Investigating Drivers of Genetic Structure in Plants: Global, Regional and Local Scales

Diana Gamba-Moreno
University of Missouri-St. Louis, dlgtk5@mail.umsl.edu

Follow this and additional works at: https://irl.umsl.edu/dissertation
Part of the Biodiversity Commons, Evolution Commons, and the Population Biology Commons

Recommended Citation
Gamba-Moreno, Diana, "Investigating Drivers of Genetic Structure in Plants: Global, Regional and Local Scales" (2020). Dissertations. 919.
https://irl.umsl.edu/dissertation/919
Investigating Drivers of Genetic Structure in Plants: Global, Regional and Local Scales

Diana L. Gamba-Moreno

M.S. Biology: Ecology and Systematics, San Francisco State University, 2013
B.S. Biology (emphasis in Botany), Universidad del Valle, Cali, Colombia, 2010

A Dissertation Submitted to The Graduate School at the University of Missouri-St. Louis
in partial fulfillment of the requirements for the degree
Doctor of Philosophy in Biology with an emphasis in Ecology, Evolution, and Systematics

May
2020

Advisory Committee

Nathan Muchhala, Ph.D.
Chairperson

Robert Ricklefs, Ph.D.

Christine Edwards, Ph.D.

María del Carmen Ulloa, Ph.D.

Copyright, Diana L. Gamba-Moreno, 2020
Abstract

Genetic structure within and among plant populations is a critical component of plant biodiversity, informing local adaptation, conservation, and incipient speciation. However, its drivers remain poorly understood, especially across different spatial scales. In my dissertation I examined factors that affect plant population genetic structure at global, regional, and local scales. At the global scale, I performed a literature review of population genetic differentiation ($F_{ST}$) in seed plants based on a 337-species dataset with data on $F_{ST}$ and species traits. Using phylogenetic multiple regressions, I found that $F_{ST}$ is higher for tropical, mixed-mating, non-woody species pollinated by small insects, and lower for temperate, outcrossing trees pollinated by wind. At the regional scale, I tested the effect of flowering asynchrony on genetic divergence between conspecific subpopulations of understory flowering plants in the Andean biodiversity hotspot. I documented flowering phenology for nine species at two sites over one year and inferred population genetic parameters with a genome-wide genotyping approach termed 2b-RAD sequencing. I found that species with higher flowering asynchrony between their subpopulations also show greater genetic divergence. At the local scale, I examined the effect of insect vs. hummingbird pollination modes on the fine-scale spatial genetic structure (SGS) of understory plants in the Andes. I focused on six species for which I confirmed putative pollinators through fieldwork and used the same genotyping technique as above. I found that insect pollination results in a stronger pattern of spatial autocorrelation among closely related individuals, relative to hummingbird pollination. Finally, I
investigated the effect of animal pollination mode and latitudinal region on plant SGS, based on a 147-species global dataset. I found that pollination by small insects is significantly associated with stronger SGS relative to pollination by large insects and vertebrates, particularly in understory plants. Likewise, species from tropical regions have significantly greater SGS than species from temperate zones. Thus, factors that affect plant population genetic differentiation are also important for plant SGS. Overall, my findings shed light on the global drivers of genetic structure in plants, and point to important mechanisms for regional genetic divergence and local genetic connectivity in Andean flowering plants.

**Keywords:** 2b-RAD sequencing, population genetic differentiation, spatial genetic structure, Andes, flowering asynchrony, pollination mode, latitudinal region.
Table of contents

Chapter I: Global patterns of population genetic differentiation in seed plants….5

Chapter II: Flowering asynchrony contributes to genetic divergence in tropical plants ...........................................................................................................................................84

Chapter III: Pollination by hummingbirds strongly decreases genetic structure between and within plant populations relative to pollination by insects.........124

Chapter IV: Impact of animal pollinators and latitudinal regions on the spatial genetic structure of plants: a global test.................................................................164
Chapter I: Global patterns of population genetic differentiation in seed plants

Running title: Drivers of genetic differentiation in plants

Diana Gamba, Department of Biology, University of Missouri at Saint Louis
dlgtk5@mail.umsl.edu

Nathan Muchhalan, Department of Biology, University of Missouri at Saint Louis
muchhalan@umsl.edu

Author for correspondence: Diana Gamba; One University Boulevard, 223R Research Hall, Saint Louis, MO 63121, USA; dlgtk5@mail.umsl.edu; +1 (314) 702-0326.

In review: Molecular Ecology

Abstract: 233 words

Total word count (main text): 6436
Abstract

Evaluating the factors that drive patterns of population differentiation in plants is critical for understanding several biological processes such as local adaptation and incipient speciation. Previous studies have given conflicting results regarding the significance of pollination mode, seed dispersal mode, mating system, growth form, and latitudinal region in shaping patterns of genetic structure, as estimated by $F_{ST}$ values, and no study to date has tested their relative importance together across a broad scale. Here we assembled a 337-species dataset for seed plants from publications with data on $F_{ST}$ from nuclear markers and species traits, including variables pertaining to the sampling scheme of each study. We used species traits, while accounting for sampling variables, to perform phylogenetic multiple regressions. Results demonstrated that $F_{ST}$ values were higher for tropical, mixed-mating, non-woody species pollinated by small insects, indicating greater population differentiation, and lower for temperate, outcrossing trees pollinated by wind. Among the factors we tested, latitudinal region explained the largest portion of variance, followed by pollination mode, mating system and growth form, while seed dispersal mode did not significantly relate to $F_{ST}$. Our analyses provide the most robust and comprehensive evaluation to date of the main ecological factors predicted to drive population differentiation in seed plants, with important implications for understanding the basis of their genetic divergence. Our study is the first that we are aware of to robustly demonstrate greater population differentiation in tropical regions.

**Keywords**: $F_{ST}$, life-history traits, latitudinal region, pollination mode.
**Introduction**

Understanding the factors that drive patterns of genetic variation among plant populations is central in biology because genetic diversity is the raw material on which evolution acts. Quantifying population differentiation, which is most frequently done using the fixation index $F_{ST}$ (Wright, 1951; see Holsinger & Weir, 2009; Meirmans & Hedrick, 2011 for a review of $F_{ST}$ and related metrics), is important for understanding the first stages of allopatric speciation (Harvey, Singhal, & Rabosky, 2019; Templeton, 1981), as well as the basis of local adaptation (Leimu & Fischer, 2008; Linhart & Grant, 1996), and provides critical information for conservation genetics (Ellstrand, 1992; Ellstrand & Elam, 1993; Kramer & Havens, 2009). Life history traits are expected to influence population genetic structure in seed plants (Duminil et al., 2007; Hamrick & Godt, 1996; Loveless & Hamrick, 1984). However, previous studies have given conflicting results as to the importance of specific traits, such as pollination mode, seed dispersal mode, mating system, and growth form (e.g., Duminil et al., 2007; Hamrick & Godt, 1996), and only one study has compared patterns of $F_{ST}$ variation between latitudinal regions (Dick, Hardy, Jones, & Petit, 2008).

Furthermore, little is known about the relative importance of these factors. Below, we discuss prior evidence for each of these factors in turn, and then detail our approach to test them all together in a single analysis that also accounts for phylogenetic relatedness.

Pollination mode is predicted to affect population genetic structure, because pollen dispersal is critical to moving alleles between plant populations.
Previous reviews have lumped different pollination mutualists together as animal pollination and compared them to wind pollination (Hamrick, Godt, & Sherman-Broyles, 1992; Loveless & Hamrick, 1984), revealing that wind tends to reduce genetic structure. Although the idea has not been tested on a broad scale, it has long been thought that different types of animal pollinators should also lead to differences in population genetic structure due to differences in their movement patterns and pollen carry-over capacity (Castellanos, Wilson, & Thomson, 2003). In fact, direct measures of pollen dispersal reveal that volant vertebrates and large bees transport larger proportions of pollen from individual trees to longer geographic distances than small insects (Dick et al., 2008). Given these results, we predict that small insects restrict gene flow among plant populations and increase $F_{ST}$, compared to large insects, vertebrates, or wind.

Seed dispersal mode is also expected to influence plant population genetic structure because, like pollination mode, it directly affects the movement of alleles and thus gene flow among populations. Strong evidence suggests that limited dispersal increases fine-scale spatial genetic structure in plants (Gelmi-Candusso, Heymann, & Heer, 2017) and in other organisms (Aguillon et al., 2017), which in consequence might scale up and lead to greater population genetic structure (Hamrick & Trapnell, 2011). In fact, reviews of the alloseyme literature suggest that seed dispersal by wind and ectozoochory results in lower $F_{ST}$ than dispersal by gravity and endozoochory due to greater gene flow among populations from long distance dispersal events (Hamrick & Godt, 1996; Hamrick, Murawski, & Nason, 1993). However, Duminil et al. (2007) found that
dispersal mode was not a significant predictor of \( F_{ST} \). The lack of consistency among studies encourages further work with larger sample sizes to fully understand the role of seed dispersal mechanisms on population genetic structure.

Unlike pollination and seed dispersal modes, the effect of mating system on plant population genetic structure has been well-established in previous broad-scale studies (Duminil et al., 2007; Loveless & Hamrick, 1984), which suggest that it is the most important predictor of \( F_{ST} \) variation. Mating system affects inbreeding, which lowers within-population variation, inflating between-population \( F_{ST} \) values (Charlesworth, 2003). Duminil, Hardy, and Petit (2009) found that the outcrossing rate and the inbreeding coefficient, which measures biparental inbreeding and selfing, are both significant predictors of \( F_{ST} \) in seed plants. Both selfing and inbreeding increase inbreeding depression and induce purging of deleterious alleles, reducing effective population size and increasing genetic drift, which can ultimately lead to fixation of different alleles in different populations (Angeloni, Ouborg, & Leimu, 2011; Wright, Ness, Foxe, & Barrett, 2008). In contrast, outcrossing increases gene flow within populations, potentially intensifying pollen-mediated gene flow among populations, which counteracts genetic drift and thus decreases population genetic structure (Duminil et al., 2009; Ellstrand, 2014).

