Estimating seed dispersal distances with incomplete genetic data: new methods, power analyses and a case study of the tropical tree *Tabebuia rosea*

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Marta Vargas Timchenko

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Helene Muller-Landau

December 2012

Acknowledgments

I would like to thank my advisor Helene Muller-Landau and committee members Jeannine Cavender-Bares and Rebecca Montgomery for their helpful guidance and advice. I thank the E. Bermingham lab members at STRI, especially Kristin Saltonstall for comments on manuscript and methods. I would also like to acknowledge F. Andy Jones for help with sample collections, Emily Moran for advice on modeling, Noelle Beckman for advice on manuscript, Pablo Ramos the seed trap data collection, and David Brassfield for tree reproductive status data. I gratefully acknowledge funding from the National Science Foundation (DEB 0453445 and DEB 0453665), the University of Minnesota, the Smithsonian Tropical Research Institute, and the Fulbright-LASPAU scholarship program.

Abstract

Genetic data linking seeds directly to parents through maternal seed tissue are often hailed as the best way to obtain information on seed dispersal distances. However, DNA quality in maternally derived seed tissue is often low, leading to high rates of genotyping errors, and usually much discarding of data. This study tested and applied methods for gleaning information on seed dispersal distances from incomplete and error-prone genetic data, using the tropical tree *Tabebuia rosea* as a case study. Genotyping error rates were calculated using PEDANT and then incorporated these rates into a model to estimate seed dispersal distances using all available data. Simulations were used to evaluate the effects of both genotyping error rates and the number of seeds genotyped upon dispersal estimates. Simulations showed that for our system, even datasets as large as 1000 genotyped seeds, more than any previous genetic study of seed dispersal distances, are not sufficient to estimate true dispersal distances within 10% with 90% confidence. Realistic levels of genotyping errors for molecular markers with the resolution of those used here decrease the information content, such that approximately twice as many seeds are needed to obtain the same precision. Results demonstrate the importance of calculating error rates, and the value of including incomplete genetic data in analyses in order to increase power and obtain better parameter estimates. The data set used to estimate the dispersal parameters for *T. rosea* in this study (181 genotyped seeds) proved to be insufficient for the model to provide usefully precise estimates of seed dispersal parameters. The genetic data did provide useful information on directionality, showing significant bias towards due South, consistent with the prevailing northerly winds during the seed dispersal season of the focal species.

Table of Contents

List of Tables

List of Figures

Introduction

Seed dispersal is an important ecological process that involves the movement of seeds from the maternal source to a deposition site. Such movement creates spatial patterns of dispersed seeds that can strongly influence population and community structure. The patterns of seed deposition, created by the dispersal process, serve as an initial blueprint upon which post-dispersal processes such as density-dependent mortality and competition act to shape plant communities (Janzen 1970, Connell 1971, Dalling et al 1998). Despite its importance, seed dispersal remains poorly understood because it is difficult to get direct estimates of dispersal distances in natural populations. In many if not most populations, there is extensive overlap in seed shadows among neighboring conspecific plants, which complicates the assignment of dispersed seeds to their maternal sources (Jones and Muller-Landau 2008).

Genetic methods as a tool for studies of seed dispersal

In general, seed dispersal can be studied directly, by determining the origin of dispersed seeds in the landscape or using indirect methods such as inverse modeling. Some ways of looking at seed dispersal directly include examining post-dispersal variation in seed densities around isolated trees (e.g., Dalling et al 1998), relocating marked seeds around their sources (Van Der Wall 1994, Turchin 1998, Xiao et al 2006), or radiotracking individual seeds (Pons and Pausas 2007). But most often it is done indirectly, by looking at the spatial patterns of dispersed seeds and inferring their origin. Ever since Ribbens et al (1994) introduced inverse modeling methods for fitting models for seed dispersal to

seed deposition patterns in stands with multiple source plants, these methods have subsequently been widely applied (Muller-Landau et al 2002, Clark et al 2005).

The development of higher resolution microsatellite and AFLP markers made it possible to use genetic markers to establish the identity of the sources of seeds collected postdispersal (Dow et al 1996, Godoy 2001, Robledo-Arnuncio and Garcia 2007, Jones et al 2005). These markers enable direct assignment of dispersed seeds or seedlings to parents and thus direct calculation of seed dispersal distances for individual seeds or seedlings. Such development led to wider use of genetic markers in seed dispersal studies, and was seen by many as a breakthrough that would revolutionize studies of seed dispersal (reviewed in Ashley et al 2010). Thereafter many studies of seed dispersal have used paternity analyses on seeds (direct estimation) or established seedlings (indirect estimation). For example, in an early paper, Dow et al (1996) use microsatellites to infer dispersal distances for both seeds and pollen in *Quercus*. More recently, similar approaches have been taken by others (Dow et al 2008, Sebben et al 2011, Choo et al 2012). One obvious drawback of this technique is that it requires knowledge of the genotypes of both parents, and in many cases cannot differentiate mother from father, thus potentially confounding seed and pollen dispersal distances. Typically these studies assume that the mother is always the closest of two or more parent matches, an assumption that is almost certain to be incorrect in some cases.

More recently, a number studies have applied molecular techniques to maternally derived tissue on the seed or diaspore in order to establish the identity of the maternal source of seeds collected postdispersal (Godoy 2001, Jones et al 2005, Robledo-Arnuncio and Garcia 2007, Smouse et al 2012). The advantage of this technique compared to traditional paternity analysis is that there is no possibility of confusing paternal and maternal parents: it more definitively identifies the mother tree, and thus the seed dispersal distance. However, this method suffers from the potentially low quality of DNA in maternally derived tissues, and can result in genotyping errors as described below.

What are the advantages of using genetic methods vs. inverse modeling? In some cases, genetic markers have revealed cryptic dispersal that could easily be missed using other methods – for example, the deposition of seeds under conspecifics that are not the mother plant (Bullock et al 2006). High-resolution genetic methods can also potentially distinguish seeds immigrating from outside the studied population, providing more accurate measurement of long-distance dispersal (Jones and Muller-Landau 2008). However, as shown by Robledo-Arnuncio and Garcia (2007), and Jones and Muller-Landau (2008), genetic data alone does not protect against major errors in estimation of seed dispersal distances – inverse modeling is still needed to provide unbiased estimates for most sampling designs.

The treatment of genotyping errors

Though genetic methods have great potential to aid studies of seed dispersal, they also have inherent drawbacks: they are expensive, labor-intensive and susceptible to errors resulting from DNA degradation, poor handling and equipment malfunction. Though ecologists often think of genetic markers as providing the final answer on genetic relatedness of studied individuals, all genetic methods are subject to error, and even small error rates have important implications for interpreting results. For example, an error rate of 1% per locus in a 12-locus genotype means that 21% of diploid genotypes contain at least one error, and this in turn means that only 62% of identity comparisons are correctly characterized as identical (Hoffman and Amos 2005). In practice, error rates can be even larger, as many studies are designed using highly polymorphic loci, and evidence suggests that those are more error prone than loci with lower allelic diversity (Hoffman and Amos 2005). Because error in estimating true genotypes can significantly affect the success of a study, the development of methods for dealing with such errors has been the subject of a number of recent papers (Bonin et al 2004, Pompanon et al 2005, Dewoody et al 2006).

The precision of the genotyping process depends greatly on the quantity and quality of the available DNA. Low DNA quality or quantity has been shown to result in a number of errors, including allelic dropout and failed amplification. Methods of direct parental assignment using seed exocarps are potentially especially prone to such errors. DNA is generally present in low quantities in seed exocarps, and those are in most instances dry

or hard and hence difficult to handle, especially for small-seeded plants. It is common practice to deal with these problems by replicating analyses (Pompanon et al 2005, Morin et al 2009) and throwing out data on incomplete genotypes and those that did not replicate correctly (Jones et al 2005). Yet even incomplete genotypes and those with high error rates provide important information on the probability of different assignments. It is not cost-effective to discard such data, especially because most seed dispersal studies are designed using a limited number of seeds and this number is generally restricted by the availability of resources for genotyping. Further, even after throwing out inconsistently replicated genotypes, it is problematic to assume the remaining data is error-free, as is commonly done (e.g., Jones and Muller-Landau 2008), as it is never possible to eliminate error completely, and just one or a few erroneous outliers can have huge weight in analyses.

To enable comparison of error rate estimates among studies, Broquet and Petit (2004) classify genotyping errors into two types: allelic dropouts, defined as one of the alleles on a heterozygote failing to amplify resulting in a false homozygote, and false alleles, which are the result of slippage amplification or cross contamination. Both types of errors are inherent to PCR based techniques. Pompanon et al (2005) review causes and consequences of genotyping errors; one of the consequences they identify is incorrect parental exclusion in paternity analysis. Johnson and Haydon (2007) further stress the need to take genotyping errors into account in parentage analysis or direct assignment studies. They find a notable lack of consistency in how errors are identified and

estimated in different studies, which leads to a lack of comparability. As a result, they develop a maximum likelihood method, implemented in the software program PEDANT, for calculating the two types of errors (namely, allelic dropout and false alleles); these methods have been used in a number of subsequent studies (Kolomyjec 2009, Hauser et al 2011, Hess et al 2012, Llorens 2012).

Given limited resources for genetic studies of seed dispersal and the inherent risks of genotyping errors, it is crucial to consider errors when planning the sampling design, and to choose methods that appropriately deal with expected and observed errors. As discussed above, failure to account for nonzero error rates can lead to flawed conclusions.

Modeling seed dispersal

Mean dispersal distance, shape of distribution of dispersal distances, and directionality are all important aspects of the seed dispersal patterns. Regardless of whether or not genetic data is used, the seed dispersal of a given species is most often characterized as the function that describes the probability distribution of dispersal distances, dispersal kernel. The dispersal kernel is commonly estimated from seed trap data using inverse modeling techniques. The shape of the kernel is as important as estimates of dispersal distances because different points of this distribution govern entirely different ecological processes acting upon the population, for instance the density dependant mortality, at the mean of the distribution, or immigration, at the tail of the distribution. In a more general sense the dispersal kernel describes the movement of seeds over space and is of major

importance for mathematical modeling of plant populations, especially during population expansion.

Though dispersal kernels are often modeled as isotropic, directionality can also be important (Skarpaas et al 2007, Sanchez et al 2011, Van Putten et al 2012). Different dispersal vectors can create different patterns of directionality. Seed dispersal by wind is expected to be strongly directional when and where winds during the season of seed dispersal are strongly directional. Seeds dispersed by mammals may also show significant directionality for any given tree, because animal movement is not random in space and depends on the surrounding neighborhood of the tree and the behavior of the animal (Garcia et al 2007).

Objectives

We develop, test and apply methods for gleaning information on seed dispersal distances from incomplete and error-prone genetic data. We review the literature on previous seed dispersal studies using direct parental assignments from genetic markers for information relevant to sampling design and the treatment of genotyping errors. We develop methods to fit seed dispersal models based upon the frequencies of individual genotypes in each seed trap and prior information on error rates, using a multinomial error distribution, the competing sources model (**CSM**, Robledo-Arnuncio and García 2007), implemented with simultaneous estimation of fecundity (Jones and Muller-Landau 2008). We validate and test the power of these methods through simulation, and we present a case study of the

tropical tree *Tabebuia rosea*, in which we estimate seed dispersal distances using all available data (genetic and seed trap). In addition, we use basic statistics to investigate directional patters of seed dispersal..

Materials and Methods

Literature review

We searched the literature for every published study in which maternal seed tissue was genotyped to enable assignment to maternal parents that appeared before September 2012. We specifically examined every paper that cited Jordano and Godoy (2001), the first study of this kind, and also conducted keyword searchers. For each study, we tabulated the number of seeds collected and genotyping error rates when reported.

Study site and species

The study was conducted on Barro Colorado Island (9º9' N, 79 º51' W), an artificial island located in Gatun Lake in the Panama Canal (Figure 1). Annual precipitation averages 2600 mm, with a dry season from mid-December through mid-April. The island is covered by lowland tropical moist forest, approximately half old-growth, and half old secondary forest 80-120 years old.

