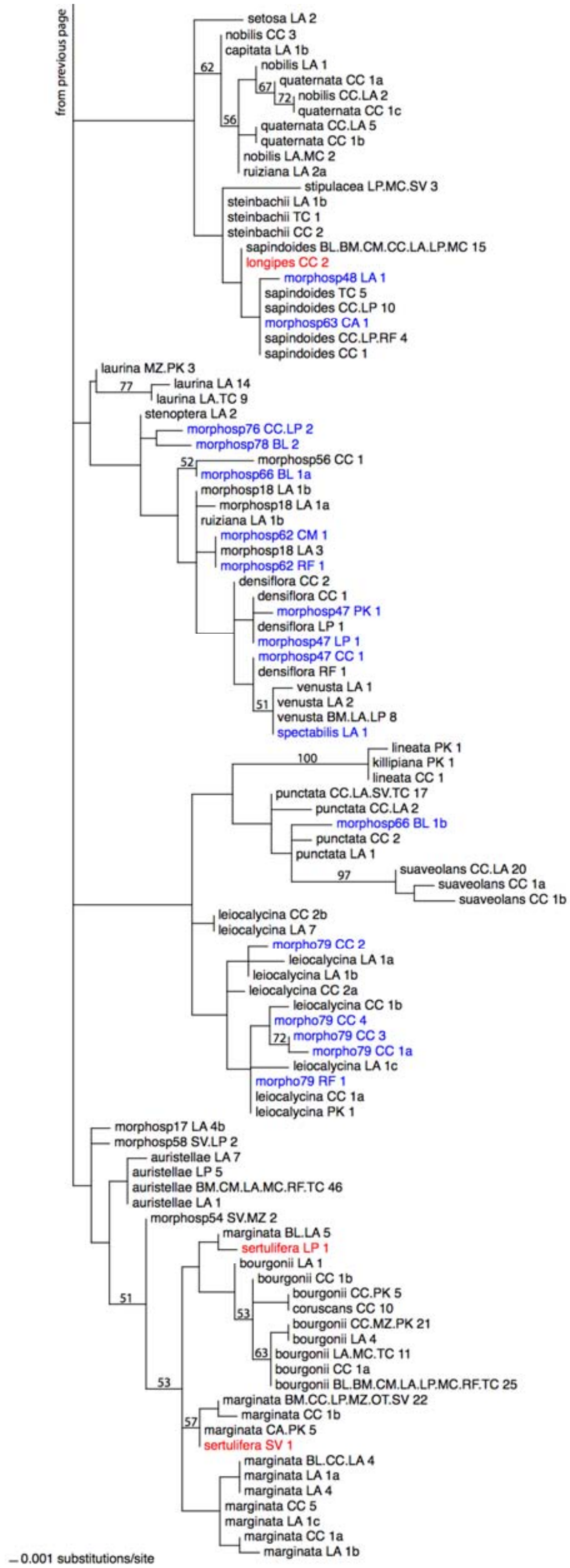
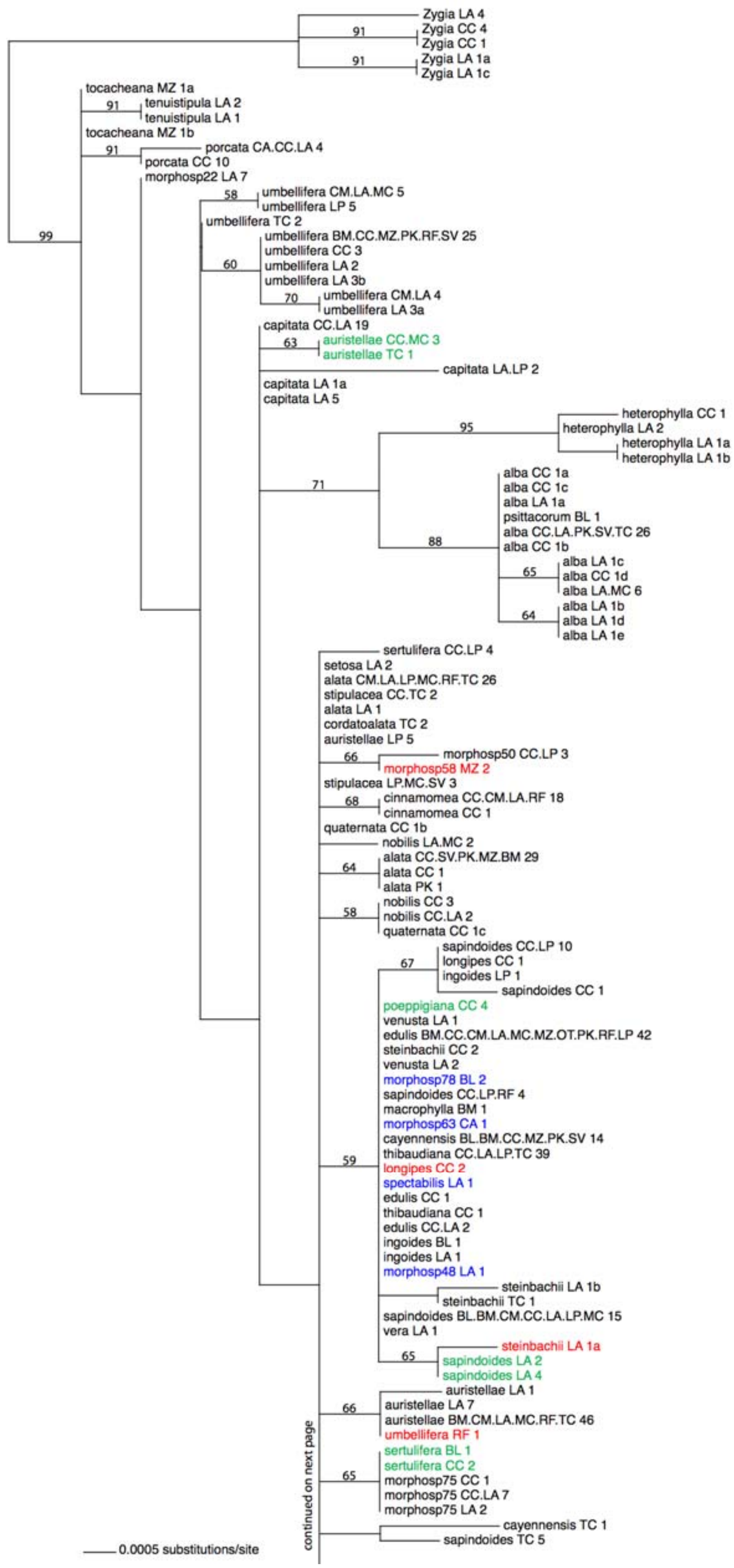

from previous page
setosa LA 2
nobilis CC 3
capitata LA 1b
nobilis LA 1
quaternata CC 1a
nobilis CC.LA 2
quaternata CC 1c
quaternata CC.LA 5
quaternata CC 1b
nobilis LA.MC 2
ruiziana LA 2a
stipulacea LP.MC.SV 3
steinbachii LA 1b
steinbachii TC 1
steinbachii CC 2
sapindoides BL.BM.CM.CC.LA.LP.MC 15
longipes CC 2
morphosp48 LA 1
sapindoides TC 5
sapindoides CC.LP 10
morphosp63 CA 1
sapindoides CC.LP.RF 4
sapindoides CC 1
laurina MZ.PK 3
laurina LA 14
laurina LA.TC 9
stenoptera LA 2
morphosp76 CC.LP 2
morphosp78 BL 2
morphosp56 CC 1
morphosp66 BL 1a
morphosp18 LA 1b
morphosp18 LA 1a
ruiziana LA 1b
morphosp62 CM 1
morphosp18 LA 3
morphosp62 RF 1
densiflora CC 2
densiflora CC 1
morphosp47 PK 1
densiflora LP 1
morphosp47 LP 1
morphosp47 CC 1
densiflora RF 1
venusta LA 1
venusta LA 2
venusta BM.LA.LP 8
spectabilis LA 1
lineata PK 1
killipiana PK 1
lineata CC 1
punctata CC.LA.SV.TC 17
punctata CC.LA 2
morphosp66 BL 1b
punctata CC 2
punctata LA 1
suaveolans CC.LA 20
suaveolans CC 1a
suaveolans CC 1b
leiocalycina CC 2b
leiocalycina LA 7
morpho79 CC 2
leiocalycina LA 1a
leiocalycina LA 1b
leiocalycina CC 2a
leiocalycina CC 1b
morpho79 CC 4
morpho79 CC 3
morpho79 CC 1a
leiocalycina LA 1c
morpho79 RF 1
leiocalycina CC 1a
leiocalycina PK 1
morphosp17 LA 4b
morphosp58 SV.LP 2
auristellae LA 7
auristellae LP 5
auristellae BM.CM.LA.MC.RF.TC 46
auristellae LA 1
morphosp54 SV.MZ 2
marginata BL.LA 5
sertulifera LP 1
bourgonii LA 1
bourgonii CC 1b
bourgonii CC.PK 5
coruscans CC 10
bourgonii CC.MZ.PK 21
bourgonii LA 4
bourgonii LA.MC.TC 11
bourgonii CC 1a
bourgonii BL.BM.CM.LA.LP.MC.RF.TC 25
marginata BM.CC.LP.MZ.OT.SV 22
marginata CC 1b
marginata CA.PK 5
sertulifera SV 1
marginata BL.CC.LA 4
marginata LA 1a
marginata LA 4
marginata CC 5
marginata LA 1c
marginata CC 1a
marginata LA 1b
62
67
72
56
77
52
51
100
97
72
51
53
63
53
57
0.001 substitutions/site

Figure 8: Phylogenetic Tree for ITS Dataset

Consensus of maximum likelihood bootstrap analysis for the nuclear ITS locus for all sequences used in the concatenated analysis. The numbers above the branches indicate the number of 100 bootstrap replicates that supported a given node (only given if >50). The taxon labels are followed by two letter codes, which stand for the locations at which that sequence was found (see Fig. 6) and a number, which indicates the number of total individuals in the dataset with that sequence. The last lower case letter provides a unique identifier where necessary for comparison with Figures 7 and 9. If a taxon's name was changed based on the results of phylogenetic analyses, it is colored: green indicates species that were incorrectly lumped with another species; blue indicates species that were incorrectly split from another species; red indicates straight mistakes in identification.


Zygia LA 4
91
Zygia CC 4
Zygia CC 1
91
Zygia LA 1a
Zygia LA 1c
tocacheana MZ 1a
91
tenuistipula LA 2
tenuistipula LA 1
tocacheana MZ 1b
91
porcata CA.CC.LA 4
porcata CC 10
morphosp22 LA 7
58
umbellifera CM.LA.MC 5
umbellifera LP 5
umbellifera TC 2
99
umbellifera BM.CC.MZ.PK.RF.SV 25
umbellifera CC 3
60
umbellifera LA 2
umbellifera LA 3b
70
umbellifera CM.LA 4
umbellifera LA 3a
capitata CC.LA 19
63
auristellae CC.MC 3
auristellae TC 1
capitata LA.LP 2
capitata LA 1a
capitata LA 5
heterophylla CC 1
95
heterophylla LA 2
heterophylla LA 1a
heterophylla LA 1b
alba CC 1a
alba CC 1c
71
alba LA 1a
psittacorum BL 1
alba CC.LA.PK.SV.TC 26
88
alba CC 1b
alba LA 1c
65
alba CC 1d
alba LA.MC 6
alba LA 1b
64
alba LA 1d
alba LA 1e
sertulifera CC.LP 4
setosa LA 2
alata CM.LA.LP.MC.RF.TC 26
stipulacea CC.TC 2
alata LA 1
cordatoalata TC 2
auristellae LP 5
66
morphosp50 CC.LP 3
morphosp58 MZ 2
stipulacea LP.MC.SV 3
68
cinnamomea CC.CM.LA.RF 18
cinnamomea CC 1
quaternata CC 1b
nobilis LA.MC 2
alata CC.SV.PK.MZ.BM 29
64
alata CC 1
alata PK 1
nobilis CC 3
58
nobilis CC.LA 2
quaternata CC 1c
sapindoides CC.LP 10
67
longipes CC 1
ingoides LP 1
sapindoides CC 1
poeppigiana CC 4
venusta LA 1
edulis BM.CC.CM.LA.MC.MZ.OT.PK.RF.LP 42
steinbachii CC 2
venusta LA 2
morphosp78 BL 2
sapindoides CC.LP.RF 4
macrophylla BM 1
morphosp63 CA 1
cayennensis BL.BM.CC.MZ.PK.SV 14
thibaudiana CC.LA.LP.TC 39
59
longipes CC 2
spectabilis LA 1
edulis CC 1
thibaudiana CC 1
edulis CC.LA 2
ingoides BL 1
ingoides LA 1
morphosp48 LA 1
steinbachii LA 1b
steinbachii TC 1
sapindoides BL.BM.CM.CC.LA.LP.MC 15
vera LA 1
steinbachii LA 1a
65
sapindoides LA 2
sapindoides LA 4
auristellae LA 1
66
auristellae LA 7
auristellae BM.CM.LA.MC.RF.TC 46
umbellifera RF 1
sertulifera BL 1
sertulifera CC 2
65
morphosp75 CC 1
morphosp75 CC.LA 7
morphosp75 LA 2
cayennensis TC 1
sapindoides TC 5
continued on next page
0.0005 substitutions/site