Growth form is also an important predictor of population genetic structure. Broad-scale analyses (Duminil et al., 2009; Hamrick et al., 1992) have found strong associations between growth form and \( F_{ST} \), with woody plants tending to
have lower $F_{ST}$ than herbaceous plants. The mechanism that causes this association is unclear, however, and may actually be driven by correlations between growth form and other factors. For example, Duminil et al. (2009) found that growth form only affects $F_{ST}$ indirectly, through its influence on outcrossing rate ($t_m$) and inbreeding coefficient ($F_{IS}$); woody growth form is associated with greater $t_m$ and lower $F_{IS}$. However, Hamrick and Godt (1996) reviewed the allozyme literature for over 300 species and found that when considering outcrossing plants, woody plants show lower levels of $F_{ST}$ than herbs, which suggests that growth form directly affects gene flow among populations, decreasing population genetic structure. This could be because in trees greater geographic distance is presumably required for genetic differences to be detected among populations than in herbs, given that trees are larger than herbs. Thus, when considered at similar geographic scales, we predict that herbs have populations with greater genetic differentiation than trees.

Finally, the latitudinal region in which a plant occurs could also affect its population genetic structure due to differences among regions in spatial and climatic landscapes. In general, geographic heterogeneity and seasonal asynchrony over short distances are considerably higher in the tropics than in the temperate zones (Esquerré, Brennan, Catullo, Torres-Pérez, & Keogh, 2019; Ricklefs, 1977; Stein, Gerstner, & Kreft, 2014), which may act to disrupt mating among conspecific subpopulations, and thus limit gene flow (Martin, Bonier, Moore, & Tewksbury, 2009; Quintero, González-Caro, Zalamea, & Cadena, 2014). Additionally, genetic drift could have a more prominent role in the tropics.
than in the temperate zones, due to the fact that most species in the tropics occur at low population densities and thus should have lower effective population sizes than in temperate zones (Dick et al., 2008; ter Steege et al., 2013). In fact, although their sample size was limited and phylogenetic autocorrelation was not accounted for, Dick et al. (2008) found that tropical trees have on average higher F\textsubscript{ST} values than temperate trees. Given all of the above effects, we predict that F\textsubscript{ST} is higher in the tropics than in the temperate zones.

Previous studies have not included all of the aforementioned factors together when modeling patterns of population genetic structure in seed plants (Duminil et al., 2007; Hamrick et al., 1992; Hamrick & Godt, 1996; Loveless & Hamrick, 1984; Nybom & Bartish, 2000). Furthermore, the most thorough study of F\textsubscript{ST} in seed plants was over a decade ago (Duminil et al., 2007) and thus could not take advantage of the wealth of population genetic studies published since then. Here we reviewed publications to assemble a 337-species database of seed plants with the goal of evaluating the factors predicted to best explain variation in plant population genetic structure. We focused on studies that used nuclear markers because their genetic structure should reflect both pollen and seed movement (due to biparental inheritance), unlike chloroplast markers, which only reflect seed movement (due to maternal inheritance) (McCauley, 1994). We examined five ecological factors, including pollination mode, seed dispersal mode, mating system, growth form, and latitudinal region, while controlling for phylogenetic autocorrelation. We also accounted for variables pertaining to the sampling scheme that have been shown to affect F\textsubscript{ST} values for plants (Nybom &
Bartish, 2000) and other systems (Blasco-Costa & Poulin, 2013; Pascual, Rives, Schunter, & Macpherson, 2017; Riginos, Douglas, Jin, Shanahan, & Treml, 2011); namely, genotyping technique, distance between populations, and sample size. Using multiple regressions, we asked: (Q1) What set of life history traits promote population divergence in seed plants? (Q2) Do patterns of variation in $F_{ST}$ differ between latitudinal regions? (Q3) What are the relative importance of these factors in explaining variation in $F_{ST}$?

**Materials and methods**

**Data collection**

We constructed an $F_{ST}$ dataset through a systematic search in google scholar (key words: “genetic structure”, “population differentiation”, “population genetics”, “genetic diversity”, “population gene flow”) for articles published up until June 2018. The search yielded 356 peer-reviewed publications on seed plants for which measures of population genetic structure ($F_{ST}$) based on nuclear markers were available. When multiple studies reported $F_{ST}$ values for the same species, we recorded the $F_{ST}$ from the study with the largest geographic range, as this may better represent the genetic diversity found in the species (Cavers et al., 2005). By this criterion, we compiled a dataset that included 337 unique species. We extracted information for the predictor variables directly from the publications, and infrequently complemented this, where necessary, with information from peer-reviewed literature on the studied species (see Appendix S1 and Table S1 in Supporting Information). Predictor variables were included in
multiple regressions to explain variation in \( F_{ST} \) values (see section \( F_{ST} \) models).

We included three factors that pertained to the sampling scheme of each study and that can potentially affect \( F_{ST} \) (Nybom, 2004; Nybom & Bartish, 2000): genetic marker used, maximum distance between populations, mean sample size per population. We used them to construct a null model to be compared against models with our factors of interest. Factors of interest consisted of five categorical variables with 2–4 levels: mating system (outcrossing, mixed-mating), growth form (non-woody, shrub, tree), pollination mode (large insects, small insects, vertebrates, wind), seed dispersal mode (animal, gravity, wind), and latitudinal region (tropics, sub-tropics, temperate). Below we explain the \( F_{ST} \) estimates and all eight factors used in this study in greater detail.

**\( F_{ST} \) estimates**

We collected \( F_{ST} \) and \( F_{ST} \) analogs as measures of genetic differentiation (Holsinger & Weir, 2009; Meirmans & Hedrick, 2011) which we collectively refer to \( F_{ST} \) throughout this paper. Assuming an island model of migration-drift equilibrium, Wright (1951) developed a theoretical framework for studying the gene frequency variation among subpopulations through the fixation indices, i.e. F-statistics. In this model, \( F_{ST} \) is the degree of gene differentiation among subpopulations for genes that have only two alleles. Nei (1973) expanded the model for polymorphic genes, and proposed \( G_{ST} \) as a measure of the gene diversity partitioned among subpopulations, relative to the total gene diversity of the population. Subsequently, Weir & Cockerham (1984) proposed a standard
measure of genetic structure $\theta$ based on Wright (1951). The statistic $\theta$ is estimated per and across loci, and represents the correlation of genes, or coancestry, among individuals in a given population. Excoffier, Smouse, and Quattro (1992) proposed AMOVA (Analysis of Molecular Variance) and corresponding statistic $\phi_{ST}$; the proportion of genetic diversity partitioned among populations. Finally, Hedrick (2005) proposed a standardized measure of population differentiation, $G'_{ST}$, which accounts for the level of heterozygosity of the marker used for genotyping individuals ($G'_{ST} = G_{ST \text{overall}} / G_{ST \text{max}}$).

The most common statistic in our dataset was $\theta$. When $\theta$ was reported per loci, we took the mean across loci as the global $F_{ST}$ for that species. The AMOVA derived $\phi_{ST}$ was also common. Some studies reported both $\theta$ and $\phi_{ST}$, in which case we used $\phi_{ST}$ as it likely better represents genetic structure among populations (Hey & Pinho, 2012). The statistics $\theta$ and $\phi_{ST}$ were, however, frequently almost equivalent. Another common measure was $G_{ST}$; when reported for multiple pairs of populations, we used the mean across all pairs. A few studies reported $G'_{ST}$. It was not possible to back-transform $G'_{ST}$ to $G_{ST}$ because such studies did not report the maximum possible $G_{ST}$ in their data (Hahn, Michalski, Fischer, & Durka, 2016). Even though $G'_{ST}$ potentially yields a higher value than $G_{ST}$ (or $\theta$ and $\phi_{ST}$) based on the same data (Hedrick, 2005; Meirmans & Hedrick, 2011), we still included $G'_{ST}$ values, reasoning that any trend of variation in population genetic structure due to the variables here tested should still be present.
**Molecular markers**

$F_{ST}$ values can be strongly affected by the genotyping technique implemented (Nybom, 2004; Nybom & Bartish, 2000; Meirmans & Hedrick, 2011), thus, we included this factor in our null model. In our database, the majority of studies used nuclear microsatellites (140 species), followed by allozymes (114 species). Fewer studies used dominantly inherited markers, including Amplified Fragment Length Polymorphism (60 species), Random Amplification of Polymorphic DNA (16 species), and Inter-Simple Sequence Repeat (7 species).

**Distance between populations**

Greater distance between populations should correspond to greater genetic differentiation based on an isolation by distance model (Wright, 1943). Thus, we also included in our null model the maximum distance between populations used in each study. We calculated this based on the coordinates of the two most distant populations. When this was not available, we used the scale bar of maps showing sampled populations. Distance varied from 0.01–9900 km (mean=703 ± 1077 SD).

**Mean sample size per population**

The maximum value that $F_{ST}$ can take decreases when the within-population expected heterozygosity increases. Thus, a general concern is that large sample sizes are required because small samples can overestimate $F_{ST}$.
(Holsinger & Weir, 2009; Kalinowski, 2005; Willing, Dreyer, & van Oosterhout, 2012). We accounted for this potential bias by including the mean sample size per population in our null model. Across the studies, this sample size ranged from 3 to 285 individuals per population, with an overall mean of 40.12 (± 44.9 SD).

**Pollination mode**

Species were coded as pollinated by wind, small insects, large insects, or vertebrates. Small insect pollinators included small Hymenoptera (i.e., *Trigona* and *Melipona* bees and wasps), Diptera (i.e., hoverflies and gnats), Coleoptera (i.e., small curculionids), Hemiptera (i.e. Anthocoridae and Miridae), and Thysanoptera (i.e., thrips). Large insects included large bees (i.e., honeybees, bumblebees, carpenter bees, euglossine bees) and Lepidoptera (i.e., hawk moths and yucca moths, monarch butterflies). We included honeybees in the large insect category based on evidence showing that honeybees have flying and pollen carry-over capacity similar to bumblebees (Cresswell, Bassom, Bell, Collins, & Kelly, 1995; Escaravage & Wagner, 2004). Vertebrates included bats, hummingbirds, and other nectarivorous birds such as honeyeaters and sunbirds. Some instances of vertebrate pollination were more generalized, with visitors including a combination of bats, birds, rodents, and/or marsupials.

**Seed dispersal mode**

Species were coded as dispersed by wind, animals, or gravity. Plants
adapted to wind dispersal presented fruits or seeds that were particularly light and/or winged. For those plants adapted to animal dispersal, exploratory analyses showed that different types of animal dispersal were not significantly different (results not shown). Thus, we kept the animal dispersal category broad, including plants with fruits or seeds dispersed by endo-, ecto-, or syn-zoochory. Plants with no adaptations for vector-mediated seed dispersal were coded as gravity dispersed. Based on the information reported in publications with \( F_{ST} \) and trait data, we did not find evidence of secondary movement of fruits or seeds by biotic agents. In some instances, however, water may play a secondary role in dispersing seeds that fall under mother plants, as in the mangrove species \textit{Avicennia} spp. and \textit{Rhizophora} spp., and for \textit{Beta vulgaris} L., \textit{Casuarina cunninghamiana} Miq., \textit{Cocos nucifera} L., and \textit{Primula nutans} Georgi, as well as for many forest trees after floods or inhabiting riparian sites (Levine & Murrell, 2003; Nilsson, Brown, Jansson, & Merritt, 2010).