Tabebuia rosea (Bignoniaceae) is a common Neotropical canopy tree ranging from Mexico to Colombia and Venezuela, and is also present in the West Indies. It grows at

elevations from sea level to 1,500 m. In Panama, it is a characteristic species of tropical dry, tropical moist, and premontane moist forests (Tosi 1971). *T. rosea* is frequent on Barro Colorado Island, with a density of 1.4 trees over 10 cm dbh per ha in the 50 ha Forest Dynamics Plot. This species is wind-dispersed and shade-intolerant. It is more abundant in the younger forest on Barro Colorado. Leaves fall in the dry season right before flowering and the tree produces trumpet-like pink flowers that are most abundant in March and April (sometimes in February). Seeds mature in the late dry season and are released in the late dry and early rainy season. The fruit is a dehiscent capsule 25-35 cm long; the seeds are hyaline wings (Croat 1978).

Tree and seed censuses

The study area included five rectangular plots, each 6 ha in area, located around telemetry towers (part of the Automated Radio Telemetry System, ARTS). The plots extended 100 m east, west, and south of the towers, and 200 m to the north (during the dry season, prevailing winds come from the north, so the longer plot extension in this direction was designed to incorporate more potential parent trees of seeds arriving near the tower). Arrangement of the elements within the plot is showed in Figure 2.

In these plots, all trees over 20 cm DBH were identified, tagged and mapped in 2004. Reproductive status of each tree of the focal species was assessed in 2005, 2006 and 2007 through visual observation of the crown (from the ground using binoculars) during the fruiting season. Reproductive status was scored using an ordinal scale: 0 for

nonreproductive, 1 to represent one quarter of the crown filled with seeds, 2 and 3 to represent two and three quarters respectively, or 4 to represent a crown completely filled with fruit.

In each plot, 49 square seed traps made of mesh and PVC pipes, each with area 0.5 m^2 and located at a height of 1.0 m above the ground were arrayed in a 7 x 7 grid in the central ha around the radio tower in each plot. The traps were censused weekly between March 2005 and April 2007, and all seeds of the focal species that fell during that period were collected. Maternal tissues of the collected seeds (seed wings or exocarps) were removed and the seeds themselves were discarded. Individual seed wings were placed in microcentrifuge tubes and stored at -20 C degrees until DNA was extracted.

In order to identify the adult sources of dispersed seeds, all potential parent trees of the focal species in the plots were sampled and genotyped. Green leaves were collected from all available *T. rosea* adults using a sling shoot. Leaves were dried using Drierite and stored upon extractions. Initial leaf collection occurred in 2008 (collected by F.A. Jones); several trees were recollected in November 2009 due to problems with amplifying DNA from the original collections. Leaves were successfully collected from 68 trees; another 25 trees that were alive at the time of the original plot tree censuses died or were not reproductive at the time of seed or leaf collection. Note that trees were only sampled within the five 6-ha plots, not taking into account potential parents outside the plots.

DNA extractions and genotyping

We used a modified CTAB extraction protocol, optimized for acidic plant tissues, for adult leaf tissues (Doyle and Doyle 1990, Warude 2003). We homogenized 100-200 mg of leaf tissue in 2 mL tubes in a TissueLyzer (Qiagen) with two stainless steel beads. Powdered tissues were incubated with 800 *u*l of modified CTAB buffer solution at 65° C for 30 minutes, extracted with an equal volume of chloroform:IAA (24:1), and precipitated from the aqueous layer using chilled ethanol and 3M sodium acetate. DNA was washed using 70% ethanol then resuspended in TE buffer after drying.

Maternally derived tissue from the seed wing of each seed (all available tissue) was homogenized in 96 well plates in TissueLyzer with two stainless steel beads. DNA was extracted from the resulting powder using the DNease Plant 96 Plate Kit (Qiagen), with a final elution volume of 50 *u*l.

Genotyping was performed using 4 microsatellite loci developed for *Tabebuia aurea* (Tau17, Tau21, Tau 22 and Tau 31; Braga et al 2007). The polymerase chain reaction (PCR) cocktail included 1x buffer, 2mM $MgCl_2$, 100 *u*M each dNTP, 1.29% BSA, 1% DMSO, 0.36 *u*M of each primer, 0.05 units of AmpliTaq Gold polymerase (Applied Biosystems) and 1 ul DNA template for adults and 2 *u*l for seeds. This increase in DNA template for seed wings was to compensate for lower concentration of extracts. The thermal cycle was 12 min at 94° C, 35 cycles of 30 seconds at 94° C, 30 seconds at 56° C, and 45 seconds at 72° C, and a final elongation step of 2 minutes at 72° C. At least

two replicates were run for each parent tree and seed coat; however, some samples were run up to 4 times due to low PCR success or suspected contamination. The SSR primers were labeled with fluorescent dyes and electrophoresed on an ABI Automated Sequencer 3130XL with LIZ 500 size standard. Genotypes were scored using GeneMapper 3.7 (Applied Biosystems). Basic genetic statistics were calculated using FSTAT 2.9.3. To ensure we used chose only informative microsatellite primers we included in this calculation 87 seeds collected in traps located in the radio towers within the plots, those where not used in dispersal kernel estimation.

Identity analysis

In order to calculate the number of genotyped seeds that matched adults inside the plots, we used the identity test option in Cervus 3.0 (Kalinowski et al 2007). All complete and incomplete genotypes were included. As no adult trees shared the same multilocus genotype, parentage assignment of seeds was not biased by assignment to the closest trees (Jones and Hubbell 2006). Partial matching with one locus mismatch was allowed because high error rates in the data were suspected (see below). This kind of partial matching is less powerful for incomplete genotypes, as they provide less information, but could be very useful for complete ones.

Error rate calculations

To estimate error rates, we replicated PCR and genotyping. At least two and as many as six replicates were run for each individual extract (each individual seed or adult).

Replicate data were used to calculate maximum likelihood estimates of allele drop and false alleles probabilities for each locus and for each population (seeds or adults) using PEDANT (Johnson and Haydon 2007). We first made a matrix of all possible replicate genotype pairs for a given sample. To avoid overrepresentation of samples having more replicates, we created a master dataset that included multiple copies of pairwise comparisons from samples with fewer replicates to achieve equal number of pairwise comparisons for every sample. We then used random subsets of the data for each sample in different runs of PEDANT.

Patterns of dispersal distance and direction in genetically matched seeds

For each genotyped seed that was matched to a particular maternal parent at four or three loci, we calculated the angle direction and the distance between the mother tree and the trap where the seed was collected. No mismatches where allowed. This angle direction (in radians) was used to construct circular histograms in order to look for directional patterns. Tree to seed distance was used to construct histograms to look at the distribution of seed dispersal distances for those matches. Mean circular direction, circular standard deviation, mean dispersal distance and standard deviation where calculated for full matches at four or three loci. As seeds that fall right under the mother tree are likely to show less directionality, we also did these analysis including only seeds with dispersal distances greater than 20 m.

Inverse modeling of seed production and dispersal

We used maximum likelihood to estimate seed production and dispersal parameters from data on seed arrival, seed and tree locations, and seed and tree observed genotypes, given genotyping errors. Our analysis parallels the Bayesian approach used by Moran and Clark (2011).

Seed production, or fecundity, of each tree was assumed to be proportional to its reproductive status times its basal area. That is, the fecundity f_{qi} of tree i in plot q was modeled as

$$f_{qi} = Br_{qi}D_{qi}^2$$

(eq .1)

where B is a fitted fecundity parameter, r_{qi} is the reproductive status (0, 1, 2, 3 or 4), and D_{qj} is the diameter of tree i in stand q in units of millimeters. For many trees, $r = 0$ because no seeds were observed during the reproductive tree censuses. To avoid potentially very problematic errors of omission, for these individuals we assigned the value $rD^2 = 1$.

We assume seeds disperse according to an isotropic dispersal kernel taking the form of a 2-dimensional Student's T distribution with degrees of freedom parameter set to 3 (Clark et al 1999, Muller-Landau et al 2008). Specifically, the two-dimensional dispersal kernel p giving the probability of seed arrival per unit area is

$$P(d) = \frac{1}{\pi u \left(1 + \frac{d^2}{u}\right)^2}$$

(eq. 2)

where d is the distance from the source, and u is a fitted parameter.

Expected total seed rain into trap j in stand q, θ_{qj}, is thus calculated as the sum of all source trees within the plot, plus expected immigrate seed rain from outside the plot:

$$\theta_{qj} = A\left[\sum_{I_q} Br_{qi} D_{qi}^2 p\left(d_{qij} \middle| u_s\right) + O_{qj}\right]$$

(eq. 3)

where A is trap area, d_{qij} is the distance between tree i and trap j in plot q, and the second term O_{qj} gives the expected seed rain per unit area from outside the plot. Expected seed rain from out-of-plot trees is calculated by assuming that the average density and fecundity of trees outside the plot is equal to the average within-plot density and fecundity, summed across all the plots.

We assumed the observed number of seeds in trap j in stand q, S_{qj}, follows a Poisson distribution around the expected:

$$S_{qj} \sim Pois\left(\theta_{qj}\right)$$

(eq. 4)

Under this assumption, the likelihood of all the observed seed arrivals into traps in traps in plot q, L_{qu} (the subscript u standards for ungenotyped, as this takes no account of genetic data), is

$$L_{qu} = \prod_{Jq} Pois\left(S_{qj}|\theta_{qj}\right)$$

(eq. 5)

To take account of the genotype data, we calculated the expected proportion of seeds with observed genotype k in trap j in stand q, π_{qjk}:

$$\pi_{qjk} = \frac{\sum_{I_q} f_{qi} p(d_{qij}|u_s) \prod_L P(G^o_{qjkl}|G^o_{qil;}, e_{1l}, e_{2l}) + O_{qj} \sum_z c(z) \prod_L P(G^o_{qjkl}|G^o_{z;}, e_{1l}, e_{2l})}{\sum_{I_q} f_{qi} p\left(d_{qij}\middle|u_s\right) + O_{qj}}$$

(eq. 6)

where G^o_{qjkl} is the observed genotype k in trap j in stand q at locus l, G^o_{qil} is the observed genotype of adult i in stand q at locus l, z is an index for all possible multi-locus genotypes, and $c(z)$ is probability of observing genotype z given observed allele frequencies at each locus in the adult population (assuming no linkage). The summation in the first term of the numerator is over all mother trees i in the set I_q of mother trees in plot q, and the product in the first term is over all loci l in the set L of all loci. The second term in the numerator reflects the probability of obtaining the observed genotype in a seed from an out-of-plot adult. Note that trap area is not present because it cancels out. After expanding the fecundity terms, the fecundity parameter B also cancels, leaving

$$\pi_{qjk}$$

$$= \frac{\sum_{I_q} r_{qi} D_{qi}^2 p\left(d_{qij} \middle| u_s\right) \prod_L P\left(G_{qjkl}^o \middle| G_{qi;}^o, e_{1l}, e_{2l}\right) + O_{qj} \sum_z c(z) \prod_L P\left(G_{qjkl}^o \middle| G_{z;}^o, e_{1l}, e_{2l}\right)}{\sum_{I_q} r_{qi} D_{qi}^2 p\left(d_{qij} \middle| u_s\right) + O_{qj}}$$

(eq. 7)

L_{qg} , the likelihood of the observed genotypes and locations within stand q given dispersal parameter and genotyping errors, is then

$$L_q = \prod_{Jq} \prod_{Kqj} Pr\left(W_{qjk} | \pi_{qjk}\right)$$

(eq. 8)

where W_{jk} is the actual proportion of seeds with genotype k in trap j in stand q and K_{qj} is the set of all genotypes found in traps in plot q.