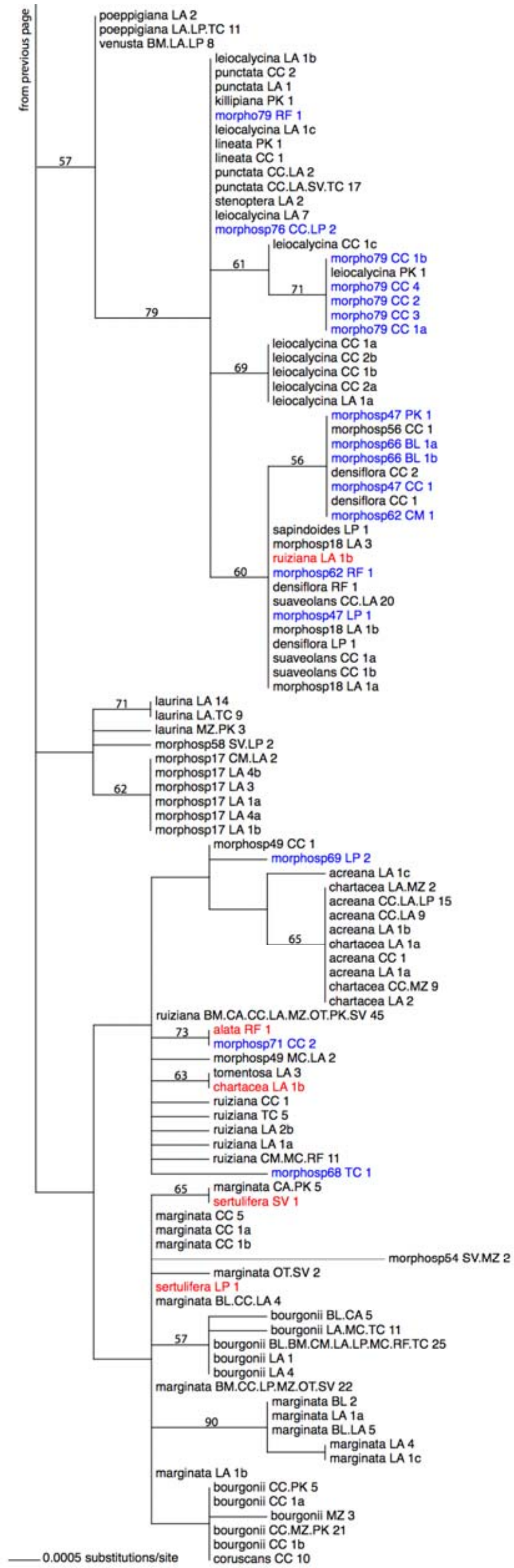

from previous page
poeppigiana LA 2
poeppigiana LA.LP.TC 11
venusta BM.LA.LP 8
leiocalycina LA 1b
punctata CC 2
punctata LA 1
killipiana PK 1
morpho79 RF 1
leiocalycina LA 1c
lineata PK 1
lineata CC 1
punctata CC.LA 2
punctata CC.LA.SV.TC 17
stenoptera LA 2
leiocalycina LA 7
morphosp76 CC.LP 2
leiocalycina CC 1c
morpho79 CC 1b
leiocalycina PK 1
morpho79 CC 4
morpho79 CC 2
morpho79 CC 3
morpho79 CC 1a
leiocalycina CC 1a
leiocalycina CC 2b
leiocalycina CC 1b
leiocalycina CC 2a
leiocalycina LA 1a
morphosp47 PK 1
morphosp56 CC 1
morphosp66 BL 1a
morphosp66 BL 1b
densiflora CC 2
morphosp47 CC 1
densiflora CC 1
morphosp62 CM 1
sapindoides LP 1
morphosp18 LA 3
ruiziana LA 1b
morphosp62 RF 1
densiflora RF 1
suaveolans CC.LA 20
morphosp47 LP 1
morphosp18 LA 1b
densiflora LP 1
suaveolans CC 1a
suaveolans CC 1b
morphosp18 LA 1a
57
79
61
71
69
56
60
laurina LA 14
laurina LA.TC 9
laurina MZ.PK 3
morphosp58 SV.LP 2
morphosp17 CM.LA 2
morphosp17 LA 4b
morphosp17 LA 3
morphosp17 LA 1a
morphosp17 LA 4a
morphosp17 LA 1b
71
62
morphosp49 CC 1
morphosp69 LP 2
acreana LA 1c
chartacea LA.MZ 2
acreana CC.LA.LP 15
acreana CC.LA 9
acreana LA 1b
chartacea LA 1a
acreana CC 1
acreana LA 1a
chartacea CC.MZ 9
chartacea LA 2
65
ruiziana BM.CA.CC.LA.MZ.OT.PK.SV 45
alata RF 1
morphosp71 CC 2
morphosp49 MC.LA 2
tomentosa LA 3
chartacea LA 1b
ruiziana CC 1
ruiziana TC 5
ruiziana LA 2b
ruiziana LA 1a
ruiziana CM.MC.RF 11
morphosp68 TC 1
73
63
marginata CA.PK 5
sertulifera SV 1
marginata CC 5
marginata CC 1a
marginata CC 1b
morphosp54 SV.MZ 2
marginata OT.SV 2
sertulifera LP 1
marginata BL.CC.LA 4
bourgonii BL.CA 5
bourgonii LA.MC.TC 11
bourgonii BL.BM.CM.LA.LP.MC.RF.TC 25
bourgonii LA 1
bourgonii LA 4
marginata BM.CC.LP.MZ.OT.SV 22
marginata BL 2
marginata LA 1a
marginata BL.LA 5
marginata LA 4
marginata LA 1c
marginata LA 1b
bourgonii CC.PK 5
bourgonii CC 1a
bourgonii MZ 3
bourgonii CC.MZ.PK 21
bourgonii CC 1b
coruscans CC 10
65
57
90
0.0005 substitutions/site

Figure 9: Phylogenetic Tree for tnD-T Dataset

Consensus of maximum likelihood bootstrap analysis for the chloroplast trnD-T locus for all sequences used in the concatenated analysis. The numbers above the branches indicate the number of 100 bootstrap replicates that supported a given node (only given if >50). The taxon labels are followed by two letter codes, which stand for the locations at which that sequence was found (see Fig. 6) and a number, which indicates the number of total individuals in the dataset with that sequence. The last lower case letter provides a unique identifier where necessary for comparison with Figures 7 and 8. If a taxon's name was changed based on the results of phylogenetic analyses, it is colored: green indicates species that were incorrectly lumped with another species; blue indicates species that were incorrectly split from another species; red indicates straight mistakes in identification.

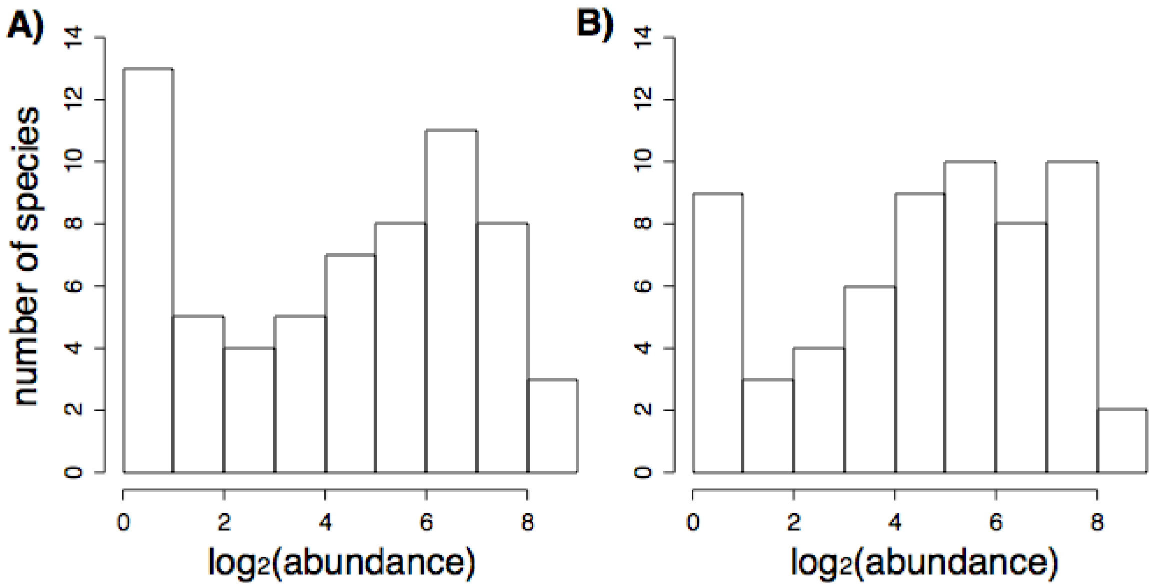


Figure 10: Distribution of Relative Abundances

The distribution of relative abundances for the entire dataset for **A)** the original species delimitations and **B)** the revised species delimitations.

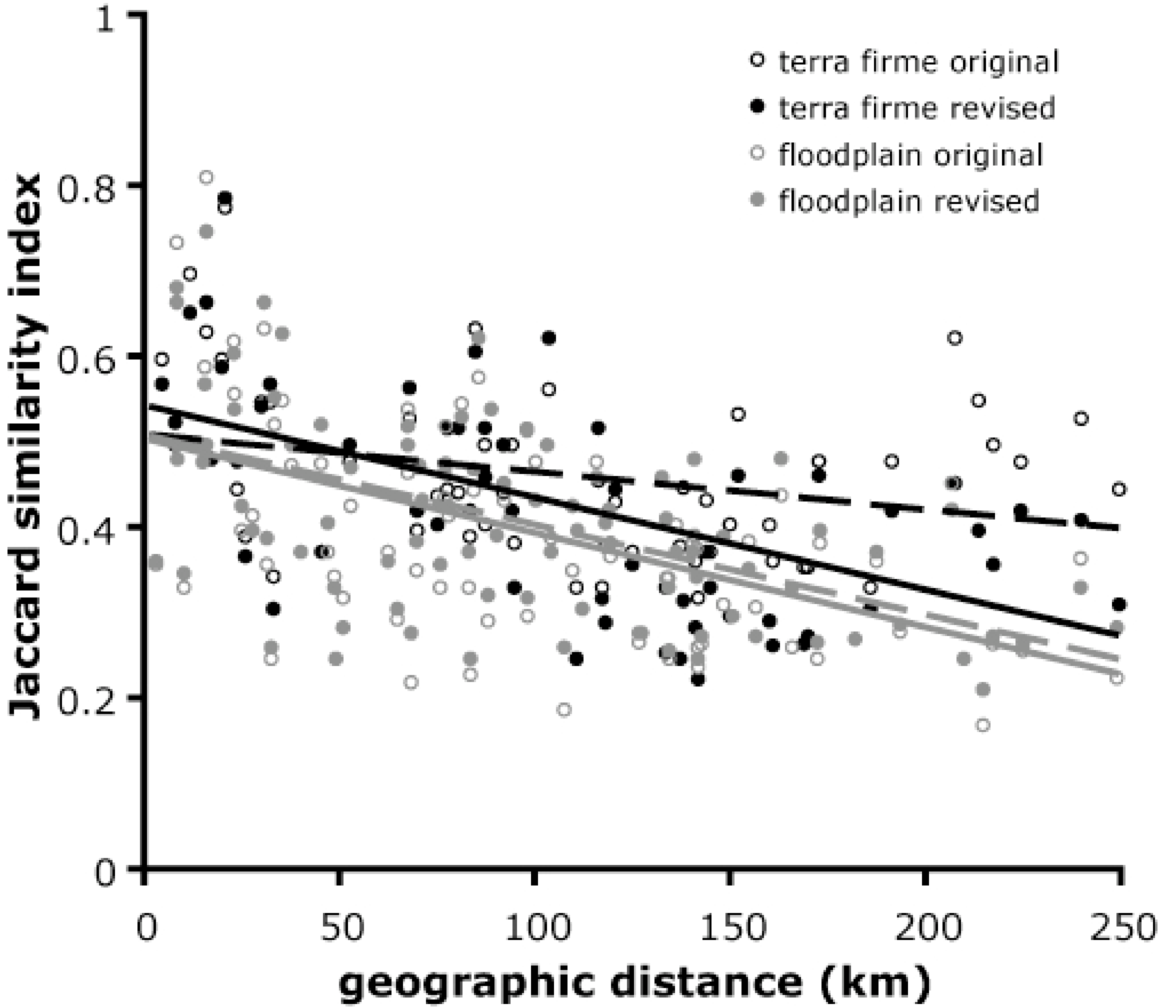


Figure 11: Distance-decay in Community Similarity (Jaccard Index)

Decline in Jaccard Similarity Index with distance in Madre de Dios for floodplain and terra firme *Inga* communities. Linear relationships were fit using reduced major axis regression and a jackknifing procedure (dashed lines=original delimitations, solid lines=revised delimitations).