**Mating system**

We coded species as selfing, mixed-mating, or outcrossing, as identified by the authors in each study. Selfing species included strictly autogamous species. They were rare (\( N=7 \)) and not included in the final 337-species dataset, due to their low sample size. Mixed-mating species included those that undergo both outcrossing and selfing to some extent, through either autogamy or geitonogamy (Goodwillie, Kalisz, & Eckert, 2005). Outcrossing species included plants that are self-incompatible, unisexual (i.e. monoecious or dioecious), or
dichogamous hermaphrodites; i.e. either having the male reproductive organs come to maturity before the female organs (protandry), or vice versa (protogyny).

**Growth form**

Species were coded as trees, shrubs, or non-woody plants. Trees included woody plants >10m tall, typically with a single trunk coming from the base. Shrubs included upright woody plants <10 m tall, typically with one or several trunks coming from the base. We also included in the shrub category hemi-parasites and hemi-epiphytes. Non-woody plants included herbs, epiphytes, and non-woody climbers. Growth form of species was often linked to habitat in that many non-woody plants and shrubs occurred in the forest understory, while many trees occurred in the subcanopy and canopy. However, non-woody plants, shrubs, and trees also occurred in open habitats like prairies. We did not include habitat as an additional predictor in our models due to its high collinearity with growth form.

**Latitudinal region**

We recorded the geographic location of each study to create an additional categorical variable for latitudinal region. Species were coded as tropical, sub-tropical, or temperate. Tropical regions included sites between the tropics of Cancer and Capricorn (23.5° north and south of the equator, respectively), which are characterized by relatively low variation in daylight and temperature throughout the year, but with large environmental heterogeneity over short
distances. Sub-tropical regions included latitudes from 23.5° to 35° (north and south). These regions have climates similar to the tropics, but with more seasonal fluctuations. Temperate regions included latitudes greater than 35° north and south. These zones are characterized by a wide range of temperatures throughout the year, and by clearly marked seasonal changes.

**Analytical framework**

Analyses were performed in R (R Core Team 2018). Prior to model testing, we performed transformations of continuous data to improve normality of model residuals (details in Appendix S2). $F_{ST}$ was transformed using Tukey’s ladder of powers transformation (Tukey, 1970) with the function `transformTukey` from the R package `rcompanion` (Mangiafico, 2018). Continuous predictors were transformed using their natural logarithm. We also estimated correlations (Plackett, 1983) and evaluated multicollinearity issues (Acock & Stavig, 1979; Fox & Monette, 1992) among predictor variables (Appendix S3). The multicollinearity tests indicated that all predictors could be included together in a multiple regression (Table S2 and Table S3).

In order to calculate and subsequently perform models that correct for phylogenetic signal (Freckleton, Harvey, & Pagel, 2002), a species-level phylogeny (Fig. S1) was produced with the R package `V.PhyloMaker` (Jin & Qian, 2019). This package prunes a custom list of species from the latest and most complete mega-tree of vascular plants (Smith & Brown, 2018) (see Appendix S4 for details). We then assessed phylogenetic signal in categorical predictors with
Abouheif's (1999) method (Jombart, Balloux, & Dray, 2010; Pavoine, Ollier, Pontier, & Chessel, 2008), and in $F_{ST}$ values with Pagel's (1999) $\lambda$ (Molina-Venegas & Rodríguez, 2017; Revell, 2012) (Appendix S5). We found that closely related species tend to be more similar than expected by chance in their mating system, growth form, pollination mode, seed dispersal mode, latitudinal region and $F_{ST}$. The highest observed Moran's $I$ was that of growth form, followed by pollination mode, latitudinal region, seed dispersal mode, and lastly mating system (Fig. S2). $F_{ST}$ values were also phylogenetically autocorrelated (Pagel's $\lambda=0.52$, $P<0.001$ and Pagel's $\lambda=0.53$, $P<0.001$ for raw and transformed $F_{ST}$ values, respectively). Given the high levels of phylogenetic signal, we implemented phylogenetically informed multiple regressions (Symonds & Blomberg, 2014) with the function ‘phylolm’ from the R package phylolm (Ho & Ané, 2014). For the fit of models, the likelihood of the parameters was calculated with a Brownian motion model of evolution (Ho & Ané, 2014) (Appendix S6).

Finally, for the categorical predictors with more than two levels we chose reference levels based on exploratory analyses with phylogenetic ANOVA and post-hoc tests (Garland, Dickerman, Janis, & Jones, 1993; Revell, 2012). We selected the level which mean was most different from that of other levels (Tables S4 and S5). Reference levels were as follow: trees for growth form, small insects for pollination mode, gravity for dispersal mode, and temperate for latitudinal region.
**Fst models**

We began our phylogenetic multiple regressions analyses of factors affecting genetic structure by constructing a null model with the sampling-scheme variables. We sequentially added the life history traits to this null model, checking whether each addition improved model fit of a multiple regression based on Akaike Information Criterion (AIC) scores (Akaike, 1974). Mating system and growth form were added together as there is ample evidence of their effect on Fst (Duminil et al., 2007; Hamrick & Godt, 1992). We then added pollination mode and seed dispersal mode, to check whether either, or both together, improved the previous model. After finding the best model explaining Fst with life history traits (Q1), we compared this model to one that included latitudinal region as an additional factor (Q2). We assessed the variance explained by each model with the R package rr2 and the function ‘R2.pred’ (Ives, 2018; Ives & Li, 2018).

We further evaluated the best-fit model through a backward stepwise model selection with the function ‘phylostep’ in the phylolm package. The functions ‘phylostep’ and ‘phylolm’ were congruent in finding the same best model.

We then evaluated the importance of each variable in this best-fit model (Q3). We used the R package rr2 and the function ‘R2.lik’ to obtain the unique contribution of each factor in terms of the amount of Fst variance explained by comparing the best-fit model with a reduced model not including the factor of interest.
Results

Taxonomic scope and phylogeny

The 337 species were distributed in 210 genera, representing 96 families in 34 orders. The majority of species (268) belonged to the Eudicots, followed by 43 Monocots, 17 Magnoliids, and 9 Gymnosperms. The families Fabaceae (mostly *Acacia*; 8 species) and Fagaceae (mostly *Quercus*; 13 species) were particularly well represented, with 37 and 26 species respectively (Table S1). The resulting phylogeny had 337 tips and 311 nodes (Fig. S1). In other words, 92% of the phylogeny was resolved, and only 26 tips (8%) belonged to polytomies. These polytomies correspond to clades for which phylogenetic information remains scarce or unclear (Stevens, 2001 onwards): *Begonia* (Begoniaceae), *Alcantarea* and *Encholirium* (Bromeliaceae), *Streptocarpus* (Gesneriaceae), *Arceuthobium* (Santalaceae), *Magnolia* (Magnoliaceae), *Piper* (Piperaceae), *Psychotria* (Rubiaceae), *Acacia* (Fabaceae), and *Sorbus* (Rosaceae).

Life history traits that promote population divergence in seed plants (Q1)

Among phylogenetic multiple regressions with the four life history traits (models 1–4, Table 1), model 4 was the best-fit, indicating that mating system, growth form, pollination mode and seed dispersal mode all influence $F_{ST}$ ($\text{AIC} = -482.3$). However, the performance of model 4 was almost indistinguishable from that of model 3 ($\Delta\text{AIC} = 2.2$), which only differed in the lack of the factor seed dispersal mode. Further evidence for the relative unimportance of seed dispersal mode can be seen in the fact that adding seed dispersal mode to model 1 (which
only has mating system and growth form) results in much less improvement of fit (models 2 vs. 1, ΔAIC=2.5) than adding pollination mode (models 3 vs. 1, ΔAIC=16.6).

**Differences among latitudinal regions (Q2)**

Adding the factor latitudinal region to models with the four life history traits notably increased fit to the data (models 5–7, AIC=−488.6 to −503.9, Table 1). This is particularly evident when comparing the best-fit models for each instance (models 4 vs. 6, ΔAIC=21.6). Model performance was indistinguishable for models 6 vs. 7 (ΔAIC=1), which only differed in the addition of seed dispersal mode. Finally, in models 5 and 7 the factor seed dispersal mode was no longer a significant predictor of F_{ST} (Table 1 and 2). Below we focus on results from model 7, as it is the most inclusive model of the factors we tested with the best fit to the data.

Figure 1 shows how the levels of each factor affect population differentiation as measured by F_{ST} values (after transformation). The effect of each factor is depicted after accounting for the effect of the other independent variables in model 7. For mating system, outcrossers tend to have lower population differentiation than mixed-mating plants (Fig. 1a). Trees tend to have significantly lower population differentiation relative to non-woody plants and shrubs, while the latter two growth forms did not differ between each other (Fig. 1b). Pollination by small insects leads to significantly greater differentiation compared to large insect, vertebrate and wind pollination, while the latter three
pollination modes did not differ between each other (Fig. 1c). Temperate zones have significantly lower $F_{ST}$ values than tropics and subtropics, and the latter two regions did not differ from each other (Fig. 1e). Finally, seed dispersal mode was not a significant predictor of population genetic differentiation. $F_{ST}$ values associated with gravity dispersal were highly variable, and although gravity dispersal results in higher $F_{ST}$ values compared to wind dispersal, this difference was not significant. Animal dispersal also resulted in highly variable $F_{ST}$ values that did not differ from other dispersal modes (Fig. 1d).

**Most important factor for explaining $F_{ST}$ (Q3)**

Of all of the factors that we analyzed, latitudinal region explained the highest percent variation (7%), higher than the life history traits in model 7 (0.9–6%, Fig 1f). Of the life history traits, mating system and pollination mode had the highest independent contribution to the variation in $F_{ST}$ values (6% each), followed by growth form (4%), while the contribution of dispersal mode was very low (0.9%) and not statistically significant (Fig. 1f).