Putting all data together, the total likelihood of the data for plot q, L_q, the combined likelihood of observing S_{qj} seeds in the traps of which a proportion W_{qjk} of the seeds have observed genotypes $k = 1 \ldots K$ is

$$L_q = \prod_{Jq} Pr\left(S_{qj} | \theta_{qj}\right) \prod_{Kqj} Pr\left(W_{qjk} | \pi_{qjk}\right)$$

(eq. 9)

The likelihood L over all plots is the product of these likelihoods for the individual plots.

In the case of our study species, there is prior information on the dispersal parameters from Muller-Landau et al (2008). We thus run some of our analyses using these results as

priors; specifically, the prior is a lognormal distribution for u with log mean 6.3 (corresponding to u=545) and log standard deviation 0.095.

Simulations

Simulated datasets were created based on real data for adult locations, reproductive status, and size; trap locations; observed genotyping error rates among seeds and adults; observed allele frequencies among adults; and dispersal parameter u=500, a round number close to the value estimated by Muller-Landau et al (2008). The number of adult trees in the simulations, their size, locations, and reproductive status is taken to be the same as the real number of adults in the sampled plots. The adult genotypes are simulated based on the calculated allele frequencies of real data trees. Then the expected number of seeds of each tree or from off plot to arrive in each trap is calculated under the assumed dispersal parameter, and the observed number is simulated as a Poisson around the expected. Offplot seed rain is simulated in the same way that it is fit, with the assumption that seed sources are essentially smeared across the landscape, with seed production per area equal to that on the plot. Seeds coming from outside the plot are assigned random genotypes drawn based on population allele frequencies using the multinomial distribution. True seed genotypes match true adult genotypes; observed seed and adult genotypes are then simulated based on observed error rates (as calculated from PEDANT), or set equal to true genotypes for the simulation in which error rates were assumed zero. All seeds arriving in traps are assumed genotyped.

We used simulations to verify the backfitting methods, and explore power as a function of sample size (number of genotyped trees) and genotyping errors (present at observed levels, or absent). To verify the backfitting methods, we simulated large datasets with and without genotyping errors, and backfit them. To assess the effect of sample size on the parameter estimation, we simulated datasets with different fecundities, in order to produce different numbers of sampled seeds, including datasets comparable in sample size to the real data for *T. rosea*. To address the importance of correctly characterizing genotyping errors, we simulated datasets with and without genotyping errors. For all simulated scenarios, we ran five replicates.

Results

Literature review

We found 12 studies involving genotyping of maternal seed tissue, all published since 2000 (Table 1). The number of seeds genotyped per study ranged from 23 to 726. Eight of the twelve studies provided no information on genotyping error rates, and no study provided information on both allele drop and false allele rates.

Tree and seed trap data

There were a total of 70 adult trees in the 5 tower plots (Table 2), of those the majority where on Schneirla and Armour plots. Of those 70 trees, 30 trees were reproductive in one or more of the three years this study lasted, and 4 in all three. The total number seeds

caught in seed traps during that same period of time was 775, of which a large majority (627) were in Armour. Seeds were very unevenly distributed among seed traps within plots (Figure 3).

Genotyping

Multilocus genotypes were obtained for 69 adults in the tower plots area, representing 98% of the potentially reproductive parents during the time of this study. In 64 of the 69 adults, complete genotypes were obtained for all 4 loci. Observed heterozygosity among the adults ranged from 0.54 to 0.79 and the number of alleles per locus ranged from 4 to 12, with a total of 29 alleles across all four loci (Table 3).

Genotypes were obtained for 181 seeds (Table 3), although many individuals did not amplify for all four loci. We suspect that these failures to amplify are due to poor DNA quality. It is, however, impossible to establish whether that failed amplification was due to mutation at the primer annealing site (Null allele). The number of alleles per locus ranged from 4 to 12 overall. Private alleles were present in both seeds and adults. Allelic richness R_s was higher in adults than in seeds for all loci. Allele size ranges were the same for both seeds and adults in three of four loci; the exception was Tau 21, where allele 213, present only in adults, resulted in an increased size range for adults. Observed heterozygosity was lower than expected heterozygosity and in all cases this difference was larger for seeds than for adults.

Identity analysis

All complete and incomplete genotyped trees and seeds were used in the identity matches (Table 4). As different numbers of loci were genotyped on seeds, full matches were performed at 4 loci for complete genotyped seeds and trees, producing perfect matches at all 4 loci. Partially genotyped seeds were matched to complete or incomplete trees (seeds matched at 3, 2, or 1 loci). Full matches at 4 loci found an exact match for 46 seeds (25.41%) (Figure 4, Table 4). For incomplete seed genotypes, we found 17, 86 and 518 seeds-adult matches when only 3, 2, or 1 seed loci where able to be matched. In those cases, the total number of seed-to adult matches was much higher than the total number of seeds genotyped because the number of seeds that match multiple adults increases as the number of matching loci go down. When one locus mismatch was allowed per seed-adult genotype identity comparison (partial match), the total number of seed-adult matches grew from 307 to 829 (Table 4). Of course, there was also an increase in the number of seeds that matched more than one adult in the plot.

In addition to being able to identify candidate parents for initially unmatched seeds, partial matching allowed us to identify loci with potentially high error rate. In all 829 partial matches between seeds with 1, 2, or 3 loci genotyped, the allowed mismatch occurred at locus TAU17 for the majority of the seed-adult matches. However this calculation does not make any distinction between one or two alleles mismatching at a given locus, or homozygote vs. heterozygote mismatch, it simply identifies the site of potential genotyping errors and candidate loci for looking for error rates.

Error rate calculations

Allelic dropout error rates ranged from 0.4% to 7.2% for adults, and 14.4 to 32.2% for seeds, while false allele rates ranged from 0% to 0.6% for adults and 0.2 to 4.7% for seeds (Table 5). For every allele and each type of error, error rates were higher for seeds than adults. The locus with the highest allelic drop for adults was Tau 22, while for seeds it was Tau 21. The highest false allele estimate was for seeds at Tau 17, adults had false allele rates of zero with the exception of Tau 17.

Patterns of dispersal distance and direction in genetically matched seeds

The mean dispersal distance for full matches at 3 and 4 loci was 55.03 m, with a standard deviation of 28.7 (Figure 5a). The longest distance between a matched seed and adult was 219.37. There was a trend for more seeds to be dispersed to the South (Figure 5b). For matches at four loci, the mean direction was 2.87 (0.91 π) radians and the sd was 1.71 (0.544 π); for four and three loci combined the mean was 3.20 (1.02 π) and the sd was 1.81 (0.58). By convention, due North is zero, and due West is $\pi/2$, due South is π, and East is $3\pi/2$. Directional patterns were very similar when seeds traveling less than 20 m were excluded.

Inverse modeling of seed production and dispersal

Simulations

Model fitting led to good estimates of the true value for very large datasets of number of genotyped seeds (~20,000). However, both precision and accuracy declined with smaller

sample sizes: confidence intervals became very large, and they were less likely to include the true value (Figure 6a). The introduction of genotyping errors also widened confidence intervals, and further decreased the probability that the confidence intervals included the true value when sample sizes were anything but very large (Figure 6b). Overall, the model fitting showed a clear bias towards overestimating *u* (and thus seed dispersal distances) for smaller sample sizes and when there were genotyping errors.

Real data

Inverse modeling to the true seed data for *Tabebuia rosea* gave an estimated dispersal parameter (*u*) of 1143 (95% confidence interval 908 to 1373), which corresponds to a mean dispersal distance of 53.1 m (Table 7). The estimated seed dispersal kernel is shown in Figure 7. When a lognormal prior for *u* based on previous fitting to non-genetic seed trap data from Muller-Landau et al (2008) with logmean 6.3 (corresponding to *u*=544) was used, the posterior mean *u* was 619.1, corresponding to a mean dispersal distance of 36.5 m.

Discussion

Interpreting the results

Genetic data are often considered definitive for estimates of seed dispersal, yet our results show that, as for all studies, small sample sizes can compromise estimates severely. Further, our simulations suggest that most studies to date have had small sample sizes

relative to what is needed to estimate seed dispersal parameters with high precision. For example, datasets of 1000 seeds, which are larger than the largest sample size reported to date in the literature (Jones et al 2005) still have wide variation among replicate simulations (range of ~200 among just 4 simulations, or 40% of the true value). The results of our simulations also show that confidence intervals are much more dependent upon sample size than upon genotyping errors. This provides evidence that it is best to (appropriately) include all genetic data, not only what is typically considered as good data (i.e. full genotypes), because even incomplete or error-prone genetic data convey useful information.

In light of those simulation results, it is clear that there are simply not sufficient genetic data to provide good information on seed dispersal distances for our case study species. This conclusion is supported by the backfits to the real data. The backfits that use informative priors for *u* result in posterior estimates that are very similar to the priors, indicating the limited power of the new genetic data collected in the current study. This suggests that the prior, based on inverse modeling with a larger, but entirely non-genetic, seed trap dataset provided much more information than the genetic data analyzed here.

Tabebuia rosea disperses its seeds at the end of the dry season and beginning of the wet season. In the dry season, strong winds (long term mean of 3.9 km/h) blow predominantly from the North to the South. In the wet season, winds are much weaker, and have no particular directionality. We expect that the longer dispersal events for *T.*

rosea will thus be largely to the South. Despite our limited sample size, our study suggests that *T. rosea* dispersal is indeed directional and from North to South, as the mean dispersal direction is almost exactly due south.

Genotyping error rates and the uses of imperfect genetic data

The error rates in our adult tree population are within the range of reported errors in similar studies. For example, in a study of gene dispersal in *Quercus* using established seedlings, Moran and Clark (2012) report error rates that range from 0.02 to 0.08 for allelic dropout and from 0.08 to 0.18 for mistyping (false allele). In another study, Zang et al (2006) report an error rate of 0.014 for allele dropout and 0.019 for false allele, using adult trees for clone identification.

The literature contains relatively little information on genotyping error rates for seed maternal tissue (Table 1). In a study of seed dispersal in the genus *Quercus* using seed pericarps, Grivet et al (2005) report that 35% of their replicated genotypes were affected by allelic dropout; they do not report false alleles. Cremer (2012) concluded that there was no evidence of allelic dropouts in seed exocarps of *Abies alba* based on the finding that the inbreeding coefficient *F* in the seed population was no greater than that of the adult population. They reached this conclusion after excluding samples with non repeatable unusual peaks, and after discarding markers that failed amplification on 10% or more of their seed samples. In their maternity analysis of *Parkia* seed exocarps, Heymann et al (2012) allowed for one mismatch to account for allelic dropout, but they

did not actually quantify it. There are no other published studies that quantify allelic dropout in seed maternal tissue and none at all that quantify false alleles. Our study shows that even when false allele rates are virtually zero for tree genotypes, seed tissues can have relatively substantial false allele rates.

Most studies to date using maternal seed tissue rely on a combination of duplicate genotyping and the discarding of unreliable genotypes. But even imperfect genetic data is informative and it doesn't just have to be discarded and left unpublished. An approach such as the one used in this paper to calculate error rates, either directly or using programs such as PEDANT, allows for inclusion of data that would otherwise be discarded. Incomplete genotypes still provide information relative to no genetic information at all. In comparison with traditional inverse modeling without genetic data, even incomplete genetic data could help identify the source of seeds in the case of overlapping seed shadows and potential immigrant seeds due to long distance dispersal. However, inclusion of such data without proper accounting of error rates can lead to overestimation of immigration rates and seed dispersal distance estimates.