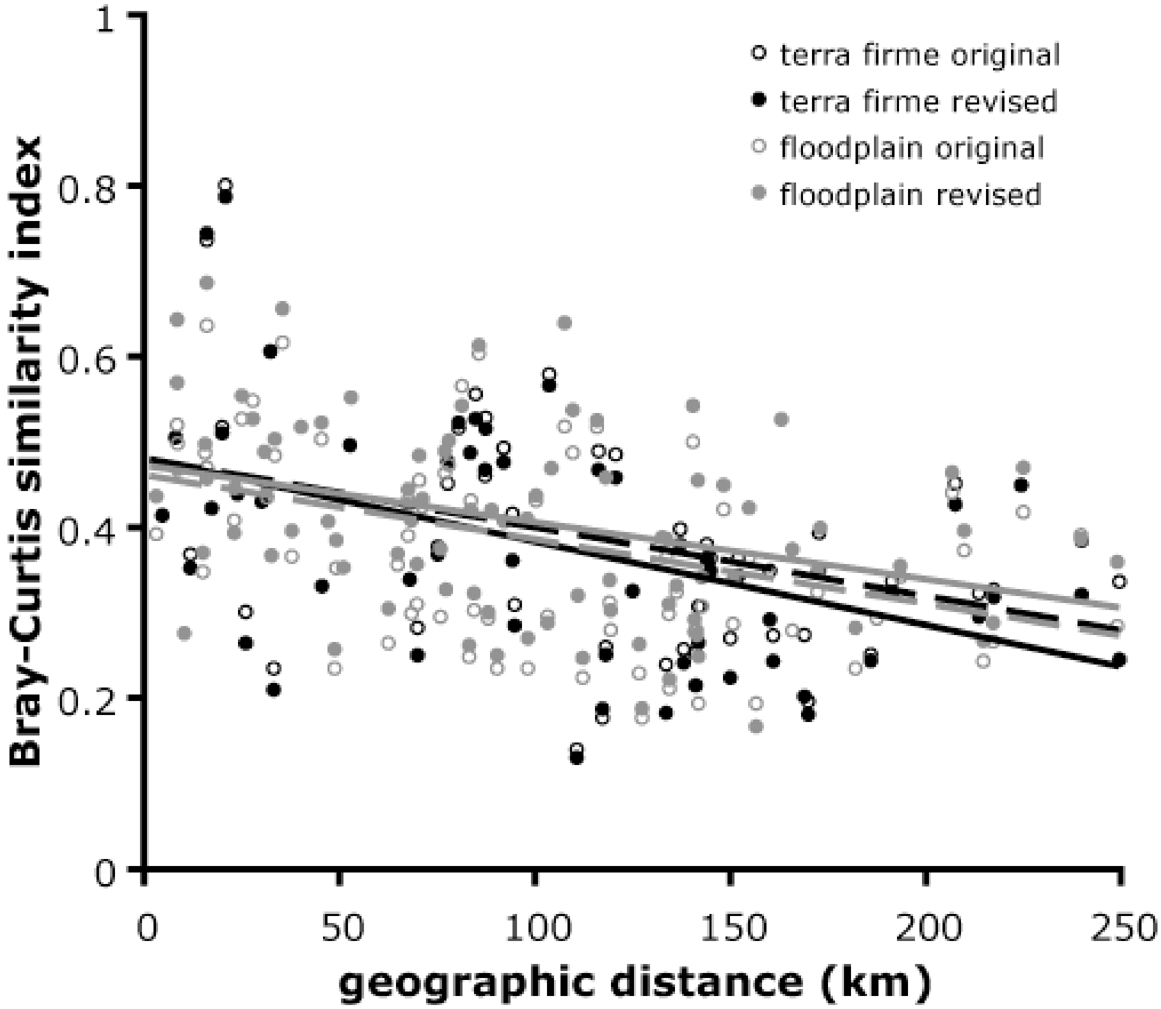


Figure 12: Distance-decay in Community Similarity (Bray-Curtis Index)

Decline in Bray-Curtis Similarity Index with distance in Madre de Dios for floodplain and terra firme *Inga* communities. Linear relationships were fit using reduced major axis regression and a jackknifing procedure (dashed lines=original delimitations, solid lines=revised delimitations).

4. Beta-Diversity, Ecological Determinism and Historical Contingency in Tropical Tree Communities

Tropical tree communities have become a model system for exploring the effects of dispersal on patterns of beta-diversity, the turnover in composition and abundance of species among communities. However, previous studies have estimated dispersal from the level of turnover whereas independent estimates of dispersal are needed to accurately interpret beta-diversity patterns. Here, for the first time we use a population genetic approach to obtain community-level estimates of seed dispersal among Amazonian tree communities. These genetic estimates are compared to analyses of beta-diversity for those same communities. We find that the relative abundances of species are highly correlated across the landscape despite severe dispersal limitation as measured by the genetic data. This suggests that non-neutral, ecological forces are determining which species are common versus rare. We also find a decline with distance in the proportion of species shared between communities. Further analyses show that this distance-decay is due to a narrow, geographic zone of particularly limited seed dispersal and high species turnover. Deep genetic breaks in many species suggest that this is a transition zone of secondary contact between rainforest communities separated for at least 175,000 years. This in turn suggests a heretofore unnoticed role for historical contingency in tropical tree community assembly.

4.1 Main Text

Determining the mechanisms behind patterns of beta-diversity, the turnover in composition and abundance of species among communities, is critical to understanding what controls the species diversity and structure of ecological communities. Both dispersal limitation and environmental gradients can be responsible for beta-diversity. In order to focus on the role of dispersal limitation, beta diversity is often examined within homogeneous environments such as upland terra firme or floodplain forest (Pitman et al. 2001, Condit et al. 2002).

Tropical rainforests can span homogeneous environments over large geographic areas (Pitman et al. 2001, Macia and Svenning 2005) and have become a model system for analyzing the impact of dispersal on beta diversity (Chave and Leigh 2002, Condit et al. 2002, Hardy and Sonke 2004, Morlon et al. 2008). Various tropical rainforest studies have found that, within single habitat types, floristic similarity declines with geographic distance (Condit et al. 2002, Tuomisto et al. 2003, Phillips et al. 2003, Morlon et al. 2008). In contrast, other studies have shown that, while there may be moderate turnover in species composition, the relative abundances of shared species are highly correlated, even over distances as great as 1500 km (Terborgh et al. 1996, Pitman et al. 2001, Vormisto et al. 2004, Macia and Svenning 2005). Under a neutral theoretical framework (Bell 2001, Hubbell 2001), the first pattern is consistent with limited dispersal between communities while the latter pattern is consistent with high rates of dispersal.

In order to determine the mechanisms underlying beta-diversity patterns, one should independently measure the level of dispersal limitation in tropical rainforest landscapes. Falsifying the high dispersal hypothesis would support the notion that dispersal limitation largely explains the observed turnover in species composition, even while ecological factors drive shared species to similar relative abundances in geographically separated communities (Terborgh et al. 1996, Pitman et al. 2001). Falsifying the low dispersal hypothesis would suggest that high dispersal may be responsible for homogenizing relative abundances across space, but that other forces, such as cryptic environmental gradients, are responsible for turnover in species composition (Nekola and White 1999, Tuomisto et al. 2003). Here, for the first time, we compare genetic estimates of dispersal between tropical tree communities with analyses of turnover in species composition and correlations in relative abundances.

We surveyed tree communities in both terra firme (10 sites) and floodplain (12 sites) along a 250 km transect in southern Amazonian Peru (see supporting online material [SOM]: field methods). Soil analyses show no evidence for an environmental gradient in terra firme along this transect (see SOM: soil analyses). There is a gradient in soil fertility in floodplain, although the variation in soil fertility within floodplain is less than that found between terra firme and floodplain (Fig. 15). Surveying communities in both habitat types allows us two independent assessments of beta-diversity patterns within relatively homogeneous environments. We focused our community surveys on members of the genus *Inga* (Fabaceae) (Pennington 1997). This is the most diverse and abundant tree genus in southern Peru (55 species, ~159 individuals >1 cm DBH per hectare, K.G.D. unpublished data) and thus provides suitable sample sizes of individuals and species for examining community-level patterns.

Both terra firme and floodplain forests show significant and nearly identical distance-decay, the decline with distance in the proportion of species shared between communities (Fig. 13a). While a gradient in soil fertility may be the cause of distance-decay in floodplain, no environmental gradients were found in terra firme that could explain distance-decay there. This suggests that either there is a cryptic environmental gradient in terra firme or some other process, such as dispersal limitation, is driving the pattern to be the same in both habitats. Both habitats also show a high correlation among communities in the relative abundances of shared species, and this correlation does not decline with distance (Fig. 13b). The distance-decay pattern is similar to that found by Condit et al. (2002) and other studies (Tuomisto et al. 2003, Hardy and Sonke 2004), while the correlation pattern is similar to that found by Terborgh, Pitman, and others (Terborgh et al. 1996, Pitman et al. 2001, Vormisto et al. 2004, Macia and Svenning 2005).

We next used a community-level population genetic approach to determine the rates of seed dispersal across this landscape. In order to focus on seed dispersal, we measured gene flow for a plastid genetic marker (the trnD-T intergenic spacer, see SOM: Laboratory Methods). The chloroplast genome is maternally inherited in most angiosperms, including all measured mimosoid legumes (Zhang 2003), and chloroplast gene flow should therefore reflect rates of seed dispersal (Latta 2006). We first examined chloroplast gene flow for both terra firme and floodplain populations between the Manu and Los Amigos river basins, which are separated by ~100km, for 20 of the 24 species shared between the two basins. In both habitat types, there is a high correlation between these two basins in the relative abundance of shared species (Fig. 16, floodplain: $r = 0.66$, terra firme $r = 0.70$). Seven species had no nucleotide polymorphisms for the surveyed marker, and it is difficult to infer seed dispersal rates for these species. Nearly all of the 13 remaining species showed high and significant differentiation between the two basins (Table 11, mean Fst=0.69+/-0.08 S.E.; for all species including those with no polymorphism, mean Fst=0.45+/-0.09 S.E.).

While the results from Fst analyses suggest limited seed dispersal, we used a coalescent approach (Hey and Nielsen 2007) to explicitly measure migration via seed dispersal between the two basins. We conducted analyses at the level of individual species and the community; community-level analysis consisted of treating each species, including species with no polymorphisms, as a separate locus in a multi-locus assessment of migration. While we do not expect that seeds from one basin disperse directly into the other basin, this analysis does reflect how often a genotype from one river basin arrives in the other basin over multiple generations of dispersal and establishment. If we detect high rates of migration, then this supports a role for dispersal in establishing the high correlation in relative abundances between the two basins, while low rates of migration would mean that other forces are responsible for this pattern.

The inferred rates of migration between the two basins are extremely low, both at the level of individual species and the community. For all of the species with nucleotide polymorphism (and for which coalescent analyses were thus possible), the most probable migration rates are near zero and, in most cases, not significantly different from zero (Table S1, Fig. S2). Limited sample size (19-32 individuals per species) means that we may have missed individuals with immigrant genotypes, but the consistent pattern across all species suggests that migration rates are in fact very low. The community-level analysis, which additionally includes species with no nucleotide polymorphism, also finds that the most probable estimates of migration rate are near zero (Table S2, Fig. S2).

Coalescent and Fst analyses of a nuclear marker, the internal transcribed spacer, reveal that there is ample nuclear gene flow between the two river basins (Table 11). This suggests that pollen dispersal is maintaining the evolutionary and ecological identity of these species in the face of highly limited seed dispersal (Cannon and Manos 2003, Latta 2006).

This finding of strong dispersal limitation signifies that high dispersal is not responsible for the high correlation in relative abundances of species shared between these two basins (Fig. S2). Clearly, some ecological forces, such as variation in competitive ability or susceptibility to disease, are driving species to similar relative abundances in different communities. In discussing data from Amazonian tree communities, several authors (Hubbell 2001, Condit et al. 2002) have suggested that the most common species may be 'non-neutral', or competitively superior, while the rest of the species are neutral and drift randomly in abundance. Although *Inga* is abundant at the genus level, individual species are generally rare relative to the rest of the community; no *Inga* species is among the 20 most common species in Madre de Dios (Pitman et al. 2001). This study shows that even among the less common species in a community, there is a regular dominance hierarchy that persists across 250 km.

In order to determine if dispersal is highly limited among all communities in this landscape, or just between the Manu and Los Amigos basins, we obtained chloroplast DNA sequences from populations along the entire transect for 8 widespread *Inga* species. We found that chloroplast gene flow rates are lower between the Manu and Los Amigos basins than anywhere else in the landscape (Fig. 14b). Community-level coalescent analyses (see SOM: Population Genetic Analyses) show that average migration rates within the Los Amigos and Manu basins are 5-6 times higher than between these two basins, and average migration rates between the Los Amigos and Tambopata basins are more than twice as high (Table 12).

We also find that the greatest turnover in species composition, in both floodplain and terra firme, occurs between the Manu and Los Amigos basins (Fig. 14c and d). Based on these results, we tentatively divide southern Peru into two regions, Manu and Los Amigos-Tambopata, which are separated by a transition zone of increased species turnover and reduced seed dispersal (Fig. 14a). Perhaps the most striking result is that average chloroplast gene flow rates between the Manu region and northern Peru, 915 km apart, and between the Los Amigos-Tambopata region and central Bolivia, 620km apart, are higher than across the transition zone within southern Peru, which is less than 100 km wide (Fig. 14).