**Influence of variables in the null model**

Variables in the null model were significant predictors of $F_{ST}$ in all multiple regressions (Table 1) and in model 7 (Table S6). Distance had the highest independent contribution (8%), compared to genetic marker and mean sample size (4% each). In general, $F_{ST}$ values become larger when the geographic scale of studies increases. In contrast, $F_{ST}$ values decrease with larger mean sample
sizes of individuals per population. Codominant markers (microsatellites and allozymes) tend to underestimate $F_{ST}$ values, while dominant markers (AFLP and RAPD) overestimate them. ISSR markers did not differ from others.

Discussion

Here we provide the most robust and comprehensive evaluation to date of factors driving population genetic differentiation in seed plants. We largely found support for our hypothesis of factors that significantly influence $F_{ST}$ and several intriguing patterns emerge from our analyses. Overall, we found higher $F_{ST}$ for tropical, mixed-mating, non-woody species pollinated by small insects, and lower $F_{ST}$ for temperate, outcrossing, trees pollinated by wind. Latitudinal region was the most important predictor for $F_{ST}$ relative to the others tested. Mating system and pollination mode had equal contributions for explaining $F_{ST}$. Growth form was also a key factor influencing $F_{ST}$, while seed dispersal mode was not important in our most inclusive model (Table 2, Fig. 1).

Influence of latitudinal region on $F_{ST}$

Population differentiation was higher in the tropics and subtropics than in temperate regions (Fig. 1e). This result supports the idea that patterns of local diversity, such as the partitioning of genetic diversity among plant populations, cannot be explained in isolation from the geographic and historic processes of each region (Ricklefs, 1987, 2004, 2006). Some factors that may contribute include regional differences in seasonality, macroevolution, and geography,
differences which have more generally been hypothesized to contribute to the latitudinal diversity gradient (i.e. increased species richness closer to the equator) (Mittelbach et al., 2007; Rolland, Condamine, Jiguet, & Morlon, 2014; Schemske, Mittelbach, Cornell, Sobel, & Roy, 2009). Below we discuss some of these ideas, including the ‘asynchrony of seasons hypothesis’ (ASH) (Martin et al., 2009), the ‘time/area hypothesis’ (Fine & Ree, 2006), and the ‘niche conservatism hypothesis’ (Kerkhoff, Moriarty, & Weiser, 2014).

One compelling explanation for the regional differences in $F_{ST}$ is based on the idea that the tropics can have highly asynchronous rainfall patterns over small spatial scales (Martin et al., 2009). Given that most plants time their flowering to seasons (Crimmins, Crimmins, & Bertelsen, 2011; Gaudinier & Blackman, 2019), and that seasons are largely determined by rainfall in the tropics, small-scale differences in rainfall potentially disrupt gene flow and cause high population differentiation over short distances compared to the temperate zones. This is the aforementioned ASH, and our analyses support the prediction of higher population differentiation in the tropics. We note that the tropics and subtropics did not differ in $F_{ST}$, and that these regions have comparable climatic patterns (Sitnikov, 2009), thus the ASH may extend to subtropical regions.

Higher $F_{ST}$ in the tropics/subtropics than in the temperate zones can also be due to the different history of plant lineages in each region. The ‘time/area hypothesis’ (Fine & Ree, 2006) and the ‘niche conservatism hypothesis’ (Kerkhoff et al., 2014) allude to the idea that tropical clades are older and tend to live in the same environments throughout their evolutionary history, while
temperate clades diversified more recently after switching to novel environments once cooling began in the Oligocene. Thus, most temperate species likely expanded their populations fairly recently post-glaciation (34 Mya), resulting in lower population differentiation due to recent gene flow maintaining cohesion. In contrast, tropical species may have been in the same place longer and their populations have had more time to isolate due to dispersal limitations and build up genetic differentiation (Kisel & Barraclough, 2010; Smith et al., 2014). Tropics and subtropics share strong floristic affinities (Sarmiento, 1972), which corresponds to the similar $F_{ST}$ between them.

Finally, gene flow is likely more restricted in the tropics due to its heterogeneous orogeny and rich fluvial systems. Such geographic differences have also been hypothesized to contribute to the latitudinal diversity gradient (e.g., Smith et al., 2014; Wallace, 1854). This argument becomes particularly compelling in combination with the fact that temperature does not vary as extremely through the year in the tropics. Given this, different subpopulations would be expected to evolve narrower physiological niches that adapt them to particular altitudinal zones, and a similarly sized mountain would impose a greater barrier to dispersal, and thus to gene flow among subpopulations, in tropical than in temperate regions (Ghalambor, 2006; Janzen, 1967).

Thus, overall, our results are in line with hypotheses that suggest greater species diversity in the tropics is due to higher speciation rates rather than lower extinction rates. While the specific mechanisms differ, including those mentioned above and others (see Mittelbach et al., 2007), these hypotheses all posit greater
population-level differentiation that then scales up to faster speciation rates in a model of allopatric or parapatric speciation. Direct tests on the influence of population differentiation on speciation rates are necessary in order to establish that population differentiation is a rate-limiting step of the speciation process (Harvey et al., 2019). Such tests are scarce and have only focused on vertebrates, finding a positive association in New World birds (Harvey et al., 2017), and no association in Australian lizards (Singhal et al., 2018). We encourage similar tests in seed plants at a global scale. Nevertheless, ours is the first study that we are aware of to clearly document such a pattern of greater population differentiation in the tropics for seed plants (see Martin & McKay, 2004 for a study in vertebrates).

**Influence of pollination mode on F_{ST}**

We found that pollination mode plays a key role in population differentiation, contrary to the findings of the latest review of F_{ST} and species traits in seed plants (Duminil et al., 2007). Specifically, species pollinated by small insects have significantly higher F_{ST} than those with other pollination modes. This pattern is likely due to reduced gene flow among plant populations. In fact, small insects have a lower pollen carry-over capacity than bumblebees and vertebrates (Dick et al., 2008; Rhodes, Fant, & Skogen, 2017), and studies of pollinator movement show that euglossine bees, hawkmoths, and bats can all travel long distances, even across fragmented habitats (Brunet, Larson-Rabin, & Stewart, 2012; Finger, Kaiser-Bunbury, Kettle, Valentin, & Ghazoul, 2014;
Janzen, 1971; López-Uribe, Oi, & Del Lama, 2008; McCulloch et al., 2013; Skogen, Overson, Hilpman, & Fant, 2019). Our results show that wind, large insects, and vertebrates have homogenizing effects on plant $F_{ST}$, which are statistically indistinguishable. Taken together, these patterns suggest that plants pollinated by small insects might be more sensitive to habitat fragmentation; the inability of these pollinators to connect distant fragments may decrease genetic diversity within populations, and along with it the ability to adapt in response to anthropogenic change.

One important caveat is that the limited information on pollination systems for many species necessitated a relatively coarse-grained division of pollination mode into broad taxonomic groups. This approach overlooks potential behavioral differences within these groups. For instance, within the vertebrate pollination category, territorial hummingbirds likely move pollen much shorter distances than trap-lining hummingbirds (Betts, Hadley, & Kress, 2015; Ohashi & Thomson, 2009), and bats may carry pollen more efficiently (Muchhala & Thomson, 2010) and to longer distances than hummingbirds (Lemke, 1984, 1985; Tello-Ramos, Hurly, & Healy, 2015).

**Influence of mating system on $F_{ST}$**

Our results provide additional support for the idea that mating system is a strong predictor of $F_{ST}$ (Fig. 1a), even in the presence of other factors (Duminil et al., 2007). Mating system associates with $F_{ST}$ because any amount of inbreeding (through mixed-mating) increases homozygosity within a subpopulation, and
reduces its effective population size, leading to increased population structure due to genetic drift. In contrast, outcrossing maintains genetic cohesion within and among subpopulations, decreasing genetic drift and reducing population structure (Charlesworth, 2003). Because populations of mixed-mating species are often highly differentiated, they will likely have populations with unique genetic diversity. Accordingly, conservation efforts for them should maximize the number of populations protected to maximize genetic diversity to increase their chances to adapt to environmental change (Ellstrand & Elam, 1993).

**Influence of growth form on F<sub>ST</sub>**

We found that trees have populations with significantly lower F<sub>ST</sub> than both shrubs and non-woody plants (Fig. 1b). Even though most trees are outcrossing in our dataset, our results show that growth form contributes to the variation in F<sub>ST</sub> independently from mating system, contrary to the findings of Duminil et al. (2007, 2009). The inherent difference in scale between growth forms may contribute to this pattern: a given geographic distance between subpopulations may restrict gene flow much more for an herb than for a tree. In fact, neighborhood size, i.e. the spatial extent of closely related individuals, is larger in trees than shrubs and herbs (Vekemans & Hardy, 2004). Furthermore, trees usually have greater longevity than shrubs and non-woody plants (Duminil et al., 2009), which may increase the chances of gene flow between tree subpopulations, more than for other growth forms. Finally, the fact that growth form and habitat are tightly linked may also contribute; many non-woody plants
and shrubs in our dataset occur in the forest understory, while many trees reach the canopy. Givnish (2010) and Theim, Shirk, and Givnish (2014) hypothesized that the understory imposes more limits to gene flow than the canopy because of the sedentary lifestyle of animal mutualists in the understory.

**Seed dispersal and F\text{ST}**

Our results did not support the hypothesis that gravity-mediated seed dispersal increases population differentiation compared to wind or animal dispersal (Givnish, 2010) (Fig. 1d). This is in line with previous findings suggesting that the genetic structure of nuclear markers is largely driven by pollen flow (Petit et al., 2005; Sork, Nason, Campbell, & Fernandez, 1999; Skogen et al., 2019), and that the effect of seed dispersal is only detectable in the population genetic structure of chloroplast genes (Duminil et al., 2007). However, we note that gravity dispersal resulted in highly variable F\text{ST} values, potentially due to unrecorded secondary seed vectors. F\text{ST} values for animal dispersal were also highly variable, which suggests that different animals could have different effects on population differentiation. Thus overall, as with vertebrate pollination, we suspect that more fine-scaled classifications of dispersers may improve our understanding of their effects on plant population genetic structure. Testing this idea, however, requires more detailed data on animal dispersal modes, which can be difficult to characterize. For example, in our study many species have a mix of seed dispersers, including small to large mammals and birds (like most Arecaceae, Fabaceae, Fagaceae, Myrtaceae,
Sapotaceae, among others), making it difficult to assign plants to a disperser-specific taxonomic affiliation or foraging behavioral trait.