Though some genetic data are more useful than none, the value of genetic data of any kind may have been overstated by previous papers. The results of our simulations suggest that the cost-effectiveness of genetic approaches to seed dispersal should be carefully evaluated before planning a sampling design. Our simulation results show that even without genotyping errors, large sample sizes in genotyped seeds may be needed in

order to obtain precise estimates of seed dispersal parameters. Genetic analyses remain relatively labor-intensive and costly compared with other types of data collection and analyses. In some cases, the gain of genetic information relative to the cost and effort involved will not justify the use of molecular methods in seed dispersal studies. If resources only enable genotyping of small numbers of seeds, more may be gained from other approaches such as inverse modeling from seed trap data alone. Previous seed dispersal studies using direct assignment through maternally derived tissues had sample sizes ranging from 23 to 726 genotyped seeds (Table 6). We found that simulations with sample sizes in this range result in dispersal estimates with wide confidence intervals, even in the absence of genotyping errors. The details of our results are specific to our case study landscape and species; that is, similar sample sizes will provide more or less information depending on the arrangement of trees and traps, and the dispersal distances of seeds. Nonetheless, our results suggest, at the very least, that sample sizes for genetic studies need to be considered more carefully. Simulation-based power analysis, such as those we present here, should be employed in advance as part of the design of any seed dispersal study involving genetic analyses. Too often, sample sizes in previous studies have been based simply on the resources available and the number of seeds that can be analyzed with those resources.

Conclusions, implications and future directions

The data set used to estimate the dispersal parameters for *T. rosea* in this study (181 genotyped seeds) proved to be insufficient for the model to provide usefully precise

estimates of seed dispersal parameters. Better estimates require larger datasets. According to the results of our simulations study, even datasets as large as 2000 genotyped seeds may not be sufficient in the presence of realistic levels of genotyping errors for molecular markers with the resolution of those used here. Further, the results of this study show that error rates should be calculated and incomplete genetic data incorporated in analyses in order to increase power and obtain better parameter estimates.

Seed maternal tissue is difficult to work with and high rates of error can be expected from the generally low quality DNA obtained. The dearth of published error rates specifically for such tissues is likely to mislead inexperienced investigators concerning the challenges they will encounter if they embark on such analyses. As our study shows, even when false allele rates are negligible for adult genotypes, they can be important in the case of maternally derived seed tissues, and allelic dropout rates can be much higher for seed than adult tissue (Table 6). Genetic analyses matching maternal seed tissue to adults have been heralded for over a decade as a superb new tool that will revolutionize our understanding of seed dispersal (Godoy and Jordano 2001, Grivet et al 2005, Pairon et al 2006, Garcia et al 2007, Hanson et al 2007, Terakawa et al 2009, Karubian et al 2010, Hansenkamp et al 2011, Cremer et al 2012, Heymann et al 2012, Smouse et al 2012). The fact that less than fifteen studies have thus far reported results of such analyses is, we suspect, due not to a lack of more attempts to use this much-hyped technique, but rather to many unpublished failures related to high genotyping error rates.

We recommend that simulation models can routinely be used for power analyses to calculate necessary sample size in order to achieve enough resolution to estimate dispersal parameters to the desired precision. When only small sample sizes are possible, alternative methods such as inverse modeling from seed traps alone should be considered. Studies should be designed to accurately quantify error rates for adults and separately for seeds/seedlings using subsets of data, to avoid costly duplicated genotyping of all samples. The effort and resources that are saved by not duplicating all genotypes can in turn be used to genotype more seeds to increase the sample size.

Table 1. Literature review of studies using maternally derived tissues for direct assignment of seeds to source trees using microsatellites, number of genotyped seeds, number of genotyped loci and error measures and rates, nr indicates not reported

Article	Sp or genus	N seeds	N loci	Error measure	Error rate
Godoy & Jordano 2001	*Prunus mahaleb*	95	9	nr	nr
Grivet et al 2005	*Quercus lobata*	215	3	allele drop	35%
Jones et al 2005	*Jacaranda copaia*	726	4	nr	nr
Pairon et al 2006	*Prunus serotica*	188	6	nr	nr
Garcia et al 2007	*Prunus mahaleb*	549	10	nr	nr
Hanson et al 2007	*Dipterix panamensis*	23	4	dominant error rate	6.10%
Terakawa et al 2009	*Myrica rubra*	387	9	nr	nr
Karubian et al 2010	*Oenocarpus batahua*	313	7	average per locus error rate	7.4%
Hansenkamp et al 2011	*Fagus sylvatica*	99	6	nr	nr
Cremer et al 2012	*Abies alba*	661	3	nr	nr
Heymann et al 2012	*Parkia*	133	7	nr	nr
Smouse et al 2012	*Quercus lobata*	169	6	null allele error	2%

Table 2. Tree and seed sample size information by plot.

	Trees					Seeds	
Plot	Number of trees	Number of reproductive trees 2005	Number of reproductive trees 2006	Number of reproductive trees 2007	Number of genotyped trees	Number of seeds captured	Number of genotyped seeds
Armour	23	7	9	6	23	627	136
Schneirla	38	3	15	2	37	118	27
Pearson	7	2	2	2	7	19	15
Drayton	1	0	1	0	0	11	2
Zetek	1	0	1	0	0	0	1
Total	70	11	26	11	67	775	181

Table 3. Genetic diversity statistics for each locus, for adult trees and seeds: Size is the size range in bp; *N* is the number of individuals; *A* is the number of alleles; A_r is allelic richness; *P* is the number of private alleles; H_o is the observed heterozygosity; H_e is the expected heterozygosity; F_{is} is Wright's inbreeding coefficient; *I* is the probability of identity.

Locus	Size	*A*	A_r	*N*	*P*	H_o	H_e	F_{is}	*I*
					Trees				
Tau17	154-189	12	11.969	67	1	0.761	0.861	0.117	0.037
Tau21	213-232	5	5	65	1	0.569	0.625	0.09	0.137
Tau22	152-178	4	4	68	0	0.544	0.586	0.071	0.262
Tau31	259-275	8	7.956	68	0	0.794	0.793	-0.002	0.077
					Seeds				
Tau17	154-189	11	9.488	238	0	0.752	0.808	0.069	0.062
Tau21	216-232	5	4.715	184	1	0.196	0.515	0.62	0.285
Tau22	152-178	6	4.922	263	2	0.319	0.52	0.386	0.227
Tau31	259-275	8	7.159	186	0	0.594	0.678	0.12	0.163

Table 4. Number of seed to adult full and partial matches (one locus mismatch) per number of seed loci genotyped.

Number of seed loci genotyped	Number of seeds genotyped	Number of exact matches to trees	Number of partial matches to trees	Number of seeds that match more than one tree for full matches	Number of seeds that match more than one tree for partial matches
4	103	46	132	0	34
3	38	17	141	16	25
2	34	86	556	21	33
1	6	158	0	3	0
all	181	307	829	40	92

Table 5. Maximum Likelihood estimates of error rate probabilities (and 95% CI), as calculated using PEDANT.

Allele	Allele Dropouts (lower CI, upper CI)		False Alleles (lower CI, upper CI)	
	Adults	Seeds	Adults	Seeds
Tau 17	0.037 (0.023, 0.052)	0.194 (0.174, 0.214)	0.006 (0.002, 0.013)	0.047 (0.038, 0.058)
Tau 21	0.016 (0.005, 0.037)	0.322 (0.247, 0.397)	0.000 (0.000, 0.004)	0.002 (0.0004, 0.005)
Tau 22	0.072 (0.005, 0.037)	0.144 (0.119, 0.173)	0.000 (0.000, 0.005)	0.010 (0.006, 0.015)
Tau 31	0.004 (0.0008, 0.010)	0.207 (0.177, 0.239)	0.000 (0.000, 0.002)	0.006 (0.002, 0.012)

Table 6. Dispersal and fecundity parameters estimated from the field data for *Tabebuia rosea,* with and without prior information based on Muller-Landau et al 2008. Both fits started with initial values of *u*=150 and F=0.4.

	u	F
Prior distribution for *u*	Mean (SE)	Mean (SE)
None	1143 (135)	0.064 (0.0026)
Lognormal with logmean X and logsd Y	619 (26)	0.056 (0.0021)

Figure 1. Study area map, showing the Island of Barro Colorado with the positions of the five telemetry towers around which the plots are located.

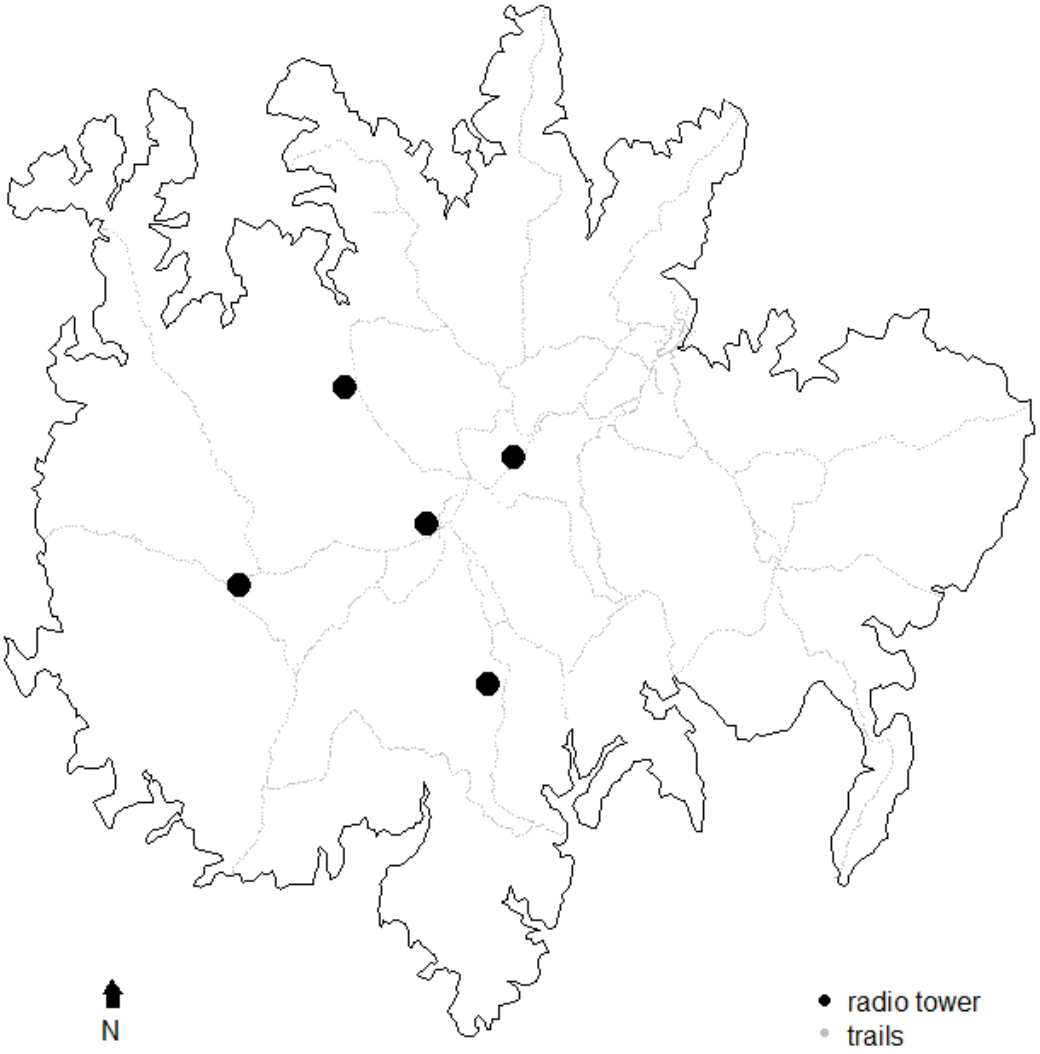


Figure 2. Diagram of the field sampling design for each plot, showing the approximate locations of the 49 seed traps relative to the central tower and the plot border. Trees were censused in the entire area of each plot. Plots extend farther to the north than to other directions because prevailing winds during the seed dispersal season are from the north.