Recent research suggests a potential cause for this transition zone. Based on geologic and paleoecological evidence, Olivier (2007, 2008) has suggested that the Fitzcarrald Arch (Espurt et al. 2007), which runs through the southern Peruvian Amazon, experienced a dry climate in the Pliocene and/or Pleistocene. Rainforest taxa may have been displaced by species that prefer drier conditions. Therefore, extant populations of rainforest species in the area may represent recent colonists from the north, southeast, or other directions that arrived after the onset of favorable climatic conditions. If a diverse set of colonists is arriving from multiple directions, there may be zones of high species turnover and low gene flow where they meet. Coalescent analyses suggest that populations in the Manu and Los Amigos-Tambopata regions were separated at least 175,000 years ago (see SOM: Population Genetic Analyses), which is roughly commensurate with the timing of this putative historical scenario.

Further research is needed to confirm if this scenario is the best explanation for the observed transition zone in southern Peru. Either way, the discovery of this zone has important implications for our understanding of patterns of beta-diversity in the Amazon. Under a neutral, equilibrium framework with steady-state dispersal limitation (Hubbell 2001), species turnover rates should be constant across the landscape and community similarity should decline smoothly with geographic distance (either linearly or logarithmically depending on the spatial scale; Condit et al. 2002, Chave and Leigh 2002). In fact, researchers have interpreted smooth distance-decay in community similarity as evidence for a neutral, equilibrium scenario (Hardy and Sonké 2004). However, we have shown that there can appear to be smooth distance-decay (Fig. 13a), even when turnover is not constant across the landscape (Fig. 14c & d). We do not find evidence for distance-decay when analyses are restricted to either side of the transition zone (i.e. within the Manu or Los Amigos-Tambopata regions, see Table 13). This suggests that this narrow zone of high species turnover is responsible for the pattern of distance-decay in community similarity that we observe in southern Peru (see SOM: Beta-Diversity Analyses for further support).

As with previous authors (Chave and Leigh 2002, Condit et al. 2002, Tuomisto et al. 2003), we attribute distance-decay in the similarity of tropical tree communities to dispersal limitation. However, we suggest, at least in southern Peru, that the dispersal limitation is historical in nature. Given sufficient time and environmental stability, species may disperse across the transition zone in southern Peru and erase any signature of distance-decay. Species are capable of seed dispersal over large distances as indicated by the high chloroplast gene flow rates between the Manu region and northern Peru and between the Los Amigos-Tambopata region and southern Peru (Fig. 2b). The large geographic range of many *Inga* species also suggests broad dispersal capability (Pennington 1997). Many of the species in this study, including some that are missing from either side of the transition zone (e.g. *Inga auristellae* and *poeppigiana*), have ranges that extend to the eastern Amazon basin and even Central America. Thus, the transition zone has not limited the absolute range of species, but it is likely responsible for patchy distributions and distance-decay within southern Peru.

Whether similar transition zones are responsible for patterns of distance-decay observed elsewhere in Amazonia is unknown at this point. A transition zone in tree community similarity has recently been observed in northern Peru (Pitman et al. 2008), but no estimates of dispersal were made across this zone. Patton et al. (2000) found a genetic break that was geographically coincident for multiple mammal species in the Brazilian Amazon and that may correspond to a zone of increased species turnover (Gascon et al. 2000). Interestingly, both of these transition zones correspond to the underlying Iquitos Arch, which, together with this study, suggests a general role for geologic arches in causing reduced dispersal and increased turnover in species composition. Further integration of phylogeographic, community ecological, and geologic studies are needed to address the role of arches in generating biodiversity patterns across the Amazon basin.

Our results clearly show that high dispersal is not responsible for the high correlation in relative abundances of species shared across southwestern Amazonian tree communities. Non-neutral processes, such as differences in competitive ability, must be driving species, even rare ones, to achieve similar relative abundances in different communities. Our results do suggest that dispersal limitation is responsible for turnover in species composition among communities. However, this is not steady-state dispersal limitation in an equilibrium framework, but instead may reflect historical processes of extinction and re-colonization. Thus, historical contingency may guide the assembly of Amazonian tree communities, while ecological factors subsequently act to determine the dominance hierarchy of the assembled species.

4.2 Supplementary Materials and Methods

4.2.1 Field Methods

Floodplain communities were surveyed at CC, PK, SV, OT, BM, BL, CA, RF, CM, MC, CI, and TC. Terra firme communities were surveyed at MZ, CC, PK, SV, BM, RF, CM, MC, CI, and TC. Abbreviations in Figure 14a and throughout are as follows: MZ=Maizal, CC=Cocha Cashu, 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, CI=CICRA, TC=Tambopata. We surveyed *Inga* communities at each site through a combination of 25 x 25m plots and transects randomly scattered over a 4 km^2 area. Within each plot or transect, we identified and measured height and diameter at breast height (1.3m, DBH) for every individual of the genus *Inga* >1.3m in height. All individuals were collected for a genetic voucher in silica gel and most individuals, including all of uncertain identification, were also collected for an herbarium voucher. Representative vouchers of all species have been deposited in herbaria at Duke University (DUKE), La Universidad Agraria La Molina in Lima, Peru (MOL), and Kew Botanic Garden (K). Survey effort varied among sites with anywhere from 69 to 282 individuals being surveyed (mean = 134.8 +/- 13.7 S.E.).

4.2.2 Soil Analyses

In order to assay environmental variation, we collected one to several soil samples at each community survey site. Individual soil samples consisted of 5 bulked soil cores taken over a 0.06 to 0.25 ha area. Soil samples were analyzed at the Clemson University Agricultural Services Laboratory for pH, concentration of multiple base cations (K, Ca, Mg, Zn, Mn, Cu, B, Na), nitrate nitrogen (NO3), phosphorous, cation exchange capacity, and base saturation. Soil samples were analyzed at the North Carolina State Soil Services Laboratory for particle size distribution. If multiple soil samples were collected at a given site, the values of these variables were averaged to provide a single set of values for each community.

In order to reduce the dimensionality of the data, we conducted a principal component analysis on all the measured soil variables. Principal Component Axis 1 (PCA1) explained 49% of the variation in the soil data and was significantly correlated with pH, base cation concentrations, cation exchange capacity, base saturation, phosphorous concentration, and percent sand. This axis largely differentiated floodplain soils from terra firme soils (Fig. 15). Floodplain soils had higher pH, higher base cation concentrations, higher cation exchange capacity, higher base saturation, higher phosphorous concentration, and less sand (greater silt and clay) than terra firme soils. PCA2 explained 15% of the variation in the soil data, PCA3 explained 10% of the variation, and all other axes individually explained less than 8% of the variation.

We probed for environmental gradients in floodplain and terra firme by assessing spatial autocorrelation for these principal component axes. We calculated the Euclidean distance between communities for each of the first three principal component axes described above and used these values to construct three matrices of environmental dissimilarity between communities. We then used Mantel tests to assess the correlation between these matrices and a matrix comprised of geographic distances between communities. Only the distance matrix for PCA1 in floodplain communities showed a significant correlation with geographic distance ($r=0.46$, $p=0.011$). In general, our results are consistent with Pitman (2000) and Phillips et al. (2003) who reported large differences between floodplain and terra firme soils in southern Amazonian Peru, but little soil variation within each habitat type.

4.2.2 Beta-Diversity Analyses

To measure the similarity of communities in terms of species composition, we used the Jaccard Similarity Index (corrected for maximum possible values as in Pitman et al. 2005). The similarity of communities in terms of abundance was taken simply as the correlation coefficient (r) between communities in the relative abundances of shared species. Following Terborgh et al. (1996) and Pitman et al. (2001), we also calculated the Spearman's rank correlation (r_s) between communities. The results with r_s are equivalent to those with the standard correlation coefficient (r) and are not presented here. We analyzed patterns of beta diversity by plotting indices of similarity against the linear geographic distance between communities. The strength and significance of these relationships was assessed using Mantel tests. Log transforming either similarity indices or geographic distance led to little or no improvement in the fit of these relationships.

In order to assess the correlation in relative abundances between the Manu and Los Amigos basins as a whole we combined multiple community surveys and calculated the overall relative abundance of species in each basin. To represent the Manu floodplain community, we combined floodplain surveys from CC, PK, SV, and OT. To represent the Manu terra firme community, we combined terra firme surveys from MZ, CC, PK, and SV. To represent the floodplain community in Los Amigos, we combined floodplain surveys from RF, CM, MC, and CI. To represent the terra firme community in Los Amigos, we combined terra firme surveys from RF, CM, MC, and CI. Relative abundances between the two basins were highly correlated in both floodplain and terra firme (Fig. 16). These correlations were higher than the average correlation between individual community surveys (Figure 13b), perhaps reflecting that lumping multiple communities reduced sampling error.

In order to analyze the effect of the transition zone on patterns of distance-decay, we assessed the Mantel correlation between geographic distance and the Jaccard Similarity Index for communities restricted to either side of the transition zone. Within either the Manu or Los Amigos-Tambopata region (Table 13), there was no evidence for significant distance-decay. However, the lack of significant results could be due in part to the low sample size of communities within regions (N=4 and 5). We therefore explored the effect of the transition zone on distance-decay using the following, additional approach.

We analyzed the residuals from the linear relationship shown in Figure 13a for two groups of pairwise comparisons: those for plots within the Manu or Los Amigos-Tambopata region versus those between the two regions. For both habitats, the residuals for comparisons within regions were significantly higher than between regions (Kruskal-Wallis test, floodplain: $p=0.0356$, terra firme $p=0.0068$). If we conduct this analysis with residuals from the relationship between the Jaccard similarity index and the logarithm of geographic distance, the results are the same (Kruskal-Wallis test, floodplain: $p=0.0304$, terra firme: $p=0.0201$). Residuals for pairwise comparisons between communities in the transition zone and other communities generally fell intermediate to the other two groups, and it is these comparisons that smoothed out the relationship between geographic distance and the Jaccard similarity index. Thus, it seems likely that the transition zone and the large compositional differences between communities on either side of the transition zone are driving the pattern of smooth distance-decay in the Jaccard similarity index observed in southern Peru.

4.2.3 Laboratory Methods

We extracted DNA from silica-gel dried leaf specimens using a modified CTAB protocol (Doyle and Doyle 1990). We used polymerase chain reaction (PCR) to amplify the trnD-T chloroplast intergenic spacer from individuals selected for analysis. We used the trnH (trnT in Demesure et al. 1995), and trng2 primers (Oh and Potter 2003). Reaction conditions were as follows: 1 cycle of 94°C for 2 min; 40 cycles of 94°C for 30 s, 55°C for 1 min, 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 Inc., Valencia, California), 2.5 μL Taq Buffer, 0.5 μL dNTP mix (10 mM concentration for each nucleotide), 1.25 μL primer 1 (10 μM concentration), 1.25 μL primer 2 (10 μM 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). Sequences were assembled using Sequencher v4.5 (Gene Codes Corp., Ann Arbor, Michigan) and manually aligned using MacClade v4.06 (Maddison and Maddison 2003). The protocol for amplification and sequencing of the ITS region was identical except that we used the ITS 4 (White et al. 1990) and ITS 5a primers (Stanford et al. 2000) and the annealing step of the PCR was performed at 52°C. All sequences used in this study have been deposited in Genbank.

The ITS sequences varied in length from 652 to 661 base pairs across species (mean=655). TrnD-T sequences were trimmed to the central region for which all sequences of a given species were readily interpretable. This region varied from 1085 to 1392 base pairs across species (mean=1225).

4.2.4 Population Genetic Analyses: Manu and Los Amigos Basins

We selected individuals from populations in the Manu and Los Amigos basins for 20 of the 24 species that are shared between the two basins (the 4 species not assayed are very rare in both basins). For 14 species, we were able to obtain sequences for the chloroplast trnD-T region and the nuclear ITS region from at least 8 individuals per population, while for 6 species, sample sizes were lower. Low sample size species are noted in Table 11. ITS sequences often had multiple ambiguous sites. We therefore used the program PHASE v2.1.1 (Stephens et al. 2001) to reconstruct haplotypes for individuals. Individuals for which we could not reconstruct haplotypes with high confidence (> 0.8 probability) were excluded from population genetic analyses. One species had no nucleotide polymorphisms for ITS (*I. porcata*), while seven species had no polymorphism for the trnD-T marker. We calculated the Fst between populations in Manu and Los Amigos using the program Arlequin v2.000 with the Tamura-Nei genetic distance formula (Schneider et al. 2000). Low sample size species did not differ significantly from high sample size species in their Fst values for either trnD-T or ITS and were thus included in the average Fst values reported in the main text.

We conducted coalescent analyses of migration between Manu and Los Amigos using the program IMa (Hey and Nielsen 2007). We restricted species-level analyses to species with high sample size (≥8 individuals per population) and nucleotide polymorphism. In preliminary analyses, species with low sample size or no polymorphism failed to converge for migration rates or any other parameters that the program IMa attempts to estimate (despite long run times and using multiple chains). This is not surprising as both classes of species have little information in their sequence data. Community-level analyses, where we treated each species as a separate locus in a multi-locus analysis, incorporated all species with high sample size, including those with no nucleotide polymorphism.