**Considerations on model inference**

Phylogenetic multiple regressions allowed us to evaluate the unique effect of each predictor on $F_{ST}$ while correcting for phylogenetic autocorrelation, which had not been accomplished in previous broad-scale studies. Additionally, we note that after adding the factor latitudinal region, the scaling parameter that corrects for phylogenetic autocorrelation ($\lambda$ fit in Table 1) became insignificant. This suggests that latitudinal region decreases the phylogenetic autocorrelation in the residuals modeled by our phylogenetic regressions (Freckleton, 2009). In fact, an alternative across-species multiple regression of model 7 (i.e., a linear model assuming phylogenetic independence) yielded identical results with indistinguishable fit to the data ($\Delta$AIC=1.9). We suspect that region captured important phylogenetic information in $F_{ST}$ and species traits; within each regional species pool, lineages share strong biogeographic and phylogenetic affinities. Put another way, we think that regional affiliation is the most important underlying factor influencing $F_{ST}$ values at a global scale, and when not included, phylogenetic signal becomes a proxy for latitudinal region due to the tendency for closely related species to occur in similar regions.

**Future directions**

Understanding how plant population genetic structure is affected by life
history traits can greatly improve management strategies for populations facing increasingly fragmented habitats due to human-accelerated global change. Our study reveals that gene flow is generally more limited in non-woody species pollinated by small insects, making them more susceptible to isolation and loss of genetic diversity. Thus, in order to preserve the largest amount of genetic diversity for species with such traits, conservation efforts should seek to maintain numerous subpopulations spanning a wide geographic extent. Future broad-scale studies of $F_{ST}$ variation could provide more even greater insights for conservation by including population densities (Murawski & Hamrick, 1991; Sork et al., 1999), effects of habitat fragmentation (Aguilar, Quesada, Ashworth, Herreras-Diego, & Lobo, 2008; Skogen et al., 2019), and the landscape context of populations (Sork et al., 1999).

Another avenue for future research involves linking patterns of genetic variation at different scales. Little is known about how factors that affect genetic patterns over fine spatial scales (i.e., within subpopulations) extend to genetic patterns over larger spatial scales (i.e., among subpopulations). Intuitively, species with greater fine-scale genetic structure (Loiselle, Sork, Nason, & Graham, 1995) should also have greater population genetic structure, but this has rarely been tested. For example, a recent review found greater fine-scale genetic structure in species with short-distance dispersers, than those dispersed by birds (Gelmi-Candusso et al., 2017), but it is unclear whether this difference would extend over larger distances. Overall, we expect that more comprehensive studies of ecological interactions, in combination with increasing amounts of
genetic data collected at various spatial scales will continue to improve our understanding of the factors that influence population genetic structure in seed plants.

Acknowledgements

We thank the researchers whose published data we used in this paper. Thanks to Robert Ricklefs, Christine Edwards, and Carmen Ulloa for advice in this study. Isabel Loza, Justin Baldwin and Sebastián Tello provided valuable help with statistical analyses. Many thanks to Justin Zweck, Justin Baldwin, Krissa Skogen, and to members of the Muchhala lab at the University of Missouri at Saint Louis for constructive discussions on a previous version of this manuscript. This research was supported by the Whitney Harris World Ecology Center at the University of Missouri at Saint Louis.

References


Data accessibility statement

Should the manuscript be accepted, the data and R scripts supporting the results will be archived in Dryad and their DOI will be included at the end of this article. No new data were used in this research because analyses were based on a literature review of published studies.

Author Contributions

DG and NM planned and designed the research. DG collected and analyzed the data. DG wrote the first draft of the manuscript. DG and NM contributed equally to substantial revisions of the manuscript.
**Table 1** Phylogenetic multiple regressions explaining variation in $F_{ST}$. In each model only the main effect of factors is considered, i.e., no interactions. AIC and $\lambda$ fit (scaling parameter to correct for phylogeny) were estimated using maximum likelihood. Underlined variables indicate that at least one of their terms was a significant factor in the corresponding model. (Thick underline: $P \leq 0.005$, thin underline: $0.005 < P < 0.05$) (next page).
<table>
<thead>
<tr>
<th>MODEL</th>
<th>Variables †</th>
<th>$R^2$</th>
<th>AIC</th>
<th>$\lambda$ fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>genetic marker</td>
<td>0.36</td>
<td>-437</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>mean sample size ‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance §</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>null model</td>
<td>0.41</td>
<td>-463.5</td>
<td>0.48</td>
</tr>
<tr>
<td>Model 2</td>
<td>null model</td>
<td>0.42</td>
<td>-466</td>
<td>0.46</td>
</tr>
<tr>
<td>Model 3</td>
<td>null model</td>
<td>0.43</td>
<td>-480.1</td>
<td>0.37</td>
</tr>
<tr>
<td>Model 4</td>
<td>null model</td>
<td>0.44</td>
<td>-482.3</td>
<td>0.35</td>
</tr>
<tr>
<td>Model 5</td>
<td>null model</td>
<td>0.42</td>
<td>-488.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Model 6</td>
<td>null model</td>
<td>0.45</td>
<td>-503.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Model 7</td>
<td>null model</td>
<td>0.46</td>
<td>-502.9</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

† yellow circle: mating system, green circle: growth form, brown circle: seed dispersal mode, red circle: pollination mode, blue circle: latitudinal region.

‡ mean sample size: natural logarithm of the mean sample size of individuals per population.

§ distance: natural logarithm of the maximum distance between populations.
Table 2 Details of model 7, the most inclusive phylogenetic model with factors of interest. Variables in bold indicate the reference level for each categorical factor. N indicates the sample size of each group without phylogenetic correction. Significant P values are in bold.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>T value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>0.59</td>
<td>0.04</td>
<td>14.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mating system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mixed-mating</strong></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcrossing</td>
<td>257</td>
<td>-0.07</td>
<td>0.01</td>
<td>-4.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Growth form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tree</strong></td>
<td>163</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-woody</td>
<td>121</td>
<td>0.09</td>
<td>0.02</td>
<td>5.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Shrub</td>
<td>53</td>
<td>0.06</td>
<td>0.02</td>
<td>3</td>
<td>0.003</td>
</tr>
<tr>
<td>Pollination mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Small insects</strong></td>
<td>176</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large insects</td>
<td>48</td>
<td>-0.06</td>
<td>0.02</td>
<td>-3.4</td>
<td>0.001</td>
</tr>
<tr>
<td>Vertebrates</td>
<td>44</td>
<td>-0.05</td>
<td>0.02</td>
<td>-2.6</td>
<td>0.01</td>
</tr>
<tr>
<td>Wind</td>
<td>69</td>
<td>-0.05</td>
<td>0.02</td>
<td>-3</td>
<td>0.003</td>
</tr>
<tr>
<td>Seed dispersal mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gravity</strong></td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animals</td>
<td>147</td>
<td>-0.003</td>
<td>0.02</td>
<td>-0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Wind</td>
<td>108</td>
<td>-0.02</td>
<td>0.02</td>
<td>-1.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Latitudinal region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Temperate</strong></td>
<td>134</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-tropical</td>
<td>78</td>
<td>0.07</td>
<td>0.02</td>
<td>4.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Tropical</td>
<td>125</td>
<td>0.09</td>
<td>0.02</td>
<td>5.4</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Fig. 1 Partial regression plots showing the effect of each factor on transformed $F_{ST}$ values after accounting for the effect of other independent variables in model 7 (i.e., adjusted $F_{ST}$). Parallel boxplots of the partial residuals are drawn for the levels of each factor along with significant differences between groups depicted by the upper horizontal grey lines according to model 7 (Table 2): (a) mating system, (b) growth form, (c) pollination mode, (d) seed dispersal mode, and (e) latitudinal region. Thick horizontal black lines are median values, boxes indicate 25% and 75% quartiles, whiskers are maximum and minimum values, white circles are outliers. (f) Relative importance of each factor ($\Delta R^2$ value); the change in $R^2$ after each individual factor is removed from model 7 (next page).
Additional supporting information that will appear in the expanded online version of this article:

**Appendix S1.** References of publications with data on F_{ST} and species traits used in this study.

**Appendix S2.** Data transformation.

**Appendix S3.** Tests of multicollinearity.

**Appendix S4.** Phylogeny.

**Appendix S5.** Phylogenetic signal.

**Appendix S6.** PhyloLM implementation.

**Fig. S1.** Phylogeny of studied species.

**Fig. S2.** Estimation of phylogenetic signal on model variables.

**Table S1.** Dataset used in this study (in Table S1.xlsx).

**Table S2.** Correlation tests between categorical variables.

**Table S3.** Estimates of the generalized variance inflation factor on predictors.

**Table S4.** Results from phylogenetic ANOVA on F_{ST}.

**Table S5.** Pairwise post-hoc tests between groups within each categorical variable, estimated after performing phylogenetic ANOVA.

**Table S6.** Details of model 7 including variables in the null model.
Appendix S1. References of publication with $F_{ST}$ data and species traits used in this study.


Broadhurst LM, Coates DJ. (2004). Genetic divergence among and diversity within two rare *Banksia* species and their common close relative in the


Molecular Ecology 15: 559–571.


Mimosoideae) in La Montaña of Guerrero, Mexico. *Genetic Resources and Crop Evolution* 52: 941–957.


**Appendix S2.** Data transformation.

We applied transformations to continuous variables in order to improve normality.

$F_{ST}$ was transformed using Tukey’s ladder of powers transformation (Tukey, 1970) with the function `transformTukey` from the R package `rcompanion` (Mangiafico, 2018). This function finds the power that makes a variable as normally distributed as possible based on the Shapiro-Wilk test (Shapiro & Wilk, 1965). Transformed $F_{ST}$ resulted in $F_{ST}^{0.275}$ (Shapiro-Wilk statistic=0.27, $P=0.7$). For continuous predictors, the best transformation to improve normality was the natural logarithm of the maximum distance between populations and the mean sample size per population.

**Appendix S3.** Tests of multicollinearity.

Because multicollinearity can complicate the identification of an optimal set of explanatory variables for a statistical model, we assessed the correlation between species traits. We calculated the Pearson Chi-Square test of independence (Plackett, 1983), which is appropriate for categorical data, between all pairs of variables. We then calculated Cramer V values, which gives
a measure of the strength of the association, using the R functions `chisq.test` and `cramerV`. Cramer V values less than 0.3 represent a moderately low association and excluding associations higher than 0.3 helps prevent multicollinearity issues (Acock & Stavig, 1979). We also estimated the variance inflation factor generalized to account for degrees of freedom of each factor (GVIF, Fox & Monette, 1992) with the R function `VIF`. GVIF values smaller than 5 are generally considered to not cause collinearity problems in model inferences. All Cramer V values were ≤0.3 and GVIF values were <2 (Table S2 and S3). Thus, multicollinearity did not affect our model inference.