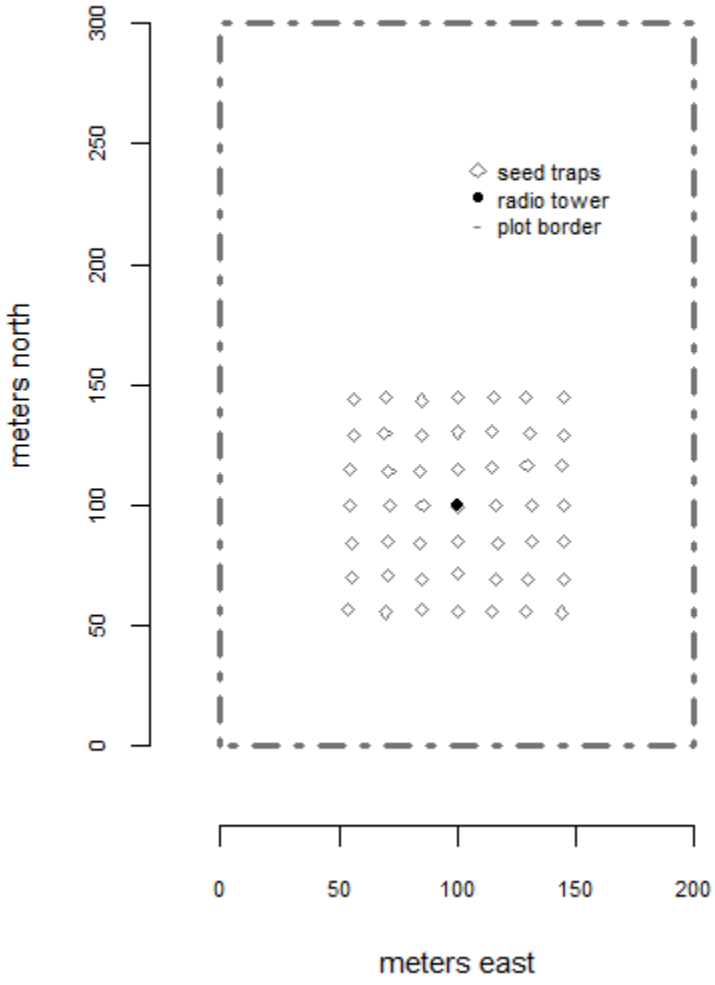
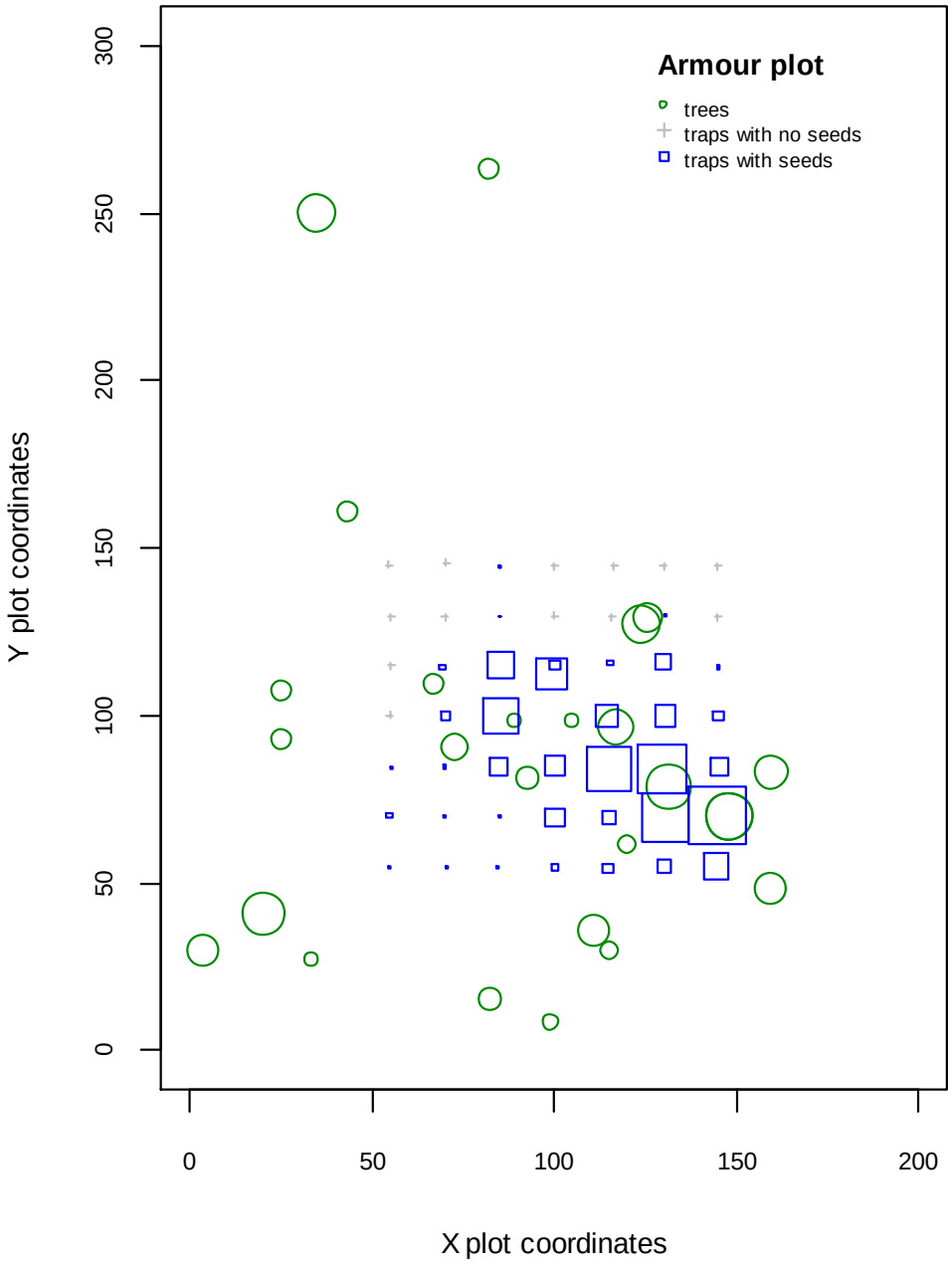
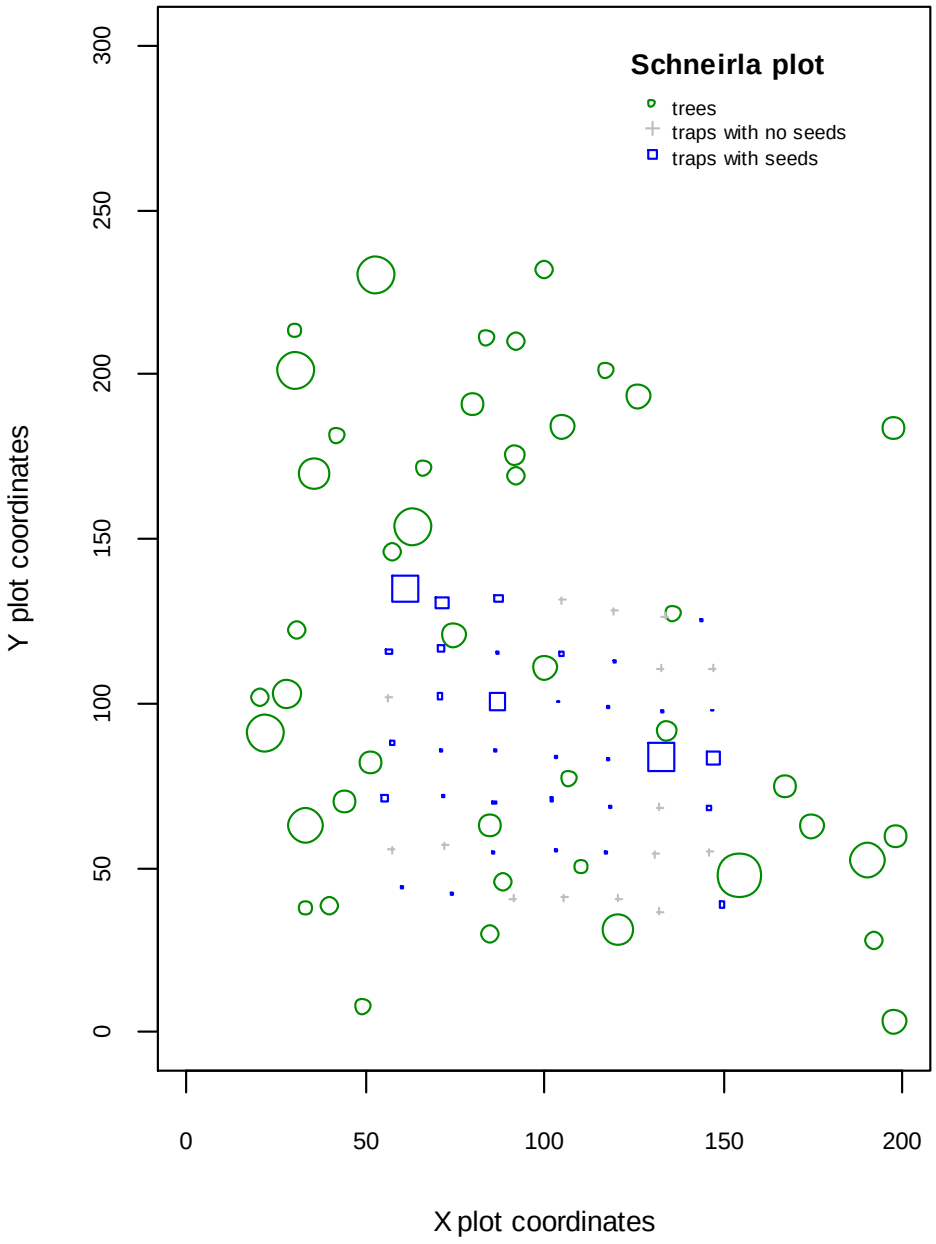


Figure 3. Maps of the adult trees of *Tabebuia rosea*, and the seed traps in each plot. The sizes of the symbols for the trees are scaled to trunk diameter (dbh), and the sizes of the symbols for traps with seeds are scaled to the total number of seeds captured in them.