The program IMa assumes no recombination among the sequences analyzed. We checked for recombination using the 4-gamete test (Hudson and Kaplan 1985) and linkage disequilibrium tests (Piganeau et al. 2004). The trnD-T locus showed no evidence of recombination for any species, as might be expected given that it is in the chloroplast genome. For the ITS marker, some species failed the 4-gamete test, but these species showed no evidence of significant recombination using linkage disequilibrium tests. This indicates either that there multiple substitutions at some sites or that there is too little information in the data (too few segregating sites) to detect recombination via linkage disequilibrium-based methods. For species with ITS sequences that failed the 4-gamete test, we used two approaches to conduct IMa analyses: 1) implementing the HKY model of sequence evolution (allowing for multiple substitutions per site) and 2) implementing the infinite sites model while treating recombination blocks as separate loci. Potential recombination blocks were determined using the program SITES (Hey and Wakeley 1997). Preliminary analyses indicated that both approaches gave equivalent results, and here we only report results using the first approach (the HKY model). For ITS analyses that showed no evidence of recombination and for all trnD-T analyses, we used the infinite sites model of sequence evolution.

For all IMa analyses, we set the upper limit for the uniform prior on divergence time (t) to a value corresponding to 10 million years. This has been estimated as the maximum age of the genus *Inga* (Richardson et al. 2001, Lavin 2006) and provides a conservative upper bound. For analyses of the trnD-T chloroplast marker, this corresponded to an average value of 15.96 for the parameter *t*, given a mutation rate of 1.3×10^{-9} substitutions per site per year (s/s/y) for chloroplast intergenic spacers in *Inga* (Richardson et al. 2001) and an average sequence length of 1225 base pairs. For analyses of the ITS marker, this corresponded to an average value of 15.32 for *t*, given a mutation rate of 2.31x10^{-9} s/s/y for ITS in *Inga* (Richardson et al. 2001) and an average sequence length of 655 base pairs. The upper limits for uniform priors on theta parameters (*q*1, *q*2, and *qA*) were adjusted based on results from preliminary runs of IMa. We generally left the upper bound on theta parameters at the default value of 10, although in some cases this was adjusted upward to a value as high as 100.

Based on preliminary runs in IMa, we set the prior upper limit for the *m*1 and *m*2 parameters at 1000. Nearly all estimates of *m*1 and *m*2 reached zero probability well before this value. For the trnD-T marker, this upper limit is equivalent to a per generation migration rate of 0.048 (i.e. 0.048 of new individuals in a generation carry an immigrant genotype). This calculation assumes the mutation rate cited above and a generation time of 30 years. *Inga* species can reproduce in as little as 3 years, if they have sufficient access to light and other resources (Pennington 1997), although in closed-canopy situations, generation time could be much longer. Nevertheless, *Inga* species do tend to be fast-growing compared to other tropical tree species, and we believe an estimate of 30 years for generation time is conservative (lower values would result in a lower calculated per generation migration rate). For ITS, a value of *m*1 or *m*2 of 1,000 corresponds to a per generation migration rate of 0.045.

We ran each analysis for a sufficient number of generations to ensure that effective sample size (ESS) values were higher than 50 for all parameters sampled over the course of the run (*t*, mutation rate scalars, etc). For most analyses, a run of 100 million generations was sufficient to pass this ESS threshold. For all analyses, we used a burn-in of 10% of the total generation time. We also examined trend lines to ensure that the program was sufficiently mixing across parameter values over the course of the run. For some ITS analyses, sufficient mixing did not occur with a single chain, and we used multiple chains with a geometric heating scheme. Sufficient mixing was obtained once the heating schemes were properly calibrated to ensure adequate swapping among chains. In order to evaluate convergence, we ran each analysis at least 2 times once the final values for priors on parameters, number of generations, and heating schemes (if necessary) were established.

For each analysis, we selected genealogies from one of the final runs to obtain joint posterior estimates for migration between communities (*m*1 and *m*2). We also conducted nested likelihood ratio tests to assess if models with migration in both directions were significantly better than models with migration in one direction or with zero migration. Both of these steps were implemented in the L mode of the IMa program (Hey and Nielsen 2007). In Table 11, we report the species-level joint posterior estimates of migration and whether those migration estimates were significantly different from zero. The average community-level assessment of migration between the Manu and Los Amigos basins is given in Table 12. We estimated the minimum divergence time between Manu and Los Amigos as the lower 95% confidence interval on the parameter *t* for the community-level analysis (which included species with no genetic variation). This estimate was converted to absolute time using the mutation rate cited above.

4.2.5 Population Genetic Analyses: Whole Transect Analyses

We selected 8 widespread and abundant species (*Inga alata, alba, auristellae, bourgonii, marginata, ruiziana, sapindoides*, and *umbellifera*) for analyses of chloroplast gene flow and migration along the transect. For each site where a given species was found, we obtained sequences of the chloroplast trnD-T marker from 3 to 6 individuals, although in a few cases, populations were only represented by 1 or 2 individuals. For each species, we calculated Fst between neighboring populations along the transect using the program Arlequin v2.000 (Schneider et al. 2000). Note that neighboring populations for some species do not necessarily correspond to neighboring survey sites. For example, *I. bourgonii* is not found at Salvador or Otorongo, so Fst between Pakitza and Boca Manu is counted as the Fst between neighboring populations. We assigned Fst values a point along the transect that is the mid-point between the two populations used to generate the estimate. When we calculated the average Fst between neighboring sites, we included all Fst values that were assigned to points between those two sites. For the *I. bourgonii* example above, the mid-point between Pakitza and Boca Manu lies between Otorongo and Boca Manu, so the Fst estimate between Pakitza and Boca Manu was incorporated into the average Fst estimate between Otorongo and Boca Manu. In order to consider chloroplast gene flow, we converted these average Fst estimates into values of M, the number of migrants per generation, using the formula M=(1-Fst)/(2*Fst) (Schneider et al. 2000). As an Fst estimate of 0 produces an M estimate of infinity, we set the maximum M value at 15 migrants per generation.

We next aimed to compare coalescent estimates of migration between other locations along the transect with that found between the Manu and Los Amigos basins (described above). However, sample sizes were not generally large enough to estimate migration between all neighboring sites. To obtain estimates of migration between communities within the Manu basin, we combined samples from MZ and CC and estimated migration between that group and the combined samples from PK and SV. To obtain estimates of migration within the Los Amigos basin, we combined samples from RF and CM and estimated migration between that group and the combined samples from MC and CI. Lastly, to obtain an estimate of migration between the Los Amigos and Tambopata basins, we combined samples from MC and CI and estimated their migration with samples from TC. For these analyses, we conducted community-level assessments of migration following the protocols outlined above, with the same prior distributions, number of generations, etc. The average joint posterior estimates of migration for each analysis are reported in Table 12.

Table 11: Species-Level Estimates of Gene Flow between Manu and Los Amigos River Basins

Fst estimates between Manu and Los Amigos river basins plus migration estimates (from coalescent analyses) from the Manu basin into Los Amigos and vice versa.

	Chloroplast (trnD-T spacer)			Nuclear (ITS)		
Species	**Fst**	**Manu-> Amigos**	**Amigos-> Manu**	**Fst**	**Manu-> Amigos**	**Amigos-> Manu**
acreana	0	NA – no genetic variation		0.09*	1.77x10-4†	5.07x10-4†
alata	1.00*	2.84x10-8	2.86x10-4†	0	6.05x10-4†	2.39x10-5†
alba	0.64*	1.72x10-7	8.03x10-5†	0.08*	1.73x10-5†	4.63x10-9
bourgonii	0.92*	6.02x10-9	2.80x10-5	0.09*	1.44x10-4†	1.85x10-4†
capitata	0	NA – no genetic variation		0.44*	1.56x10-3	4.57x10-5†
chartacea	0	NA – low sample size		0.45*	NA – low sample size	
edulis	0.91*	1.07x10-8	1.31x10-3	0	4.42x10-4†	3.64x10-6
heterophylla	0.56	NA – low sample size		0	NA – low sample size	
leiocalycina	0.41*	9.54x10-5	4.44x10-4	0.31*	2.10x10-5	4.30x10-5
marginata	0.71*	4.30x10-5†	1.09x10-4†	0.21*	5.03x10-4†	2.65x10-5
morphosp. 75	0	NA – low sample size		0.38*	NA – low sample size	
nobilis	0.05	NA – low sample size		0.06	NA – low sample size	
porcata	0.86	NA – low sample size		0	NA – low sample size	

punctata	0	NA – no genetic variation		0	1.02x10-2	5.93x10-4†
ruiziana	0.81*	5.59x10-9	4.83x10-9	0.01	9.57x10-3†	1.17x10-2†
sapindoides	0.82*	2.58x10-4	1.85x10-5	0.48*	3.63x10-5†	3.00x10-5†
steinbachii	1.00	NA – low sample size		0	NA – low sample size	
suaveolans	0	NA – no genetic variation		0	2.47x10-2†	3.49x10-4
thibaudiana	0	NA – no genetic variation		0.08*	1.12x10-3†	2.38x10-4†
umbellifera	0.29*	4.83x10-9	6.93x10-4†	0.56*	1.46x10-3	6.57x10-5†

*Indicates significant genetic differentiation between Los Amigos and Manu according to permutation analyses conducted in Arlequin v2.000 (Schneider et al. 2000).

†Indicates migration is significantly greater than zero according to nested likelihood ratio tests conducted using the program IMa (Hey and Nielsen 2007).

Table 12: Community-Level Estimates of Chloroplast Gene Flow

Average estimate of community-level migration between or within river basins (from coalescent analyses). This value is the average of the joint posterior estimate of migration from one area into another and the reciprocal estimate.

River Basins Compared	Average Migration Estimate
Between Manu and Los Amigos	1.41x10-3
Within Manu	8.59x10-3
Within Los Amigos	7.48x10-3
Between Los Amigos and Tambopata	3.10x10-3

Table 13: Distance-Decay Analyses within Regions

Analyses of distance-decay within Manu and Los Amigos-Tambopata Regions. The relationship between the Jaccard Similarity Index and geographic distance was assessed using Mantel Tests.

Region	Habitat	R^2	p-value
Manu	Floodplain	0.001	0.433
	Terra Firme	0.127	0.214
Los Amigos-Tambopata	Floodplain	0.047	0.424
	Terra Firme	0.016	0.446

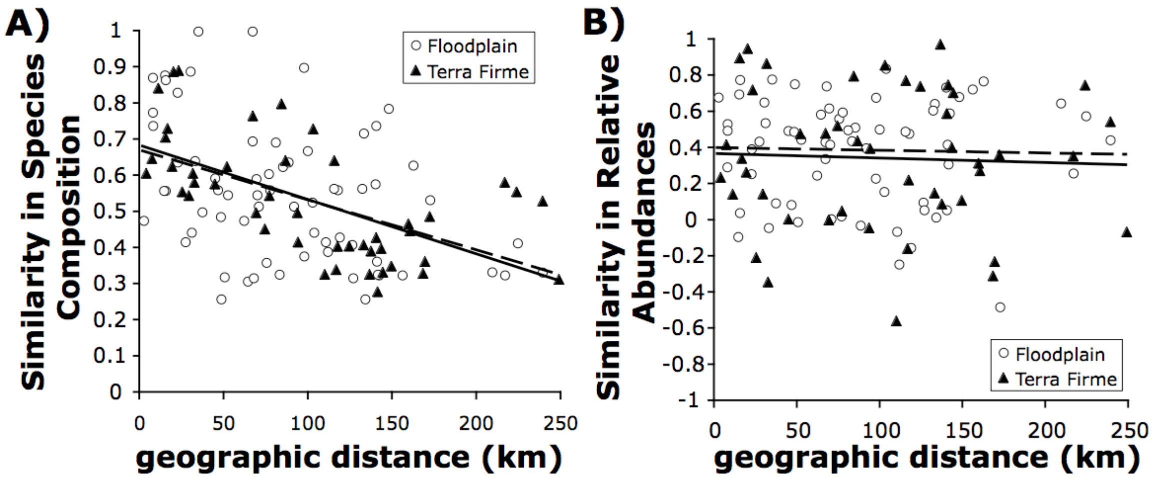


Figure 13: Relationship Between Community Similarity and Geographic Distance

The relationship between geographic distance and **A)** the Jaccard Similarity Index, which reflects the proportion of species shared between communities and **B)** r, the correlation between communities in the relative abundances of shared species. Patterns are shown for terra firme (solid triangles, solid line) and floodplain (open circles, dashed line). Best-fit linear relationships are shown, while Mantel tests were used to assess the significance of the relationship (Jaccard Similarity Index: terra firme: R^2=0.38, p=0.007, floodplain: R^2=0.17, p=0.002; correlation (r): terra firme: R^2=0.00, p=0.366, floodplain: R^2=0.00, p=0.401).