**Appendix S4.** Phylogeny.

A species-level phylogeny was produced with the R package V.PhyloMaker (Jin & Qian, 2019). This program uses as the backbone tree the latest seed plant mega-phylogeny (Smith & Brown, 2018), which is inferred from seven nuclear regions retrieved from GenBank and fossil calibrated to include branch lengths. Species are pruned from this backbone tree based on a custom species list. Species not present in the backbone tree were added as polytomies within their respective clade using the same method as Phylomatic (Webb & Donoghue, 2005), with a branch length calculation as implemented with the branch length adjuster algorithm (Webb et al., 2008). Qian & Jin (2016) showed that such approach results in phylogenies very similar to empirical species-level phylogenies. Of the 337 species in our dataset, 239 were already in the backbone tree and 98 were newly added. After these additions, V.PhyloMaker
pruned our custom phylogenetic tree to remove tips not in our dataset. Because V.PhyloMaker assigns age divergences to particular nodes in the target topology, and then places the remaining nodes evenly between them, the resulting time-calibrated tree is actually a pseudo-chronogram. Pseudo-chronograms show lower variability in branch length than well-calibrated phylogenies that use molecular clocks, yet they remain appropriate for phylogenetic comparative methods (Molina-Venegas & Rodríguez, 2017).

**Appendix S5.** Phylogenetic signal.

For categorical traits, we performed Abouheif’s method of serial independence (Abouheif, 1999), which is equivalent to Moran's $I$ when computed with a specific matrix of phylogenetic weights based on branch lengths and trait distance between tips in the phylogeny (Pavoine *et al.*, 2008). Moran’s $I$ and its significance were estimated with 1000 permutations of the dataset using the function abouheif.moran from the package adephylo (Jombart *et al.*, 2010). For continuous variables, we estimated Pagel’s $\lambda$ (Pagel, 1999) and its significance with 1000 simulations with the function phylosig from phytools (Revell, 2012). We chose Pagel’s $\lambda$ over Blomberg’s $K$ (Blomberg *et al.*, (2003)) because simulations demonstrate that Blomberg’s $K$ estimates can be highly inflated in both type I and II error when calculated using pseudo-chronograms rather than fully time-calibrated phylogenies, while Pagel’s $\lambda$ is strongly robust to branch-length biases (Molina-Venegas & Rodríguez, 2017).
**Appendix S6.** Phylolm implementation.

We performed phylogenetic multiple regression models with the function and package phylolm (Ho & Ané, 2014). We implemented the lambda phylogenetic model for the correction of the error term. The lambda parameter in this model is used to transform the error associated to the autocorrelation in the variance–covariance matrix assuming a Brownian motion model of evolution. We chose this model because it consistently had the lowest AIC value when compared to the other six methods available in phylolm. Lambda is useful for improving the fit of the phylogenetic regression, but the actual evolutionary process resulting in lambda is hard to interpret (Revell et al., 2008).

**References** (Appendix S2 – Appendix S6).


**Fig. S1.** Phylogeny produced with the R package V.PhyloMaker (See Appendix S4 for details).
**Fig. S2.** Phylogenetic signal and its significance with Moran’s *I* obtained with Abouheif’s method for categorical species traits in the dataset. Asterisks denote statistical significance based on 1000 permutations: *P*=0.001 (See Appendix S5 for details).
**Table S1.** Dataset used in this study (in file Table S1.xlsx). Abbreviations are as follow: MS, mating system; GF, growth form; PM, pollination mode; DM, dispersal mode; MaxDP, maximum distance between populations in km; MSS, mean sample size of individuals per population. When more than one publication is included per species, the first one reports the $F_{ST}$ value used in this study.

**Table S2.** Pearson Chi-squared test for correlation between categorical variables and Cramer’s V degree of association between variables. Significant P values are in bold † (next page).

† Refer to Appendix S3 for details. The strongest association was between mating system and life form; most trees are outcrossing species, while most mixed-mating species are non-woody plants. Pollination mode was significantly associated with growth form, as well as with region; wind pollinated plants are almost entirely trees from temperate regions, while vertebrate pollination is more common in non-woody tropical plants. Growth form and seed dispersal were also correlated; most gravity-dispersed plants are non-woody, most animal dispersed plants are trees, and shrubs are rarely wind dispersed. Mating system and seed dispersal were also correlated; most outcrossing plants have seeds dispersed by animals. Growth form and region were also significantly associated; most tropical and subtropical plants are trees, while most non-woody plants are from temperate regions. Lastly, seed dispersal and region significantly correlated; wind dispersal is more common in the temperate zones, while animal dispersal is more common in the tropics.
<table>
<thead>
<tr>
<th>Tested correlations</th>
<th>Chi²</th>
<th>DF</th>
<th>P value</th>
<th>Cramer's V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed dispersal – Growth form</td>
<td>51.8</td>
<td>4</td>
<td>&lt;0.001</td>
<td>0.28</td>
</tr>
<tr>
<td>Seed dispersal – Mating system</td>
<td>8.7</td>
<td>2</td>
<td>0.01</td>
<td>0.16</td>
</tr>
<tr>
<td>Seed dispersal – Pollination mode</td>
<td>11.8</td>
<td>6</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>Seed dispersal – Region</td>
<td>14.5</td>
<td>4</td>
<td>0.006</td>
<td>0.15</td>
</tr>
<tr>
<td>Growth form – Mating system</td>
<td>30.4</td>
<td>2</td>
<td>&lt;0.001</td>
<td>0.30</td>
</tr>
<tr>
<td>Growth form – Pollination mode</td>
<td>41.1</td>
<td>6</td>
<td>&lt;0.001</td>
<td>0.25</td>
</tr>
<tr>
<td>Growth form – Region</td>
<td>15.8</td>
<td>4</td>
<td>0.003</td>
<td>0.15</td>
</tr>
<tr>
<td>Mating system – Pollination mode</td>
<td>7.2</td>
<td>3</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>Mating system – Region</td>
<td>1.3</td>
<td>2</td>
<td>0.52</td>
<td>0.06</td>
</tr>
<tr>
<td>Pollination mode – Region</td>
<td>54.1</td>
<td>6</td>
<td>&lt;0.001</td>
<td>0.28</td>
</tr>
</tbody>
</table>
**Table S3.** Estimates of the generalized variance inflation factor (GVIF), and its adjusted value accounting for the degrees of freedom (GVIF^{1/(2*Df)}) for each variable in tested models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>GVIF</th>
<th>Df</th>
<th>GVIF^{1/(2*Df)}</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(distance †)</td>
<td>1.28</td>
<td>1</td>
<td>1.13</td>
</tr>
<tr>
<td>ln(MSS ‡)</td>
<td>1.50</td>
<td>1</td>
<td>1.22</td>
</tr>
<tr>
<td>Marker</td>
<td>1.71</td>
<td>4</td>
<td>1.07</td>
</tr>
<tr>
<td>Mating system</td>
<td>1.13</td>
<td>1</td>
<td>1.06</td>
</tr>
<tr>
<td>Growth form</td>
<td>1.79</td>
<td>2</td>
<td>1.16</td>
</tr>
<tr>
<td>Pollination mode</td>
<td>1.69</td>
<td>3</td>
<td>1.09</td>
</tr>
<tr>
<td>Seed dispersal mode</td>
<td>1.30</td>
<td>2</td>
<td>1.07</td>
</tr>
<tr>
<td>Region</td>
<td>1.56</td>
<td>2</td>
<td>1.12</td>
</tr>
</tbody>
</table>

† distance: maximum distance between populations in km.

‡ MSS: mean sample size of individuals per population.
Table S4. Results from phylogenetic ANOVA on each categorical variable (predictor) and $F_{ST}$ as the response variable. P values are based on 1000 simulations. Significant P values are in bold.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mating system</td>
<td>0.5</td>
<td>0.5</td>
<td>24.38</td>
<td>0.002</td>
</tr>
<tr>
<td>Residuals</td>
<td>6.83</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth form</td>
<td>0.95</td>
<td>0.48</td>
<td>25.06</td>
<td>0.014</td>
</tr>
<tr>
<td>Residuals</td>
<td>6.37</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pollination mode</td>
<td>1.17</td>
<td>0.39</td>
<td>21.03</td>
<td>0.025</td>
</tr>
<tr>
<td>Residuals</td>
<td>6.16</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispersal mode</td>
<td>0.2</td>
<td>0.1</td>
<td>4.77</td>
<td>0.21</td>
</tr>
<tr>
<td>Residuals</td>
<td>7.12</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>0.2</td>
<td>0.1</td>
<td>4.78</td>
<td>0.24</td>
</tr>
<tr>
<td>Residuals</td>
<td>7.12</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S5. Pairwise post-hoc tests between groups within each categorical variable, estimated after performing the phylogenetic ANOVA. P value corrections were done with the Holm-Bonferroni method. Significant P values are in bold (next page).
<table>
<thead>
<tr>
<th>Pairwise comparisons</th>
<th>Pairwise T values</th>
<th>Pairwise P values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mating system</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed-mating – Outcrossing</td>
<td>4.93</td>
<td><strong>0.002</strong></td>
</tr>
<tr>
<td><strong>Growth form</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tree – Non-woody</td>
<td>6.9</td>
<td><strong>0.045</strong></td>
</tr>
<tr>
<td>Tree – Shrub</td>
<td>3.71</td>
<td>0.136</td>
</tr>
<tr>
<td>Non-woody – Shrub</td>
<td>1.46</td>
<td>0.497</td>
</tr>
<tr>
<td><strong>Pollination mode</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small insects – Large insects</td>
<td>5.01</td>
<td><strong>0.006</strong></td>
</tr>
<tr>
<td>Small insects – Vertebrates</td>
<td>1.7</td>
<td>1</td>
</tr>
<tr>
<td>Small insects – Wind</td>
<td>7.18</td>
<td>0.09</td>
</tr>
<tr>
<td>Large insects – Vertebrates</td>
<td>2.54</td>
<td>0.68</td>
</tr>
<tr>
<td>Large insects – Wind</td>
<td>1.08</td>
<td>1</td>
</tr>
<tr>
<td>Vertebrates – Wind</td>
<td>3.8</td>
<td>0.68</td>
</tr>
<tr>
<td><strong>Seed dispersal mode</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gravity – Wind</td>
<td>2.93</td>
<td>0.24</td>
</tr>
<tr>
<td>Gravity – Biotic</td>
<td>2.54</td>
<td>0.35</td>
</tr>
<tr>
<td>Wind – Biotic</td>
<td>0.62</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperate – Subtropical</td>
<td>2.41</td>
<td>0.24</td>
</tr>
<tr>
<td>Temperate – Tropical</td>
<td>2.77</td>
<td>0.35</td>
</tr>
<tr>
<td>Subtropical – Tropical</td>
<td>0.01</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Table S6. Details of model 7 including variables in the null model. Variables in bold indicate the reference level for each categorical factor; the intercept of all other levels is compared to the intercept of this reference. N indicates the sample size of each group without phylogenetic correction. The $R^2$ relates to the importance of each factor to the explained variance in $F_{ST}$ after accounting for the other variables in the model. Significant P values are in bold.