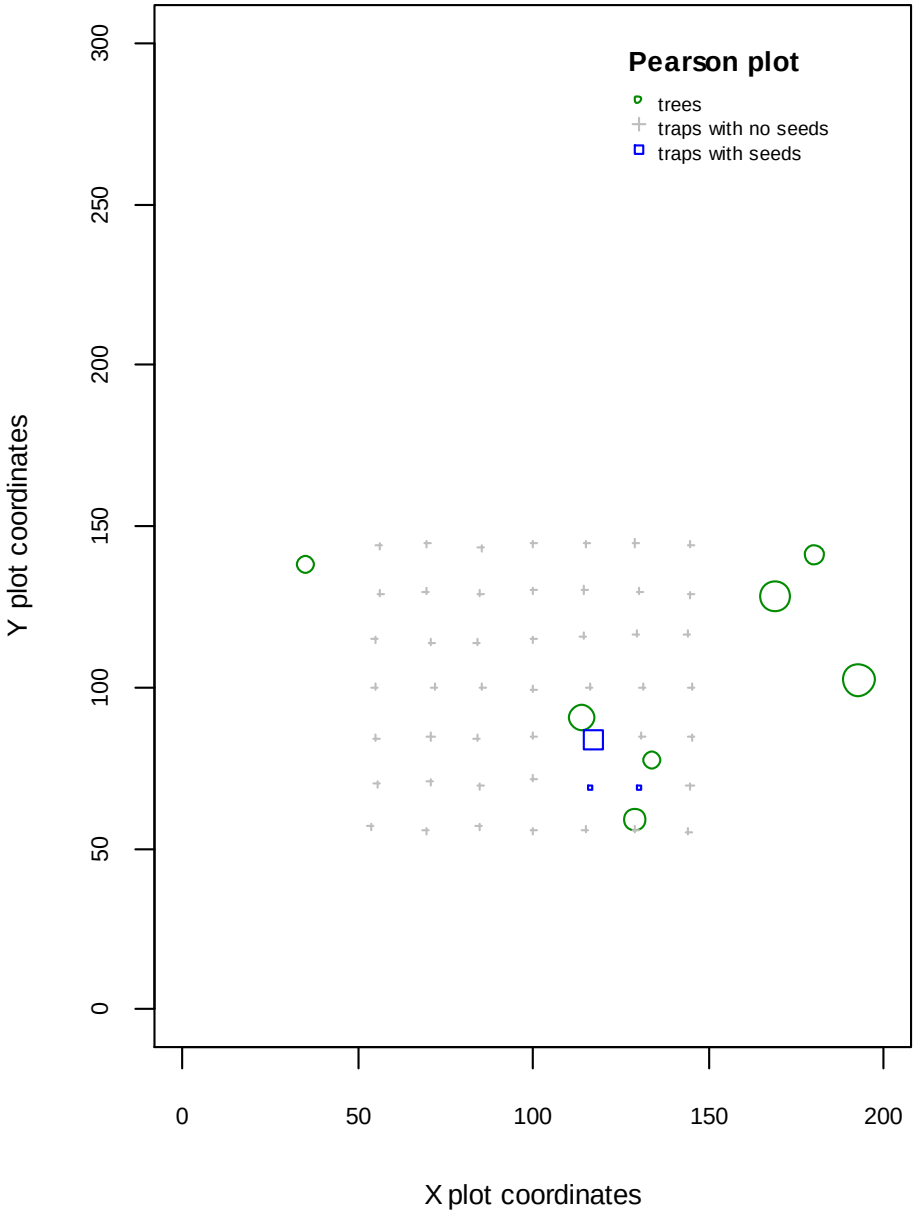
a) Armour plot



b) Schneirla plot



c) Pearson plot



d) Drayton plot

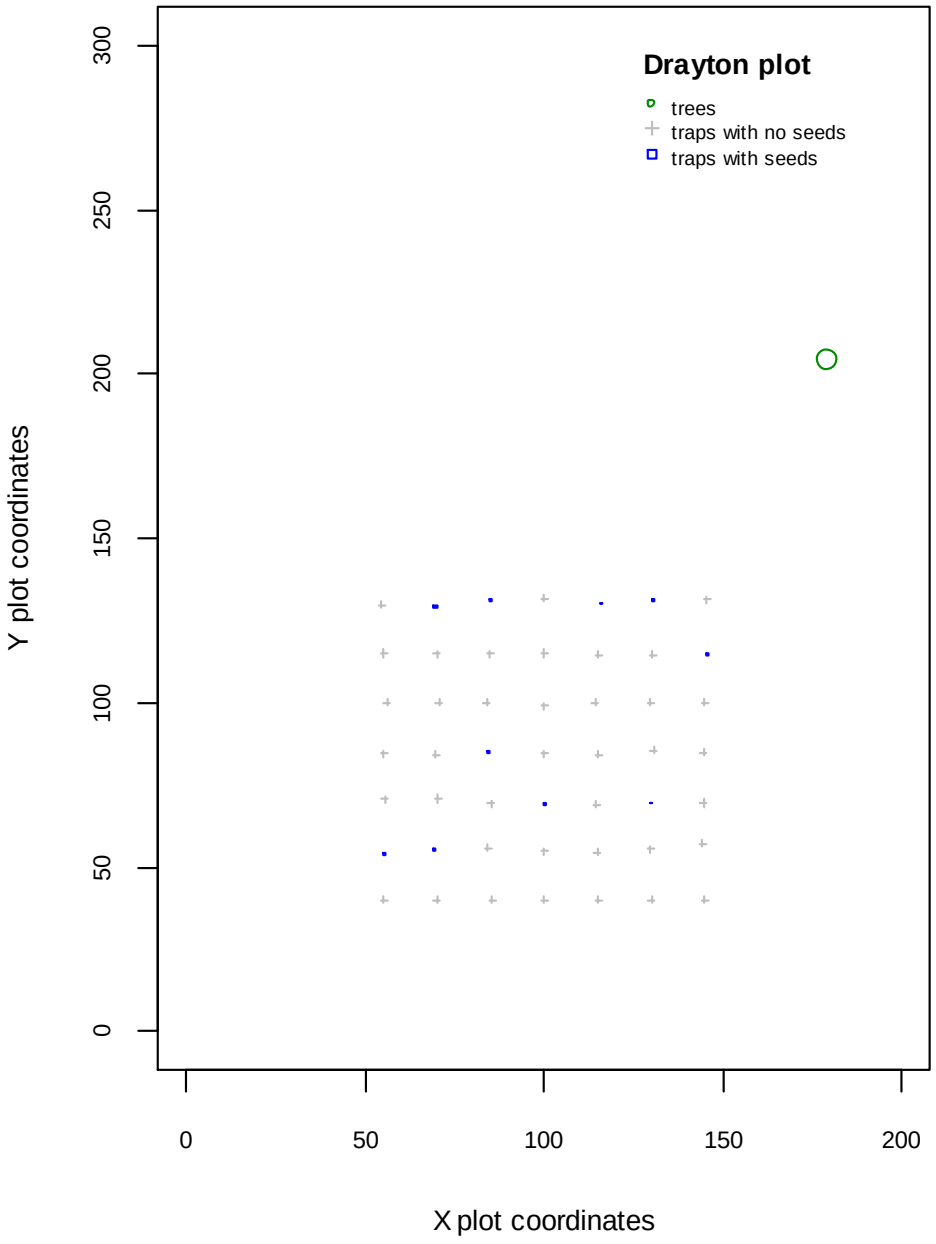
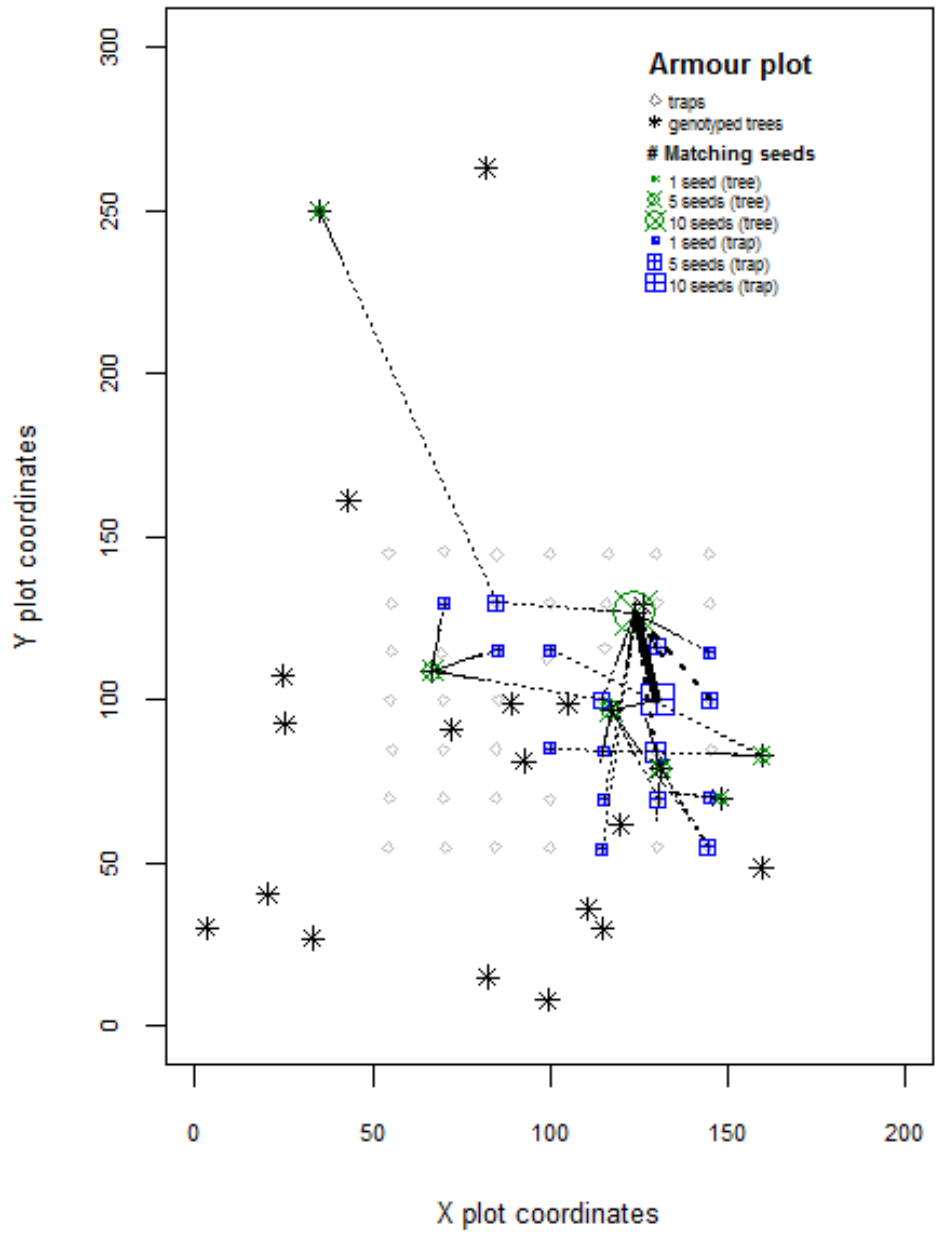
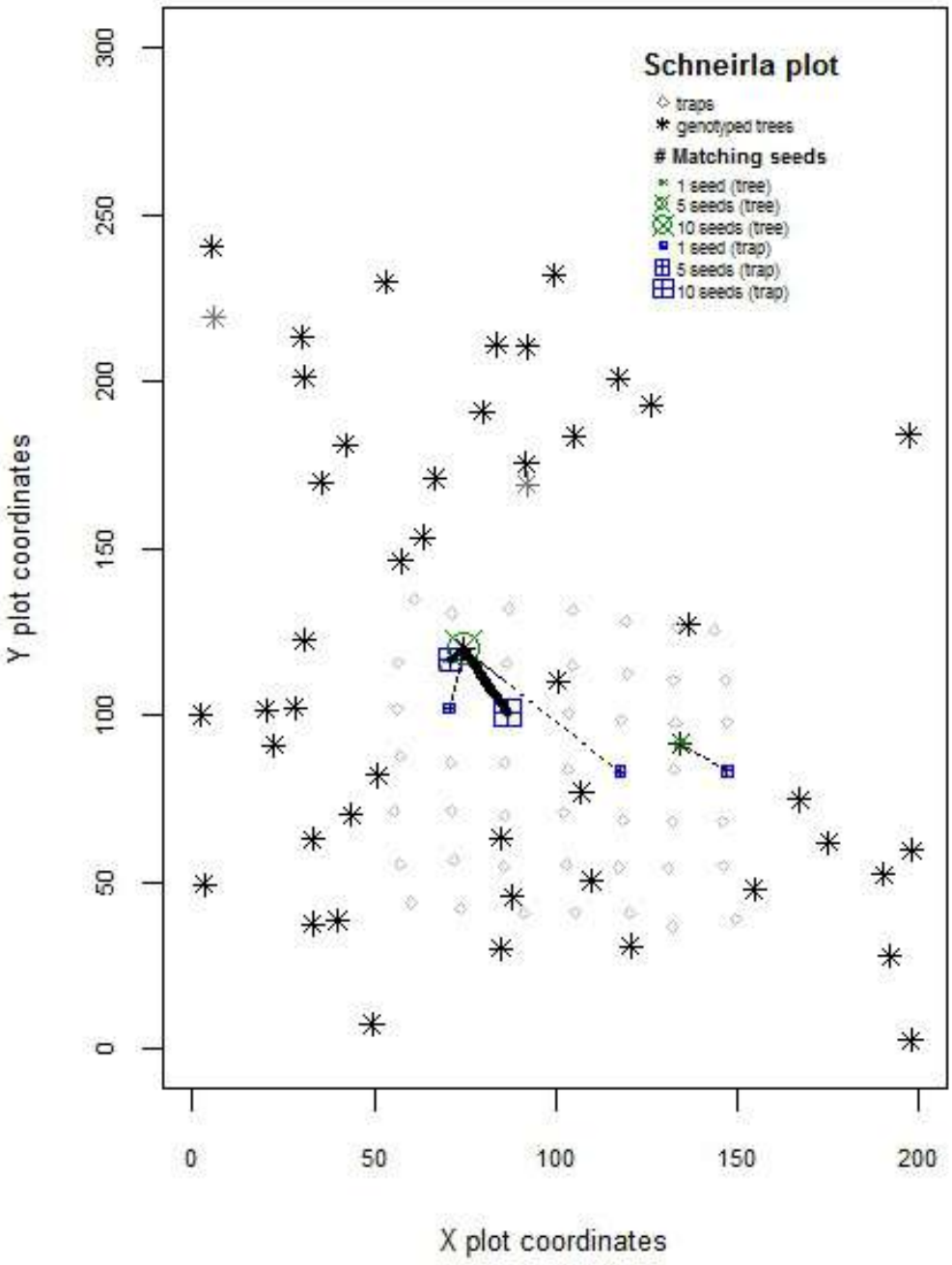