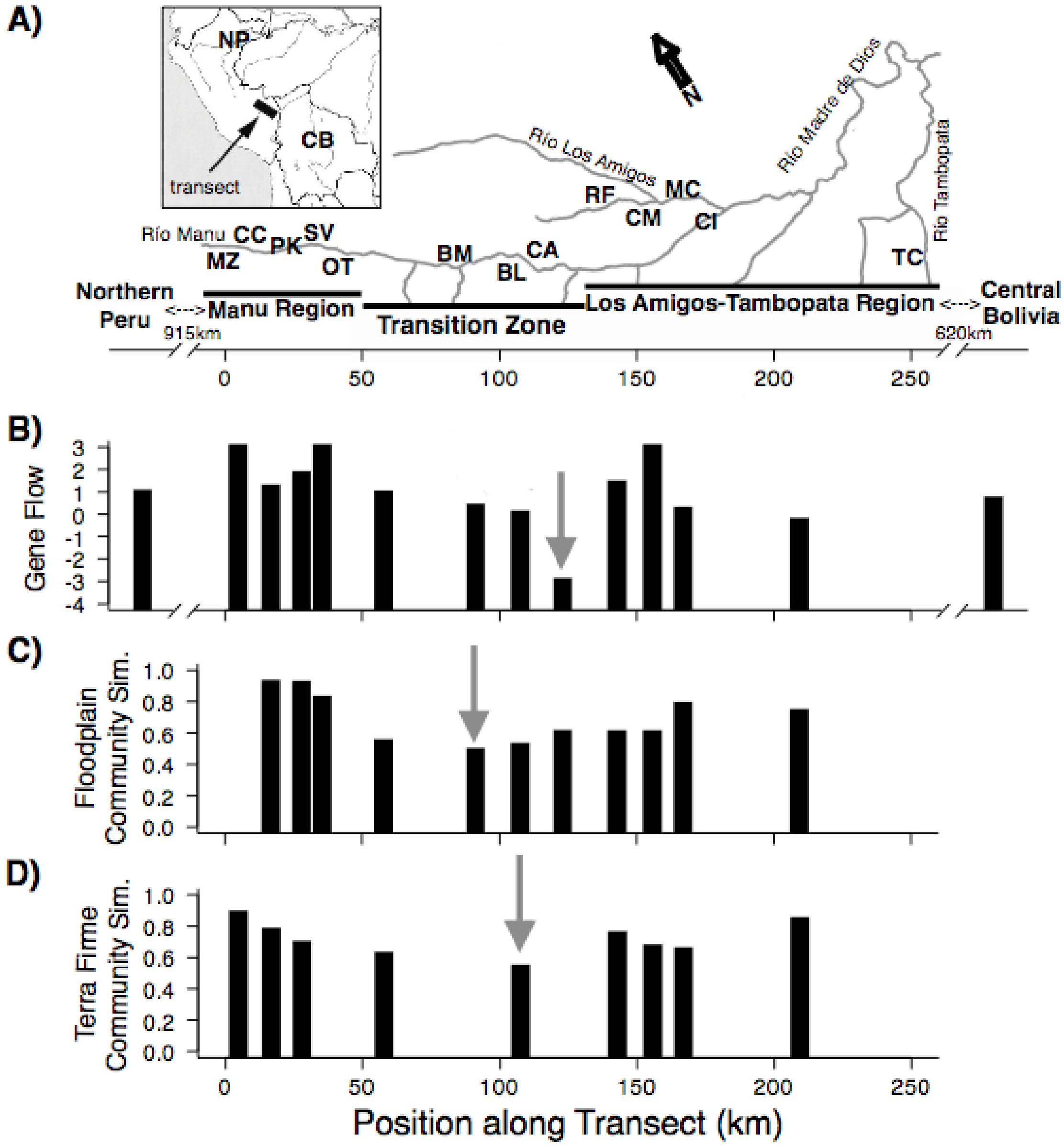


Figure 14: Gene flow and Community Similarity Across a 250km Transect

A) Map of transect showing locations of community surveys and population genetic sampling (see SOM for details); inset shows location of transect in southern Peru and genetic sampling locations in northern Peru (NP) and central Bolivia (CB). **B)** Average chloroplast gene flow, ln(M), between neighboring communities along the transect and with northern Peru and central Bolivia (extreme ends of x-axis). Bars are positioned at the mid-point between neighboring communities. **C)** Jaccard similarity between neighboring floodplain communities along the transect. **D)** same as C, but for terra firme. Gray arrows mark the point of lowest gene flow (B) or lowest community similarity (C & D) along the transect.

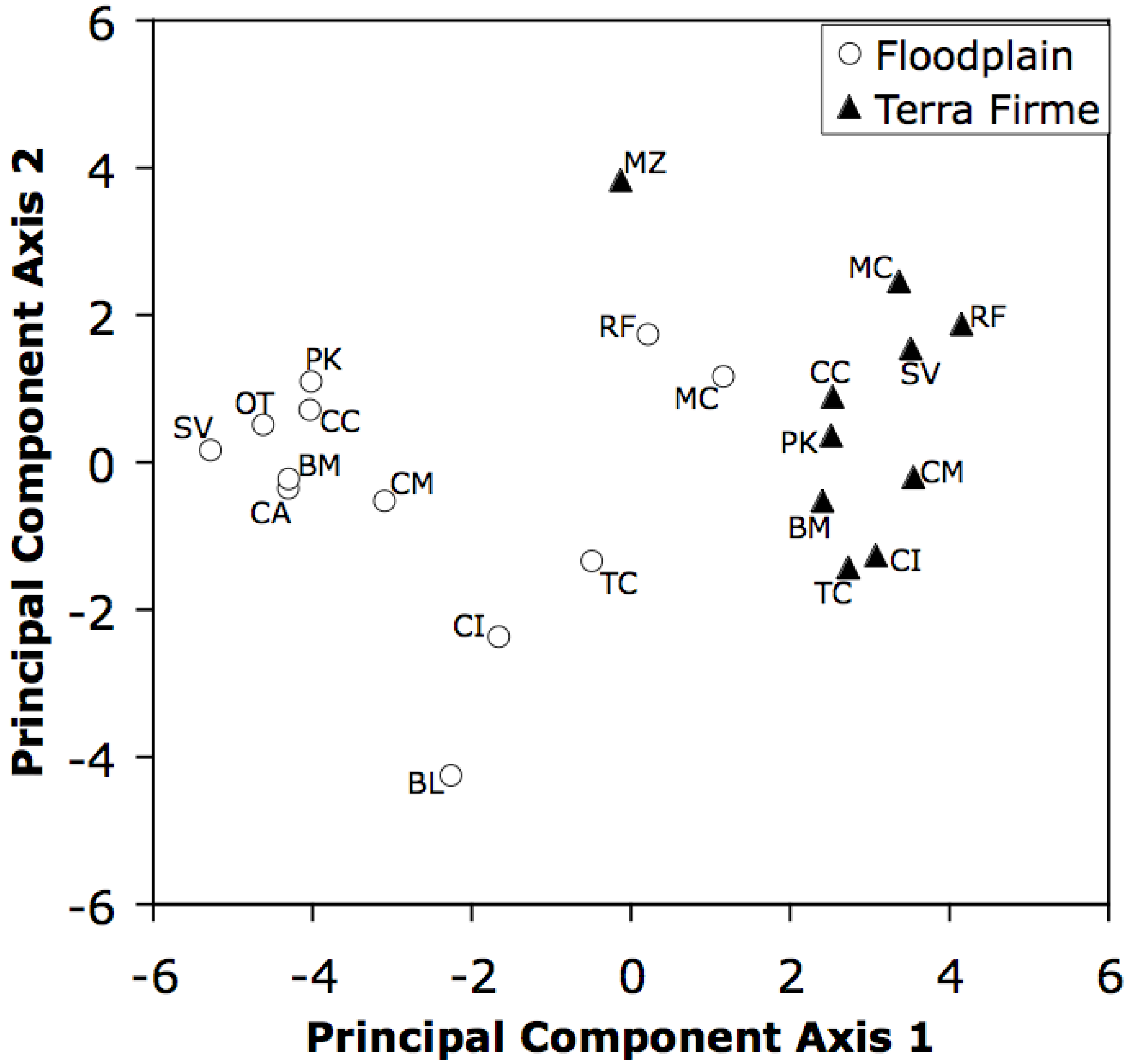


Figure 15: Soil Variation in southern Peruvian rainforests

The first two axes from a principal component analyses of 22 measured soil variables at community survey sites across southern Amazonian Peru. A map of sites is given in Figure 14 and site abbreviations are defined in SOM: Field Methods.

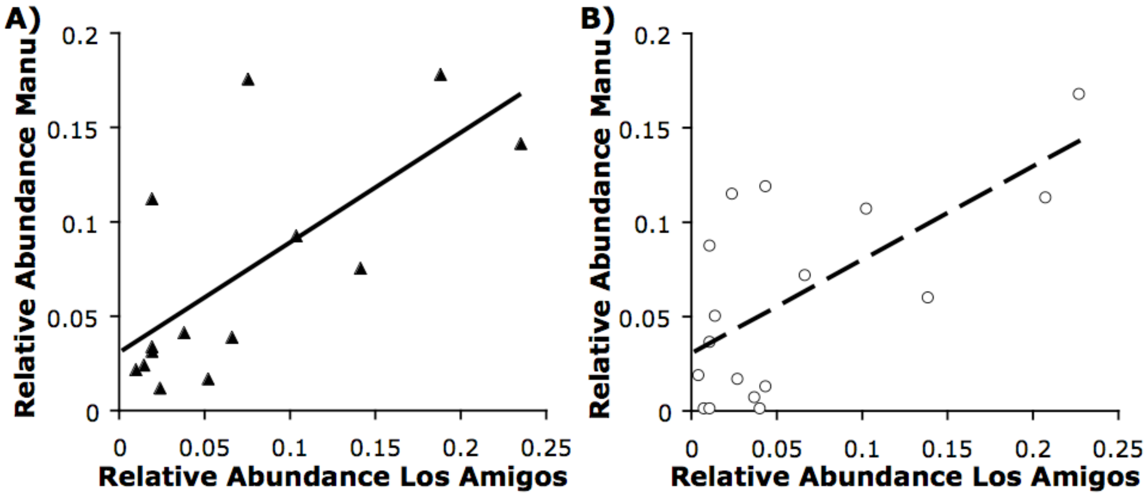


Figure 16: Correlation in Relative Abundances between Manu and Los Amigos

The relative abundances of species in the Manu river basin plotted against their relative abundance in the Los Amigos river basin for **A)** terra firme communities and **B)** floodplain communities. The two communities are significantly correlated in their relative abundances (terra firme: r = 0.70, p = 0.0055; floodplain: r = 0.66, p=0.0041).

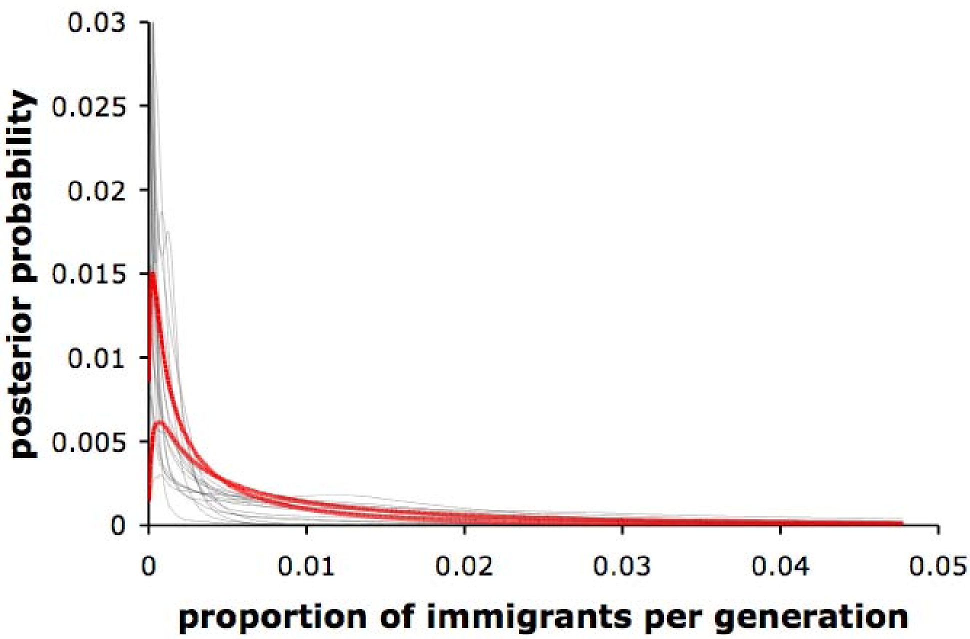


Figure 17: Migration Rates Between Manu and Los Amigos River Basins

Posterior probability distributions for the proportion of migrants per generation from the Manu river basin into the Los Amigos river basin and vice versa based on coalescent analyses of the trnD-T chloroplast locus. Gray lines represent the distributions for individual species, while the red lines represent the distribution from community-level analyses, treating each species as a separate locus in a multi-locus analysis (red line with higher peak: proportion of migrants from Manu into Los Amigos; red line with lower peak: proportion of migrants from Los Amigos into Manu)

References

Alvarez, I. and J. F. Wendel (2003). "Ribosomal ITS sequences and plant phylogenetic inference." Molecular Phylogenetics and Evolution **29**(3): 417-434.

Antongiovanni, M. and J. P. Metzger (2005). "Influence of matrix habitats on the occurrence of insectivorous bird species in Amazonian forest fragments." Biological Conservation **122**(3): 441-451.

Archaux, F., F. Gosselin, et al. (2006). "Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses." Journal of Vegetation Science **17**(3): 299-306.

Barratt, B. I. P., J. G. B. Derraik, et al. (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**(2): 583-590.