<table>
<thead>
<tr>
<th>Variable †</th>
<th>N</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>T value</th>
<th>P value</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>0.59</td>
<td>0.04</td>
<td>14.1</td>
<td>2.2E-16</td>
<td></td>
</tr>
<tr>
<td>Marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td><strong>SSR</strong></td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFLP</td>
<td>60</td>
<td>0.06</td>
<td>0.02</td>
<td>2.8</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Allozymes</td>
<td>114</td>
<td>0.01</td>
<td>0.02</td>
<td>0.1</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>ISSR</td>
<td>7</td>
<td>0.03</td>
<td>0.04</td>
<td>0.6</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>RAPD</td>
<td>16</td>
<td>0.08</td>
<td>0.03</td>
<td>2.7</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>ln(distance)</td>
<td>337</td>
<td>0.01</td>
<td>0.002</td>
<td>5.5</td>
<td>&lt;0.001</td>
<td>0.08</td>
</tr>
<tr>
<td>ln(mean sample size)</td>
<td>337</td>
<td>−0.03</td>
<td>0.007</td>
<td>−3.7</td>
<td>&lt;0.001</td>
<td>0.04</td>
</tr>
</tbody>
</table>

† SSR: simple sequence repeat (microsatellites), AFLP: amplified fragment length polymorphism, ISSR: inter-simple sequence repeat, RAPD: random amplification of polymorphic DNA. Distance: maximum distance between populations. Mean sample size: mean sample size of individuals per population.
Chapter II: Flowering asynchrony contributes to genetic divergence in tropical plants

Running title: Flowering asynchrony drives differentiation

Diana Gamba, Department of Biology, University of Missouri at Saint Louis, One University Blvd., 223R Research Hall, St. Louis, MO 63121, USA; dlgtk5@mail.umsl.edu

Alexander Linan, Center for Conservation and Sustainable Development, Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, MO 63110, USA; alinan@mobot.org

Nathan Muchhala, Department of Biology, University of Missouri at Saint Louis, One University Blvd., 223R Research Hall, St. Louis, MO 63121, USA; muchhalan@umsl.edu

Author for correspondence: Diana Gamba; One University Boulevard, 223R Research Hall, Saint Louis, MO 63121, USA; dlgtk5@mail.umsl.edu; +1 (314) 702-0326.

In review: New Phytologist

Abstract: 200 words

Total word count (main text): 4932
Abstract

Speciation rates are frequently higher in tropical clades relative to temperate counterparts, yet the underlying mechanisms behind regional differences remain poorly understood. One compelling but relatively untested idea is the 'asynchrony of seasons hypothesis' (ASH). It posits that, while seasons are relatively synchronized over large areas in temperate regions, there can be seasonal asynchrony over short distances in tropical regions due to differences in the onset of rainfall between nearby sites. Climatic seasonal asynchrony leads to reproductive seasonal asynchrony, imposing a temporal barrier to gene flow and thus promoting population genetic divergence among subpopulations, which in turn may promote speciation. Here, we focused on understory angiosperms in two cloud forest sites in northwestern Ecuador that diverge in rainfall seasonality. We tested a central prediction of the ASH: that species with higher flowering asynchrony between sites will have genetically more divergent populations. We documented flowering phenology for nine species at both sites over one year and inferred population genetic parameters with a genome-wide genotyping approach. We found a strong positive cross-species association between flowering asynchrony and population differentiation. Our results suggest that seasonal asynchrony between sites can contribute significantly to population genetic divergence, and thus potentially to speciation, in tropical angiosperms.

Key words: Andes, angiosperms, cloud forest, flowering asynchrony, population genetic differentiation, 2b-RAD sequencing.
Introduction

Understanding the spatial and temporal processes that shape the patterns of angiosperm diversity is of central interest in biology (Fedorov 1966; Davies et al. 2004; Soltis et al. 2019). One prominent pattern exhibited by many clades is that of higher diversification rates in the tropics than in the temperate zones (Mittelbach et al. 2007; Brown 2014). Phylogenetic evidence from fossil and extant species suggest that this is due to higher speciation rates—rather than to lower extinction rates—in the tropics, which are predicted to coincide with higher population genetic divergence (reviewed in Mittelbach et al. 2007). However, the underlying mechanisms responsible for higher population genetic divergence and speciation in the tropics remain largely unknown. Several explanations suggest that dispersal, and thus gene flow, is more restricted in the tropics than in the temperate zones (Salisbury et al. 2012; Schluter and Pennell 2017). Limited gene flow between populations promotes population genetic divergence, resulting ultimately in reproductive isolation and allopatric speciation (Haffer 1997; Claramunt et al. 2012).

Several factors may contribute to gene flow being more restricted in the tropics. For example, the complex topography and environmental heterogeneity of the region can limit the movement of organisms and thus gene flow, resulting in isolated subpopulations (Wallace 1854; Benham and Witt 2016). Furthermore, the low temperature seasonality in the tropics can result in subpopulations that evolve relatively narrow niches that adapt them to local conditions. If local conditions vary widely over short distances, local adaptation would further restrict
gene flow among subpopulations, and increase isolation (Janzen 1967; Ghalambor et al. 2006). Moreover, local adaptation of subpopulations might result in mismatched timing of their reproductive cycles over short distances, disrupting gene flow between subpopulations. This temporal disruption to gene flow is central to the ‘asynchrony of seasons’ hypothesis (Martin et al. 2009), a compelling but relatively untested explanation for higher rates of population genetic divergence and speciation in the tropics.

The ‘asynchrony of seasons hypothesis’ (ASH) is based on the observation that seasons in temperate zones are determined by relatively constant temperature regimes over large geographical distances, while seasons in the tropics are determined primarily by rainfall patterns, which can vary greatly over short distances. This results in a geographical mosaic of climatic seasonality in the tropics, i.e. high climatic asynchrony between nearby sites. Because organisms usually time their reproductive cycles to seasons, such climatic asynchrony could result in reproductive asynchrony, which in turn would disrupt gene flow among subpopulations and promote population genetic divergence and speciation. Thus, a central prediction of the ASH is that tropical species with higher reproductive asynchrony will have more highly genetically divergent populations. One study found support for this prediction among new world birds: seasonal asynchrony was a strong predictor of genetic distance across intraspecific pairs of individuals, after accounting for potential geographic barriers to dispersal (Quintero et al. 2014). While compelling, this study only examined seasonal asynchrony across sites, and did not document whether this in fact
corresponds with reproductive asynchrony, a task which would be somewhat daunting for birds. Reproductive cycles are relatively easy to document for angiosperms, on the other hand, by simply observing when plants are in flower. Thus, angiosperms represent a logical group for an additional test of the ASH which can more directly examine the association between reproductive asynchrony and genetic divergence.

The impact of differences in flowering time on gene flow has been evaluated among sympatric individuals of the same species (Taylor and Friesen 2017), but little is known about how differences across a species range can impact gene flow among subpopulations. A model of incipient sympatric speciation showed that asychronic flowering time among individuals quickly lead to reproductive isolation and speciation (Devaux and Lande 2008) because it results in assortative mating among individuals with overlapping flowering (also see (Hendry and Day 2005; Gaudinier and Blackman 2019)). In an allopatric scenario, flowering time should shift between sites with different seasonality as plants adapt to local conditions to maximize their reproductive success (Blackman 2017; Gaudinier and Blackman 2019). If shifts in flowering time can cause speciation in sympathy (Hendry and Day 2005), we expect they would be even more likely to cause speciation in allopatry, in line with the ASH.

Here, we examine the ASH for the first time, to our knowledge, in tropical angiosperms. We test the central prediction that species with higher reproductive asynchrony between sites should have greater population genetic divergence. We focus on two sites in northern Ecuador, located in the western slope of the
Andes. These sites are close enough to share many species, but differ in the onset of the rainy season, which we expected would promote divergent flowering time between subpopulations. For nine understory species, we documented flowering phenology at both study sites for one year. To infer population genetic divergence between sites, we used a genome-wide genotyping approach using single nucleotide polymorphisms. We then tested whether flowering asynchrony between sites explained differences in population genetic divergence across species.

**Materials and Methods**

**Study sites**

This study was performed in Golondrinas and Santa Lucía reserves, two cloud forests located in the northwestern slope of the Andean cordillera of Ecuador, in the provinces of Carchi and Pichincha, respectively (Fig. 1A). Sites are ~100 km apart from each other and range from 1500–2500 m in elevation. Rainfall seasonality was inferred from monthly precipitation data extracted from the WorldClim database at a projected resolution of 30 arcseconds (Hijmans et al. 2005). We delimited two polygons using the coordinates of our focal plants at each site (Golondrinas: 0.80–0.84 N, 78.07–78.15 W; Santa Lucía; 0.10–0.13 N, 78.59–78.64 W). Based on the area of these polygons we extracted mean monthly precipitation and calculated standard errors (Fig. 1B). The rainy season in Golondrinas extends from October to May, peaking in April, while the rainy season in Santa Lucía extends from December to May, peaking in March.
Moreover, Santa Lucía receives twice as much rainfall as Golondrinas each year. We expected that these differences in precipitation should affect the flowering phenology of some portion of our focal species, leading to asynchrony between sites.