Figure 4. Maps showing exact seed to tree matches (lines), as well as other trees and traps. Tree symbol size, trap symbol size, and line thickness are all scaled to the number of matching seeds. Only matches of seeds and trees with complete genotypes and no mismatches are shown. Note that there were no cross-plot matches.

a) Armour plot



b) Schneirla plot



c) Pearson plot

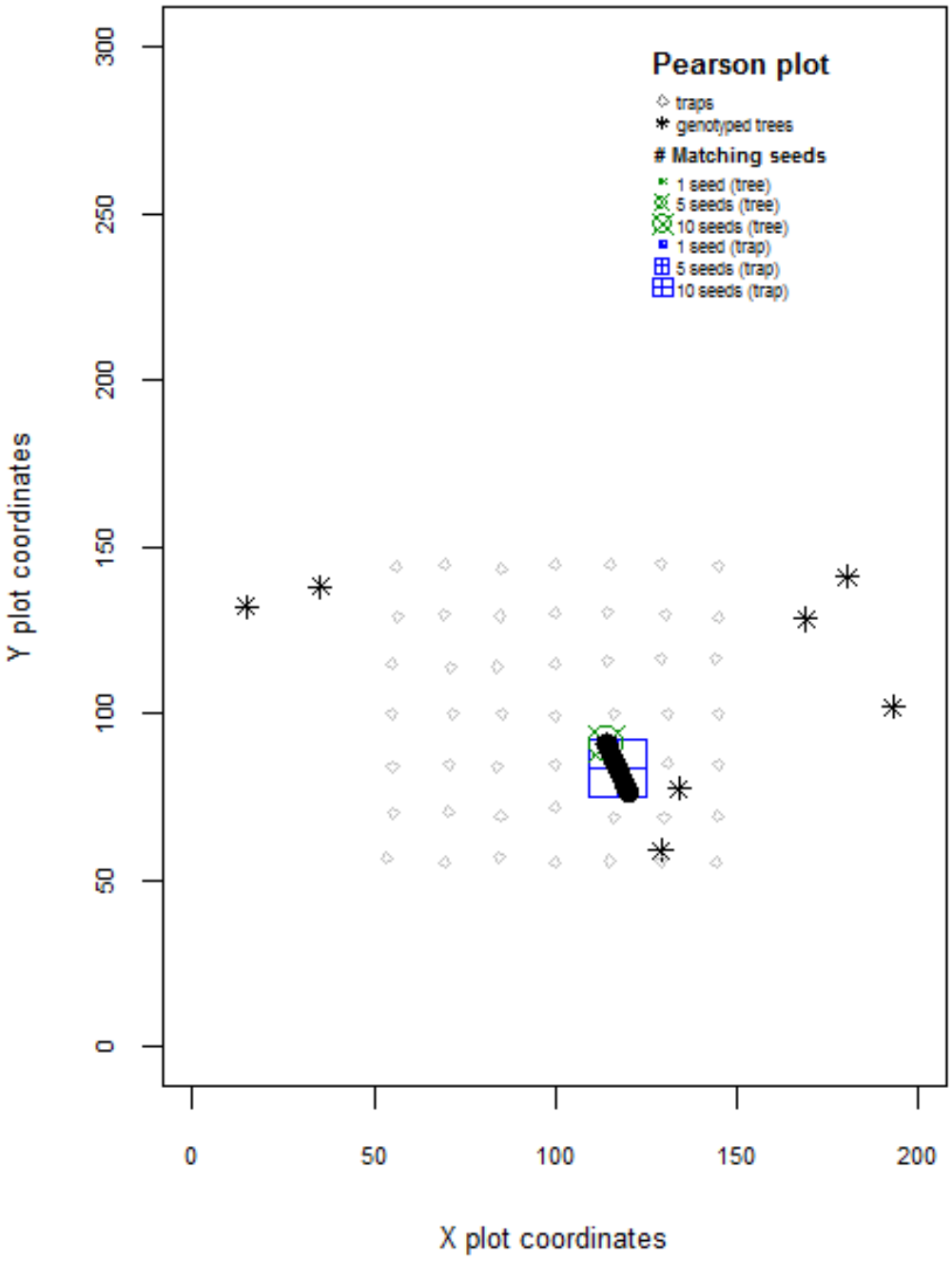
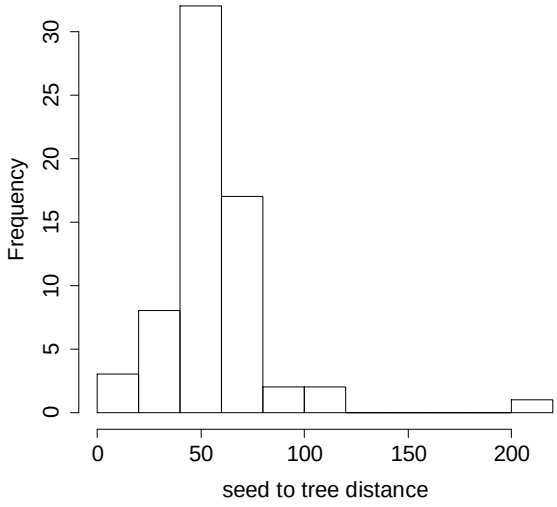
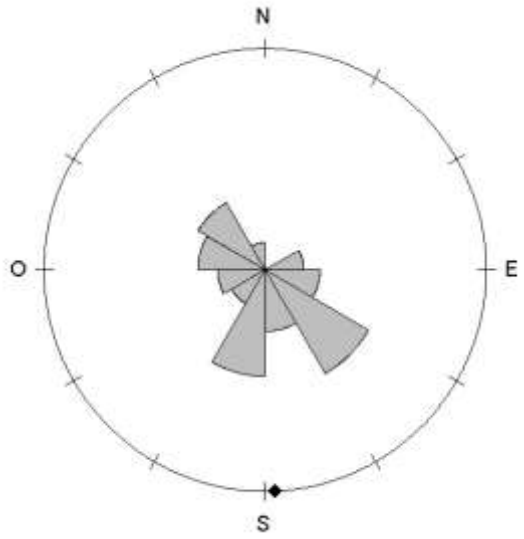


Figure 5. Patterns of seed dispersal distance and directionality based on full matches at 3 and 4 loci. a) Distribution of dispersal: Mean dispersal distance is 55.03 m, with sd 28.7. b) Directional patterns: Mean direction (indicated by the back diamond) is -1.53, or - 0.49 π, which translates as almost due south, with sd direction of 1.24 or 0.39 π (the convention is for 0 to represent East,π/2 to represent North, and so forth).



a) Distribution of dispersal distances



b) Directional patterns of full matches

Figure 6. Posterior mean values and 95% CI of the dispersal parameter (*u*) as estimated by the model from simulated datasets with or without genotyping errors (at observed rates). The confidence intervals were calculated from the estimates of one simulation with 5000 steps. a) simulation without genotyping error, b) simulation with genotyping error.

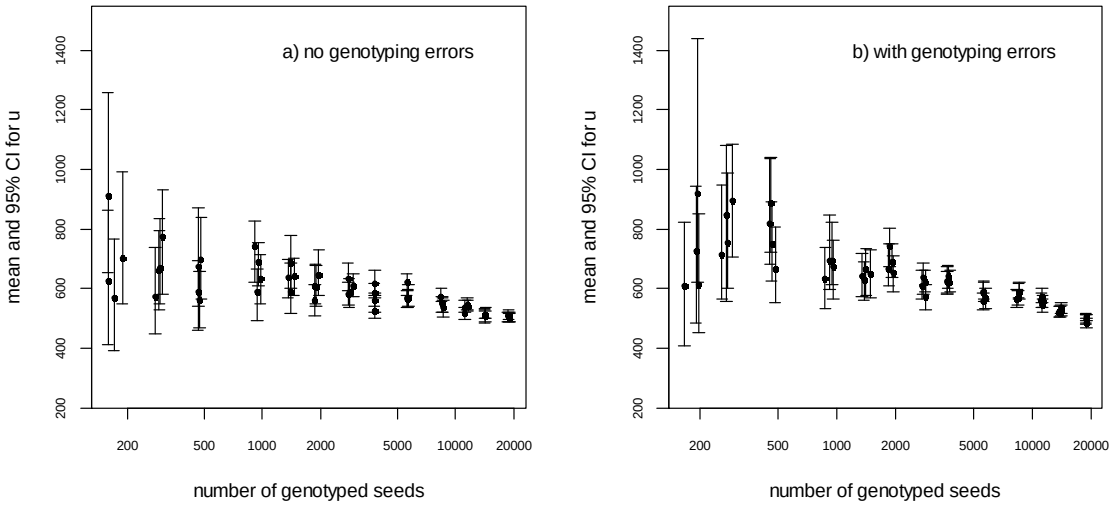
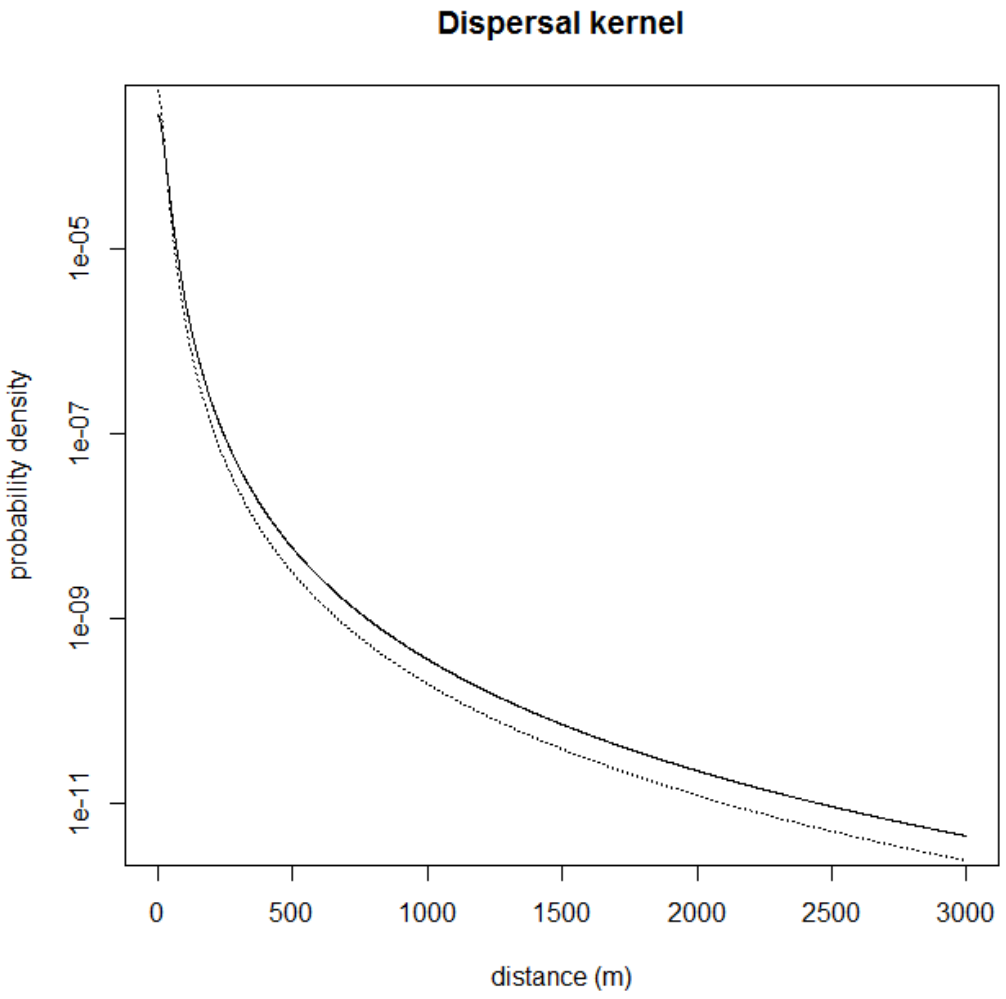


Figure 7. Estimated dispersal kernels for *T. rosea*. The solid line depicts the dispersal kernel from the data when no prior distribution was used (u=1143); the dotted line depicts the dispersal kernel when using a prior based on the results in Muller-Landau et al 2008 (u= 619).