Baum, D. A. and K. L. Shaw (1995). Genealogical Perspectives on the Species Problem. Experimental and Molecular Approaches to Plant Biosystematics. P. C. Hoch and A. G. Stephenson. St. Louis, Missouri, Missouri Botanical Garden. **53:** 289-303.

Beerli, P. and J. Felsenstein (2001). "Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach." Proceedings of the National Academy of Sciences of the United States of America **98**(8): 4563-4568.

Bell, G. (2001). "Ecology - Neutral macroecology." Science **293**(5539): 2413-2418.

Bickford, D., D. J. Lohman, et al. (2007). "Cryptic species as a window on diversity and conservation." Trends in Ecology & Evolution **22**(3): 148-155.

Bortolus, A. (2008). "Error cascades in the biological sciences: The unwanted consequences of using bad taxonomy in ecology." Ambio **37**(2): 114-118.

Brown, B., R. M. Emberson, et al. (1999). "Mitochondrial COI and II provide useful markers for Wiseana (Lepidoptera : Hepialidae) species identification." Bulletin of Entomological Research **89**(4): 287-293.

Brown, J. H. (1971). "Mammals on mountaintops: nonequilibrium insular biogeography." American Naturalist **105**(945): 467-&.

Buckley, R. (1982). "The habitat unit model of island biogeography." Journal of Biogeography **9**(4): 339-344.

Caesar, R. M., M. Sorensson, et al. (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**(5): 483-489.

Cannon, C.H. and P.S. Manos (2003). "Phylogeography of the Southeast Asian stone oaks (Lithocarpus)." Journal of Biogeography **30**(2): 211-226.

Case, T. J., J. Faaborg, et al. (1983). "The role of body size in the assembly of West Indian bird communities." Evolution **37**(5): 1062-1074.

Chase, M. W., N. Salamin, et al. (2005). "Land plants and DNA barcodes: short-term and long-term goals." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1889-1895.

Chave, J. and E. G. Leigh (2002). "A spatially explicit neutral model of beta-diversity in tropical forests." Theoretical Population Biology **62**(2): 153-168.

Chave, J., H. C. Muller-Landau, et al. (2002). "Comparing classical community models: Theoretical consequences for patterns of diversity." American Naturalist **159**(1): 1-23.

Cherry, J. L., F. R. Adler, et al. (2002). "Islands, equilibria, and speciation." Science **296**(5570).

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

Condit, R. (1998). Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and Comparison with Other Plots. Berlin, Germany, Springer-Verlag.

Condit, R., N. Pitman, et al. (2002). "Beta-diversity in tropical forest trees." Science **295**(5555): 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. Dynamics of Numbers in Populations. P. J. den Boer and G. R. Gradwell. Wageningen, Netherlands, Center for Agricultural Publication and Documentation: 289-312.

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**(4335): 1302-1310.

Connor, E. F. and E. D. McCoy (1979). "Statistics and biology of the species-area relationship." American Naturalist **113**(6): 791-833.

Cowan, R. S., M. W. Chase, et al. (2006). "300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants." Taxon **55**(3): 611-616.

Davis, J. I. and K. C. Nixon (1992). "Populations, genetic variation, and the delimitation of phylogenetic species." Systematic Biology **41**(4): 421-435.

Demesure, B., N. Sodzi, et al. (1995). "A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants." Molecular Ecology **4**(1): 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, et al. (2002). "Arthropod morphospecies versus taxonomic species: a case study with Araneae, Coleoptera, and Lepidoptera." Conservation Biology **16**(4): 1015-1023.

Diamond, J. M. (1975). Assembly of species communities. Ecology and Evolution of Communities. M. L. Cody and J. M. Diamond. Cambridge, MA, USA, Belknap Press**:** 342-444.

Diamond, J. M. and E. Mayr (1976). "The species-area relation for birds of the Solomon Archipelago." Proceedings of the National Academy of Sciences of the United States of America **73**(1): 262-266.

Doyle, J. J. and J. L. Doyle (1990). "Isolation of plant DNA from fresh tissue." Focus **12**: 13-15.

Duivenvoorden, J. F., J. C. Svenning, et al. (2002). "Ecology - Beta diversity in tropical forests." Science **295**(5555): 636-637.

Espurt, N., P. Baby, et al. (2007). "How does the Nazca Ridge subduction influence the modern Amazonian foreland basin?" Geology **35**(6): 515-518.

Faaborg, J. (1982). "Trophic and size structure of West Indian bird communities." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **79**(5): 1563-1567.

Faaborg, J. (1985). Ecological constraints on West Indian bird distributions. Ornithological Monographs. **36:** 621-653.

Fernandezpalacios, J. M. and C. Andersson (1993). "SPECIES COMPOSITION AND WITHIN ARCHIPELAGO COOCCURRENCE PATTERNS IN THE CANARY-ISLANDS." Ecography **16**(1): 31-36.

Gascon, C., J. R. Malcolm, et al. (2000). "Riverine barriers and the geographic distribution of Amazonian species." Proceedings of the National Academy of Sciences of the United States of America **97**(25): 13672-13677.

Gaston, K. J. and T. M. Blackburn (2000). Pattern and Process in Macroecology. Oxford, UK, Blackwell Science.

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**(2): 571-574.

Gotelli, N. J. and L. G. Abele (1982). "Statistical distributions of West Indian land bird families." Journal of Biogeography **9**(5): 421-435.

Grant, P. R. (1999). Ecology and Evolution of Darwin's Finches. Princeton, N.J., U.S.A., Princeton University Press.

Gurevitch, J., S. M. Scheiner, et al. (2006). The Ecology of Plants. Sunderland, MA, USA, Sinnauer Associates.

Hardy, O. J. and B. Sonke (2004). "Spatial pattern analysis of tree species distribution in a tropical rain forest of Cameroon: assessing the role of limited dispersal and niche differentiation." Forest Ecology and Management **197**(1-3): 191-202.

Hart, M. W. and J. Sunday (2007). "Things fall apart: biological species form unconnected parsimony networks." Biology Letters **3**(5): 509-512.

He, F. L. and P. Legendre (2002). "Species diversity patterns derived from species-area models." Ecology **83**(5): 1185-1198.

Hebert, P. D. N., A. Cywinska, et al. (2003). "Biological identifications through DNA barcodes." Proceedings of the Royal Society of London Series B-Biological Sciences **270**(1512): 313-321.

Hebert, P. D. N., E. H. Penton, et al. (2004). "Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator." Proceedings of the National Academy of Sciences of the United States of America **101**(41): 14812-14817.

Hebert, P. D. N., M. Y. Stoeckle, et al. (2004). "Identification of birds through DNA barcodes." Plos Biology **2**(10): 1657-1663.

Hey, J. and R. Nielsen (2007). "Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics." Proceedings of the National Academy of Sciences of the United States of America **104**(8): 2785-2790.

Hey, J. and J. Wakeley (1997). "A coalescent estimator of the population recombination rate." Genetics **145**(3): 833-846.

Hickerson, M. J., C. P. Meyer, et al. (2006). "DNA barcoding will often fail to discover new animal species over broad parameter space." Systematic Biology **55**(5): 729-739.

Hovestadt, T. and H. J. Poethke (2005). "Dispersal and establishment: spatial patterns and species-area relationships." Diversity and Distributions **11**(4): 333-340.

Hubbell, S. P. (1979). "Tree dispersion, abundance, and diversity in a tropical dry forest." Science **203**(4387): 1299-1309.

Hubbell, S. P. (2001). The Unified Neutral Theory of Biodiversity and Biogeography. Princeton, NJ, USA, Princeton University Press.

Hubbell, S. P., R. B. Foster, et al. (1999). "Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest." Science **283**(5401): 554-557.

Hubbell, S. P. a. F., Robin F. (1986). Biology, chance, and history and the structure of tropical rain forest tree communities. Community Ecology. J. a. C. Diamond, Ted J. New York, NY, USA, Harper and Rowe**:** 314-330.

Hudson, R. R. and N. L. Kaplan (1985). "Statistical properties of the number of recombination events in the history of a sample of DNA sequences." Genetics **111**(1): 147-164.

Jabot, F., R. S. Etienne, et al. (2008). "Reconciling neutral community models and environmental filtering: theory and an empirical test." Oikos **117**(9): 1308-1320.

Janzen, D. H. (1970). "Herbivores and the number of tree species in tropical forests." American Naturalist **104**(940): 501-&.

Janzen, D. H., M. Hajibabaei, et al. (2005). "Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1835-1845.

Jensen, J. L., A. J. Bohonak, et al. (2005). "Isolation by distance, web service." Bmc Genetics **6**: 6.

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**(3): 569-575.

John, R., J. W. Dalling, et al. (2007). "Soil nutrients influence spatial distributions of tropical tree species." Proceedings of the National Academy of Sciences of the United States of America **104**(3): 864-869.

Kalliola, R., J. Salo, et al. (1991). "New site formation and colonizing vegetation in primary succession on the western Amazon floodplains." Journal of Ecology **79**(4): 877-901.

Knowles, L. L. and B. C. Carstens (2007). "Delimiting species without monophyletic gene trees." Systematic Biology **56**(6): 887-895.

Knowlton, N., E. Weil, et al. (1992). "Sibling species in Montastraea annularis, coral bleaching, and the coral climate record." Science **255**(5042): 330-333.

Koptur, S. (1984). "Outcrossing and pollinator limitation of fruit set: breeding systems of neotropical Inga trees (Fabaceae: Mimosoideae)." Evolution **38**(5): 1130-1143.

Krell, F. T. (2004). "Parataxonomy vs. taxonomy in biodiversity studies - pitfalls and applicability of 'morphospecies' sorting." Biodiversity and Conservation **13**(4): 795-812.

Kress, W. J., K. J. Wurdack, et al. (2005). "Use of DNA barcodes to identify flowering plants." Proceedings of the National Academy of Sciences of the United States of America **102**(23): 8369-8374.

Lack, D. L. (1947). Darwin's Finches. Cambridge, U.K., Cambridge University Press.

Lahaye, R., M. Van der Bank, et al. (2008). "DNA barcoding the floras of biodiversity hotspots." Proceedings of the National Academy of Sciences of the United States of America **105**(8): 2923-2928.

Latta, R. G. (2006). "Integrating patterns across multiple genetic markers to infer spatial processes." Landscape Ecology **21**(6): 809-820.

Laurance, S. G. W., P. C. Stouffer, et al. (2004). "Effects of road clearings on movement patterns of understory rainforest birds in central Amazonia." Conservation

Biology **18**(4): 1099-1109.

Lavin, M. (2006). Floristic and geographic stability of discontinuous seasonally dry tropical forests explains patterns of plant phylogeny and endemism. Neotropical Savannas and Seasonally Dry Forests: Plant Biodiversity, Biogeographic Patterns, and Conservation. R. T. Pennington, J. A. Ratter and G. P. Lewis. Boca Raton, FL, USA, CRC Press**:** 433-447.

Lomolino, M. V. (1984). "Mammalian island biogeography: effects of area, isolation, and vagility." Oecologia **61**(3): 376-382.

Losos, E. C. a. L., Egbert G. (2004). Tropical Forest Diverstiy and Dynamism: Findings from a Large-Scale Plot Network. Chicago, IL, USA, University of Chicago Press.

Macarthur, R. H. and E. O. Wilson (1963). "Equilibrium theory of insular zoogeography." Evolution **17**(4): 373-&.

MacArthur, R. H. and E. O. Wilson (1967). The Theory of Island Biogeography. Princeton, NJ, USA, Princeton University Press.

Macia, M. J. and J. C. Svenning (2005). "Oligarchic dominance in western Amazonian plant communities." Journal of Tropical Ecology **21**: 613-626.

Maddison, D. R. and W. P. Maddison (2003). MacClade. Sunderland, MA, USA, Sinnauer Associates.

Martin, T. E. (1981). "SPECIES-AREA SLOPES AND COEFFICIENTS - A CAUTION ON THEIR INTERPRETATION." American Naturalist **118**(6): 823-837.

Matz, M. V. and R. Nielsen (2005). "A likelihood ratio test for species membership based on DNA sequence data." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1969-1974.

McGill, B. J. (2003). "A test of the unified neutral theory of biodiversity." Nature **422**(6934): 881-885.

McGill, B. J., B. A. Maurer, et al. (2006). "Empirical evaluation of neutral theory." Ecology **87**(6): 1411-1423.

McGuinness, K. A. (1984). "Species area relations of communities on intertidal boulders: testing the null hypothesis." Journal of Biogeography **11**(5): 439-456.

Moritz, C. (1994). "Defining Evolutionarily Significant Units for Conservation." Trends in

Ecology & Evolution **9**(10): 373-375.

Morlon, H., G. Chuyong, et al. (2008). "A General Framework for the Distance-Decay of Similarity in Ecological Communities." Ecology Letters **11**: 1-14.

Mouquet, N. and M. Loreau (2002). "Coexistence in metacommunities: The regional similarity hypothesis." American Naturalist **159**(4): 420-426.

Nanney, D. L. (1982). "Genes and phenes in Tetrahymena." Bioscience **32**(10): 783-788.

Nei, M. and W. H. Li (1979). "Mathematical model of studying genetic variation in terms of restriction endonucleases." Proceedings of the National Academy of Sciences of the United States of America **76**(10): 5269-5273.