References

Ashley, M. 2010.Plant parentage, pollination, and dispersal: how DNA microsatellites have altered the landscape. Critical Reviews in Plant Sciences **29**: 148–161.

Bonin, A., E. Bellemain, P. Bronken Eidesen, F. Pompanon, .C Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. Molecular Ecology **13**: 3261–3273.

Braga, A. C., A. M. Reis, L. T. Leoi, R. W. Pereira, and R. G. Collevatti. 2007. Development and characterization of microsatellite markers for the tropical tree species *Tabebuia aurea* (Bignoniaceae). Molecular Ecology Notes **7**: 53–56.

Broquet, T., and E. Petit. 2004. Quantifying genotyping errors in noninvasive population genetics. Molecular Ecology **13**: 3601–3608.

Bullock, J.M., K. Shea, and O. Skarpaas. 2006. Measuring plant dispersal: an introduction to field methods and experimental design. Plant Ecology **186**: 217–234.

Choo, J., T. E. Juenger, and B. B. Simpson. 2012. Consequences of frugivore-mediated seed dispersal for the spatial and genetic structures of a neotropical palm. Molecular Ecology **21**: 1019–1031.

Clark, C. J., J. R. Poulsen, B. M. Bolker, E. F. Connor, and V. T. Parker. 2005. Comparative seed shadows of bird-, monkey-, and wind-dispersed trees. Ecology **86**: 2684-2694.

Clark, J. S., M. Silman, R. Kern, E. Macklin, and J. Hiller Islambers. 1999. Seed dispersal near and far: patterns across temperate and tropical forests. America **80**: 1475–1494.

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 Populations **29**: 312.

Cremer, E., B. Ziegenhagen, K. Schulerowitz, C. Mengel, K. Donges, R. Bialozyt, E. Hussendörfer, and S. Liepelt. 2012. Local seed dispersal in european silver fir (*Abies alba* Mill.): lessons learned from a seed trap experiment. Trees **26**: 987–996.

Croat, T.1978 Flora of Barro Colorado Island. Stanford University Press.

Dalling, J. W.; S. Hubbell, and K. Silvera. 2001. Seed dispersal, seedling establishment and gap partitioning among tropical pioneer trees. Journal of Ecology **86** : 674-689.

Dewoody, J., J. D. Nason, and V. D. Hipkins. 2006. Mitigating scoring errors in microsatellite data from wild populations. Molecular Ecology Notes **6**: 951–957.

Dow, B. D., and M. V. Ashley. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Molecular Ecology **5**: 615–627.

Dow, B. D., and M. V. Ashley. 2008. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, Quercus macrocarpa. Molecular Ecology **5**: 615-627.

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

García, C, P. Jordano, and J. A. Godoy. 2007. Contemporary pollen and seed dispersal in a *Prunus mahaleb* population: patterns in distance and direction. Molecular Ecology **16**: 1947–1955.

Godoy, J. A., and P. Jordano. 2001. Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. Molecular Ecology **10**: 2275–2283.

Grivet, D., P. E. Smouse, and V. L. Sork. 2005. A novel approach to an old problem: tracking dispersed seeds. Molecular Ecology **14**: 3585–3595.

Hanson, T., S. Brunsfeld, B. Finegan, and L. Waits. 2007. Conventional and genetic measures of seed dispersal for *Dipteryx panamensis* (Fabaceae) in continuous and fragmented costarican rain forest. Journal of Tropical Ecology **23**: 635–642.

Hasenkamp, N., B. Ziegenhagen, C. Mengel, L. Schulze, H-P. Schmitt, and S. Liepelt. 2010. Towards a DNA marker assisted seed source identification: a pilot study in european beech (*Fagus sylvatica* L.). European Journal of Forest Research **130**: 513–519.

Hauser, L., M. Baird, R. A. Y. Hilborn, L. W. Seeb, and J. E. Seeb. 2011.An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. Molecular Ecology resources **11**: 150-161.

Hess, M. A., J. G. Rhydderch, L. L. LeClair, R. M. Buckley, M. Kawase, and L. Hauser. 2012. Estimation of genotyping error rate from repeat genotyping, unintentional recaptures and known parent–offspring comparisons in 16 microsatellite loci for brown rockfish (Sebastes auriculatus). Molecular Ecology Resources**12**: 1114–1123.

Heymann, E. W., K. Lüttmann, I. M. Michalczyk, P. P. Pinedo Saboya, B. Ziegenhagen, and R. Bialozyt. 2012. DNA fingerprinting validates seed dispersal curves from observational studies in the neotropical legume Parkia. PloS One **7**: e35480.

Hoffman, J. I., and W. Amos. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology **14**: 599–612.

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

Johnson, P. C.D., and D. T. Haydon. 2007. Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. Genetics **175**: 827–842.

Jones, F. A., and H. C. Muller-Landau. 2008. measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods. Journal of Ecology **96**: 642–652.

Jones, F. A., and S. P. Hubbell. 2006. Demographic spatial genetic structure of the neotropical tree, *Jacaranda copaia*. Molecular Ecology **15**: 3205–3217.

Jones, F. A., J. Chen, G-J. Weng, and S. P. Hubbell. 2005. A genetic evaluation of seed dispersal in the neotropical tree *Jacaranda copaia* (Bignoniaceae). The American Naturalist **166**: 543–555.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Molecular Ecology **16:** 1099–1106.

Karubian, J., V. L. Sork, T. Roorda, R. Durães, and T. B. Smith. 2010. Destination-based seed dispersal homogenizes genetic structure of a tropical palm. Molecular Ecology **19**: 1745–1753.

Kolomyjec, S. H., J. Y.T. Chong, D. Blair, J. Gongora, T. R. Grant, C. N. Johnson, and C. Moran. 2009. Population genetics of the platypus (Ornithorhynchus anatinus): a fine-scale look at adjacent river systems. Australian Journal of Zoology **57**: 225-234.

Llorens, T. M., M. Byrne, C. J. Yates, H. M. Nistelberger, and D. J. Coates. 2012. Evaluating the influence of different aspects of habitat fragmentation on mating patterns and pollen dispersal in the bird pollinated Banksia sphaerocarpa var. caesia. Molecular Ecology 21: 314–328.

Moran, E. V., and J. S. Clark. 2012. Between-site differences in the scale of dispersal and gene flow in red oak. PloS One **7**: e36492.

———.2011. Estimating seed and pollen movement in a monoecious plant: a hierarchical bayesian approach integrating genetic and ecological data. Molecular Ecology **20**:1248–1262.

Morin, P. A., R. G. Leduc, F. I. Archer, K. K. Martien, R. Huebinger, J. W. Bickham, and B. L. Taytor. 2009. Significant Deviations from Hardy–Weinberg equilibrium caused by low levels of microsatellite genotyping errors. Molecular Ecology Resources **9**: 498–504.

Muller-Landau, H. C., S. J. Wright, O. Calderón, R. Condit, and S. P. Hubbell. 2008. Interspecific variation in primary seed dispersal in a tropical Forest. Journal of Ecology **96:** 653–667.

Muller-Landau, H. C., S. J. Wright, O. Calderón, S. P. Hubbell, and R. B. Foster. 2002. Recruitment limitation : concepts , methods and case-studies from a tropical Forest In: Levey D.J., W. R. Silva, and M. Galetti (eds). Seed dispersal and frugivory: ecology, evolution and conservation. CAB International, Oxford: 35-53

Pairon, M., J. Mathieu, and A-L. Jacquemart. 2006. Modeling seed dispersal of black cherry, an invasive forest tree: how microsatellites may help? Canadian Journal of Forest Research **36**: 1385–1394.

Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: causes, consequences and solutions. Nature Reviews Genetics **6**: 847–59.

Pons, J., and J. G. Pausas. 2007.Acorn dispersal estimated by radio-tracking. Oecologia **153**: 903-911.

Ribbens, E., J. A. Silander, and S. W. Pacala. 1994. Seedling recruitment in forests: calibrating models to predict patterns of tree seedling dispersion. Ecology **75**:1794–1806.

Robledo-Arnuncio, J. J., and C. García. 2007. Estimation of the seed dispersal kernel from exact identification of source plants. Molecular Ecology **16**: 5098–5109.

Sanchez, J. M. C., D. F. Greene, and M. Quesada. 2011. a field test of inverse modeling of seed dispersal. American Journal of Botany **98:** 698–703.

Sebbenn, A. M., A. C. M. Carvalho, M. L. M. Freitas, S. M. B. Moraes, A. P. S. C. Gaino, J. M. da Silva, C. Jolivet, and M. L. T. Moraes. 2011. Low levels of realized

seed and pollen gene flow and strong spatial genetic structure in a small, isolated and fragmented population of the tropical tree *Copaifera langsdorffii* Desf. Heredity **106:** 134–145.

Skarpaas, O., and K. Shea. 2007. Dispersal patterns, dispersal mechanisms, and invasion wave speeds for invasive thistles. The American Naturalist **170**: 421-430

Smouse, P. E., V. L. Sork, D. G. Scofield, and D. Grivet. 2012 Using seedling and pericarp tissues to determine maternal parentage of dispersed valley oak recruits. Journal of Heredity **103**: 250–9.

Terakawa, M., Y. Isagi, K. Matsui, and T. Yumoto. 2008. Microsatellite analysis of the maternal origin of myrica rubra seeds in the feces of japanese macaques. Ecological Research **24**: 663–670.

Tosi, J. A., and FAO, Roma (Italia). 1971. Inventariación y demostraciones forestales Panamá-zonas de vida-una base ecológica para investigaciones silvícolas e inventariación forestal en la República de Panamá

Turchin, P. 1998. Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants. Sinauer, Sunderland, MA.

van Putten, B., M. D. Visser, H. C. Muller-Landau, and P. A. Jansen. 2012. Distorted-distance models for directional dispersal: a general framework with application to a wind-dispersed tree. Methods in Ecology and Evolution **3**: 642–652.

Vander Wall, S. B. 1994. Removal of wind-dispersed pine seeds by ground-foraging vertebrates. Oikos **69** : 125-132.

Warude, D., P. Chavan, K. Joshi, and B. Patwardhan. 2003. DNA isolation from fresh, dry plant samples with highly acidic tissue extracts. Plant Molecular Biology Reporter **21**: 467-467.

Xiao, Z., P. A. Jansen, and Z. Zhang. 2006. Using seed-tagging methods for assessing post-dispersal seed fate in rodent-dispersed trees. Forest Ecology and Management **223**: 18–23.

Zhang, D., S. Mischke, R. Goenaga, A. A. Hemeida, and J. A. Saunders. 2006. Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. Crop Science **46**: 2084-2092.