Nekola, J. C. and P. S. White (1999). "The distance decay of similarity in biogeography and ecology." Journal of Biogeography **26**(4): 867-878.

Newmaster, S. G., A. J. Fazekas, et al. (2006). "DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach." Canadian Journal of Botany-Revue Canadienne De Botanique **84**(3): 335-341.

Nielsen, R. and J. Wakeley (2001). "Distinguishing migration from isolation: A Markov chain Monte Carlo approach." Genetics **158**(2): 885-896.

Norden, N., J. Chave, et al. (2007). "Is temporal variation of seedling communities determined by environment or by seed arrival? A test in a neotropical forest." Journal of Ecology **95**(3): 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**(2): 203-215.

Oliver, I. and A. J. Beattie (1993). "A possible methods for the rapid assessment of biodiversity." Conservation Biology **7**(3): 562-568.

Olivier, J. (2007). Etude spatio-temporelle de la distribution de bambous dans le Sud-ouest amazonien (Sud-Pérou). Histoire, dynamique et futur d'une végétation monodominante en forêt tropicale humide. Laboratoire Evolution et Diversité Biologique. Toulouse, France, Université Paul Sabatier. **Ph.D.:** 420.

Olivier, J., T. Otto, et al. (2008). "First macrofossil evidence of a pre-Holocene thorny bamboo cf. Guadua (Poaceae: Bambusoideae: Bambuseae: Guaduinae) in south-western Amazonia (Madre de Dios — Peru)." Review of Paleobotany and

Palynology.

Pace, N. R. (1997). "A molecular view of microbial diversity and the biosphere." Science **276**(5313): 734-740.

Patton, J. L., M. N. F. Da Silva, et al. (2000). "Mammals of Rio Jurua and the evolutionary and ecological diversification of Amazonia." Bulletin of the American Museum of Natural History **244**: 1-306.

Pennington, T. D. (1997). The Genus Inga: Botany. London, U.K., Royal Botanic Gardens, Kew.

Phillips, O. L., P. N. Vargas, et al. (2003). "Habitat association among Amazonian tree species: a landscape-scale approach." Journal of Ecology **91**(5): 757-775.

Piganeau, G., M. Gardner, et al. (2004). "A broad survey of recombination in animal mitochondria." Molecular Biology and Evolution **21**(12): 2319-2325.

Pitman, N. C. A. (2000). A large-scale inventory of two Amazonian tree communities. Department of Botany. Durham, NC, USA, Duke University. **Ph.D.**

Pitman, N. C. A., H. Mogollon, et al. (2008). "Tree community change across 700 km of lowland Amazonian forest from the Andean foothills to Brazil." Biotropica **40**(5): 525-535.

Pitman, N.C.A., C.E. Ceron, et al. (2005). “Catastrophic natural origin of a species-poor community in the world’s richest forest.” Journal of Tropical Ecology **21**(5): 559-568.

Pitman, N. C. A., J. Terborgh, et al. (1999). "Tree species distributions in an upper Amazonian forest." Ecology **80**(8): 2651-2661.

Pitman, N. C. A., J. W. Terborgh, et al. (2001). "Dominance and distribution of tree species in upper Amazonian terra firme forests." Ecology **82**(8): 2101-2117.

Plotkin, J. B. and H. C. Muller-Landau (2002). "Sampling the species composition of a landscape." Ecology **83**(12): 3344-3356.

Pons, J., T. G. Barraclough, et al. (2006). "Sequence-based species delimitation for the DNA taxonomy of undescribed insects." Systematic Biology **55**(4): 595-609.

Posada, D. and K. A. Crandall (1998). "MODELTEST: testing the model of DNA substitution." Bioinformatics **14**(9): 817-818.

Pregill, G. K. and S. L. Olson (1981). "Zoogeography of West Indian vertebrates in relation to Pleistocene climatic cycles." Annual Review of Ecology and Systematics **12**: 75-98.

Queenborough, S. A., D. Burslem, et al. (2007). "Habitat niche partitioning by 16 species of Myristicaceae in Amazonian Ecuador." Plant Ecology **192**(2): 193-207.

Rach, J., R. DeSalle, et al. (2008). "Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata." Proceedings of the Royal Society B-Biological Sciences **275**(1632): 237-247.

Raxworthy, C. J., C. M. Ingram, et al. (2007). "Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos (Phelsuma) from Madagascar." Systematic Biology **56**(6): 907-923.

Reis, K. R. and D. W. Steadman (1999). "Archaeology of Trants, Montserrat. Part 5. Prehistoric avifauna." Annals of Carnegie Museum **68**(4): 275-287.

Richardson, J. E., R. T. Pennington, et al. (2001). "Rapid diversification of a species-rich genus of neotropical rain forest trees." Science **293**(5538): 2242-2245.

Ricklefs, R. E. (2000). "The relationship between local and regional species richness in birds of the Caribbean Basin." Journal of Animal Ecology **69**(6): 1111-1116.

Ricklefs, R. E. and E. Bermingham (1999). "Taxon cycles in the Lesser Antillean avifauna." Ostrich **70**(1): 49-59.

Ricklefs, R. E. and E. Bermingham (2001). "Nonequilibrium diversity dynamics of the Lesser Antillean avifauna." Science **294**(5546): 1522-1524.

Ricklefs, R. E. and E. Bermingham (2004). "Application of Johnson et al.'s speciation threshold model to apparent colonization times of island biotas." Evolution **58**(8): 1664-1673.

Ricklefs, R. E. and E. Bermingham (2004). "History and the species-area relationship in Lesser Antillean birds." American Naturalist **163**(2): 227-239.

Ricklefs, R. E. and E. Bermingham (2007). The causes of evolutionary radiations in archipelagoes: passerine birds in the Lesser Antilles. The American Naturalist. **169:** 285-297.

Ricklefs, R. E. and G. W. Cox (1972). "Taxon cycles in West Indian avifauna." American Naturalist **106**(948): 195-&.

Ricklefs, R. E. and G. W. Cox (1978). "Stage of taxon cycle, habitat distribution, and population density in the avifauna of the West Indies." American Naturalist **112**(987): 875-895.

Ricklefs, R. E. and I. J. Lovette (1999). "The roles of island area per se and habitat diversity in the species-area relationships of four Lesser Antillean faunal groups." Journal of Animal Ecology **68**(6): 1142-1160.

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**(6): 924-942.

Ronquist, F. and J. P. Huelsenbeck (2003). "MrBayes 3: Bayesian phylogenetic inference under mixed models." Bioinformatics **19**(12): 1572-1574.

Rosenzweig, M. L. (1995). Species Diversity in Space and Time. Cambridge, UK, Cambridge University Press.

Rossetti, D. D., P. M. de Toledo, et al. (2005). "New geological framework for Western Amazonia (Brazil) and implications for biogeography and evolution." Quaternary Research **63**(1): 78-89.

Ryan, P. G., P. Bloomer, et al. (2007). "Ecological speciation in South Atlantic island finches." Science **315**(5817): 1420-1423.

Savolainen, V., R. S. Cowan, et al. (2005). "Towards writing the encyclopaedia of life: an introduction to DNA barcoding." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1805-1811.

Schneider, S., D. Roessli, et al. (2000). ARLEQUIN Ver. 2.000. Geneva, Switzerland, Genetics and Biometry Laboratory, University of Geneva.

Schoener, T. W. (1976). The species-area relation wtihin archipelagos: models and evidence from island land birds. Proceedings of the 16th International Ornithological Conference. H. J. Firth and J. H. Calaby. Canberra, Australia, Australian Academy of Science.

Scott, W. A. and C. J. Hallam (2003). "Assessing species misidentification rates through quality assurance of vegetation monitoring." Plant Ecology **165**(1): 101-115.

Sheil, D. (1995). "A critique of permanent plot methods and analysis with examples from Budongo Forest, Uganda." Forest Ecology and Management **77**(1-3): 11-34.

Slatkin, M. (1993). "Isolation by distance in equilibrium and nonequilibrium populations." Evolution **47**(1): 264-279.

Slatkin, M. and L. Voelm (1991). "Fst in a hierarchical island model." Genetics **127**(3): 627-629.

Sodhi, N. S., L. H. Liow, et al. (2004). "Avian extinctions from tropical and subtropical forests." Annual Review of Ecology Evolution and Systematics **35**: 323-345.

Stanford, A. M., R. Harden, et al. (2000). "Phylogeny and biogeography of Juglans (Juglandaceae) based on matK and ITS sequence data." American Journal of Botany **87**(6): 872-882.

Stephens, M., N. J. Smith, et al. (2001). "A new statistical method for haplotype reconstruction from population data." American Journal of Human Genetics **68**(4): 978-989.

Strong, D. R. and D. A. Levin (1975). "Species richness of parasitic fungi of British trees." Proceedings of the National Academy of Sciences of the United States of America **72**(6): 2116-2119.

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**(1): 131-141.

Swofford, D. L. (2002). PAUP. Sunderland, MA, USA, Sinnauer Associates.

Terborgh, J. (1973). "Chance, habitat, and dispeasl in distribution of birds in the West Indies." Evolution **27**(2): 338-349.

Terborgh, J., J. Faaborg, et al. (1978). "Island Colonization by Lesser Antillean Birds." Auk **95**(1): 59-72.

Terborgh, J., R. B. Foster, et al. (1996). "Tropical tree communities: A test of the nonequilibrium hypothesis." Ecology **77**(2): 561-567.

Terborgh, J. W. and J. Faaborg (1980). "Saturation of bird communities in the West Indies." American Naturalist **116**(2): 178-195.

Tilman, D. (1994). "Competition and biodiversity in spatially structured habitats." Ecology **75**(1): 2-16.

Tuomisto, H. and K. Ruokolainen (2006). "Analyzing or explaining beta diversity?

Understanding the targets of different methods of analysis." Ecology **87**(11): 2697-2708.

Tuomisto, H., K. Ruokolainen, et al. (2003). "Dispersal, environment, and floristic variation of western Amazonian forests." Science **299**(5604): 241-244.

Vecchione, M., W. F. Mickevich, et al. (2000). "Importance of assessing taxonomic adequacy in determining fishing effects on marine biodiversity." Ices Journal of Marine Science **57**(3): 677-681.

Volkov, I., J. R. Banavar, et al. (2003). "Neutral theory and relative species abundance in ecology." Nature **424**(6952): 1035-1037.

Vormisto, J., J. C. Svenning, et al. (2004). "Diversity and dominance in palm (Arecaceae) communities in terra firme forests in the western Amazon basin." Journal of Ecology **92**(4): 577-588.

Webb, C. O. (2000). "Exploring the phylogenetic structure of ecological communities: An example for rain forest trees." American Naturalist **156**(2): 145-155.

White, T. J., T. Bruns, et al. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols: A Guide to Methods and Applications. M. Innis, D. Gelfand, J. Sninsky and T. White. San Diego, CA, USA, Academic Press**:** 315-322.

Whittaker, R. J. (2007). Island Biogeography: Ecology, Evolution, and Conservation. Oxford, UK, Oxford University Press.

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**(1): 69-91.

Wills, C., R. Condit, et al. (1997). "Strong density- and diversity-related effects help to maintain tree species diversity in a neotropical forest." Proceedings of the National Academy of Sciences of the United States of America **94**(4): 1252-1257.

Woodcock, S., T. P. Curtis, et al. (2006). "Taxa-area relationships for microbes: the unsampled and the unseen." Ecology Letters **9**(7): 805-812.

Wright, S. J. (1981). "Intra-archipelago vertebrate distributions: the slope of the species-area relation." American Naturalist **118**(5): 726-748.

Zhang, Q., Y. Liu, et al. (2003). "Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species." Plant and Cell Physiology **44**(9): 941-951.

Zillio, T. and R. Condit (2007). "The impact of neutrality, niche differentiation and species input on diversity and abundance distributions." Oikos **116**(6): 931-940.

Zwickl, D. J. (2006). Genetic algorithm approaches for the phylogenetic analysis of large biological datasets under the maximum likelihood criterion. Austin, University of Texas. Ph.D.

Biography

Kyle Dexter was born on August 19th, 1980 in Richmond, Virginia, U.S.A. He attended the University of Virginia (UVA), graduating in 2002 with a B.A. in Biology and Environmental Sciences. While at UVA, Kyle honed his interests in biological research, studying the behavior of ring-tailed lemurs, feeding ecology of hippopotami, and intestinal parasitism in field mice. After graduating from UVA, Kyle spent a year studying tropical tree community dynamics in the Peruvian Amazon.

While in graduate school at Duke University, Kyle received several awards and grants in support of his research. They include the following: a National Science Foundation Pre-Doctoral Fellowship, a James B. Duke Fellowship from Duke University, a National Science Foundation Dissertation Improvement Grant, and grants from the American Philosophical Society, Sigma Xi, Society of Systematic Biologists, Duke University Graduate School, Duke Biology Department, Explorer's Club, Organization for Tropical Studies, Consortium for Latin American and Caribbean Studies, and Amazon Conservation Association.

In 2008, Kyle Dexter received fellowship offers from the German Academic Exchange Service (DAAD) and the French National Scientific Research Center (CNRS) as well as a final year fellowship offer from Duke University. Kyle ultimately decided it was time to graduate and will begin a CNRS post-doctoral fellowship with Jerome Chave in Toulouse, France in December of 2008.