**Study species**

To select our focal species, we began by compiling a list of species occurring at both sites using the Tropicos.org database of the Missouri Botanical Garden. Through fieldwork, we further narrowed this list to nine perennial understory angiosperms, based on sufficient abundance in both study sites for flowering phenology surveys and population genetic work. These included *Begonia tiliifolia* C. DC. (Begoniaceae), *Besleria solanoides* Kunth (Gesneriaceae), *Burmeistera multiflora* Zahlbr. (Campanulaceae), *Centropogon solanifolius* Benth. (Campanulaceae), *Drymonia tenuis* (Benth.) J.L. Clark (Gesneriaceae), *Fuchsia macrostigma* Benth. (Onagraceae), *Gasteranthus quitensis* Benth. (Gesneriaceae), *Kohleria affinis* (Fritsch) Roalson & Boggan (Gesneriaceae), and *Meriania tomentosa* (Cogn.) Wurdack (Melastomataceae).

Based on our observations in the field, all focal species have dichogamous hermaphrodite flowers with male parts developing before female parts, except for *B. tiliifolia*, which is monoecious with male flowers developing before female flowers. Dichogamy likely reduces self-fertilization for all species, although some of them produce multiple flowers at the same time, which might result in geitonogamy. Pollination of most species is achieved by hummingbirds,
while two species are bat pollinated, and one species is insect pollinated (Muchhala 2006; Weinstein and Graham 2017; Dellinger et al. 2019) (Table 1). Seed dispersal in focal species remains largely unknown, and we were unable to detect seed dispersers from field observations. Those species with berries and fleshy capsules (Table 1) are hypothesized to be animal dispersed (Kvist and Skog 1992; Loiselle and Blake 1993), while other types of capsules are hypothesized to be gravity dispersed (Gamba et al. 2017).

**Estimation of flowering phenology**

To assess phenological patterns, we marked 10–25 individuals per species located along trails in the reserves. We selected individuals that were at least 5 m apart from each other to limit spatial autocorrelation. We recorded the number of flowers during twice-per-month surveys over one year (July 2017 through June 2018; Table S1). For each species at each site, the date with the highest number of flowers was taken as the 100% flowering peak and used to calculate the percentage of flowers for the rest of survey dates (Table S2).

**Evaluation of flowering seasonality and asynchrony**

We evaluated flowering seasonality from the twice-per-month flowering percentages with a Fourier spectral analysis using the function ‘spec.pgram’ in the stats R package in RStudio V 1.2.5019 (R Core Team 2018). Such analysis decomposes the flowering time series into sinusoidal curves representing different periodicities (Platt and Denman 1975; Zalamea et al. 2011; Quintero et
al. 2014). For each species, we evaluated the fit of the flowering data to periodicities corresponding to one peak of flowering per year with a 12-month period between peaks (i.e., annual pattern), two peaks of flowering per year with a 6-month period (biannual), and three and four peaks of flowering per year, with 4- and 3-month periods, respectively (sub-annual patterns). To evaluate whether the fit of the selected pattern for each species was greater than would be expected by chance, we constructed a null distribution of flowering times for each species by randomly resampling the flowering data 10,000 times (as in Zalamea et al. 2011). All species exhibited statistically significant phenological seasonality, exceeding the 95% quantiles of the corresponding null distributions, and this pattern was consistent between sites.

After establishing the periodicity of phenological patterns at each site, we then performed Fourier cospectral analyses to estimate the magnitude of intraspecific flowering asynchrony between sites. This analysis gives a value in radians corresponding to an angle positioning that represents the lag between flowering peaks between sites (Quintero et al. 2014). We transformed this value to degrees and subsequently to percent asynchrony, where 0° corresponds 0% asynchrony (both peaks occurring at the same time), and 180° corresponds to 100% asynchrony (the peak of flowering in one site coinciding with the valley of flowering on the other site). We also used similar Fourier analyses as outlined above to estimate rainfall seasonality and percent asynchrony between study sites, using the WorldClim data described previously (Table S3).
Genomic library preparation and sequencing

We began molecular work by extracting whole genomic DNA from silica-dried leaf tissue from 20 individuals per species from each study site. We followed the CTAB protocol (Doyle and Doyle 1987), modified slightly by incorporating additional ethanol washes of the DNA pellet. We quantified DNA with a Qubit 2.0 Fluorometer (Invitrogen, Thermo Fisher Scientific), using the manufacturer's protocol. For each of our samples with sufficient DNA, we obtained single nucleotide polymorphisms (SNPs) to use as genetic markers for population divergence inferences with the restriction site-associated DNA sequencing technique called 2b-RAD (Wang et al. 2012). We constructed 2b-RAD libraries for each individual following the protocol of (Wang et al. 2012). 500 ng of total genomic DNA were digested with a type IIb endonuclease, BcgI (New England Biolabs), which cuts DNA on both sides of a recognition site to obtain uniform 36-bp fragments scattered across the genome. Oligonucleotide Illumina sequences were ligated to these fragments with 12 double-stranded barcoded adapters, one per each column of a 96-sample plate. In order to increase sequence coverage per locus, we utilized reduced representation barcoded adapters which reduce the total number of loci sequenced. Samples with different barcoded adapters were pooled into 8 groups of 12 samples. Following initial pooling, Illumina RAD PCR primers (1–8) were incorporated into the fragments of each pool via 14 cycles of PCR amplification. Amplified pools were then purified via gel electrophoresis, and fragments of 75bp were size-selected by excising target bands from the agarose gel. We then used a Min Elute Gel
extraction kit (Qiagen) to purify target bands. Purified samples were quantified and pooled into a single library in equimolar concentrations. We generated three libraries, which together included ~15 individuals per species per study site. Libraries were sequenced on Illumina HiSeq 2500 (Brigham Young University, UT) and HiSeq 4000 (Duke University, NC) machines, to generate single-end 50 bp reads.

**Building loci and genotyping individuals**

Reads were demultiplexed using a custom script (trim2bRAD) generated by the Matz lab at the University of Austin, TX (https://github.com/z0on/2bRAD_denovo). This script trims 2b-RAD fragments from barcodes to produce one fastq file per sample. The resulting files were quality filtered with FastQC (Babraham Bioinformatics) and the FASTX-toolkit (Gordon and Hannon 2010). We discarded low quality reads and obtained sequences that were 36 bp in length, with a minimum of 90% bases having a Phred quality score of at least 20 and an input quality offset of 33. We then used the Stacks v2.3e pipeline to genotype individuals and produce a catalog of loci for each species (Catchen et al. 2013). We ran Stacks using the default parameter settings for building loci, which we considered to be appropriate for the short size of the 2b-RAD fragments. These parameter settings included a maximum distance of 2 nucleotide differences allowed between reads, a minimum depth of coverage of 3 reads required to create a stack, and a maximum distance of 4 nucleotide differences allowed to align secondary reads.
to primary stacks. We also allowed one gap between stacks before merging into putative loci. We filtered loci with the program ‘populations’ on the same pipeline. We excluded loci that were genotyped in <40% of individuals in each population. To avoid effects of linkage disequilibrium in our analyses, we only used one random SNP per locus. To prevent potential low-frequency SNP miscalls, we discarded alleles that had a frequency <5% in any locus across all individuals. To avoid repetitive or paralogous loci, the maximum number of heterozygous individuals that may be present in any locus was set to 75%. Lastly, we used the program VCFtools v0.1.16 (Danecek et al. 2011) to identify individuals with >50% missing data relative to variant sites, which we removed from subsequent analyses.

**Inference of population genetic divergence**

We used the program GenoDive v3.0 (Meirmans and Van Tienderen 2004) to calculate genetic diversity statistics. We assessed population genetic divergence between study sites for each species using the pairwise fixation index, $F_{ST}$ (Wright 1965; Nei 1977), and the allelic differentiation statistic, Jost’s $D$ (Jost 2008; Jost et al. 2018). The statistical significance of diversity statistics was assessed using 1000 random permutations of the data, while standard deviations of diversity statistics were obtained by jackknifing over loci and 95% confidence intervals were obtained by bootstrapping over loci.

To further visualize genetic divergence, we inspected genetic clustering in focal species. We conducted assignment tests using the program STRUCTURE

We examined whether the data fit to $K = 1–4$ genetic clusters using 20 replicates per $K$ with 300,000 generations used as burn in followed by 500,000 generations to achieve convergence. Optimal $K$ values were inferred using the Evanno method (Evanno et al. 2005). Results were summarized with the program CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007).

**Testing the relationship between flowering asynchrony and population genetic divergence**

We used linear regressions to test if flowering asynchrony predicts population genetic divergence between sites across our focal species. To evaluate whether this relationship was robust to different measures of population genetic divergence, we repeated analyses with either pairwise $F_{ST}$ or Jost’s $D$ as response variables. We also performed phylogenetic regressions to account for potential autocorrelation in the data due to evolutionary relationships. To this end, we extracted a species-level phylogeny containing the focal taxa (Fig. S1) from an angiosperm mega-tree (Smith and Brown 2018) in the R package V.PhyloMaker (Jin and Qian 2019). Branch lengths were inferred using the branch length adjuster algorithm in the same package (Qian and Jin 2016). We performed linear regressions of population genetic divergence on flowering asynchrony with the R function ‘lm’, and phylogenetic regressions with function ‘phylolm’ from the phylolm R package (Ho and Ané 2014). To assess the fit of
our phylogenetic models to the data, the likelihood of parameters was calculated under a Brownian motion model of trait evolution (Symonds and Blomberg 2014). We compared the fit of the models using AIC scores (Akaike 1974; Burnham and Anderson 2004). To provide a more thorough evaluation of model fit, we also measured phylogenetic signal in the error term of each linear regression (as in Revell 2010) using Pagel’s $\lambda$ (Pagel 1999).

Results

Flowering seasonality and asynchrony

Fourier spectral analyses found an annual flowering periodicity to be the most common pattern in both study sites (Table 2). Most species flowered earlier in Golondrinas than in Santa Lucía (Fig. 2), as might be expected given the earlier onset of the rainy season in Golondrinas. The only species with patterns different from annual were *B. multiflora*, in which the production of flowers was steady with three peaks in the year, and *B. solanoides*, in which we recorded two clear peaks in the year separated by periods of 0% production.

Among the 7 annually-flowering species, there was variation in the extent to which they were also flowering in other parts of the year, which can be summarized as three general patterns: 1) constant flower production at >30% of highest flower count throughout the year (*B. tiliifolia*), 2) constant flower production at >10% of highest flower count throughout the year (*D. tenuis* and *G. quitensis*), and 3) discrete flower production, with periods of 0% production lasting 1–4 months (*C. solanifolius, F. macrostigma, K. affinis, M. tomentosa*).
Fourier cospectral analyses identified a range of flowering asynchrony values across the species, from 2.7–87.0% (Table 2, mean = 26.1 ± 29.5 SD).

*Burmeistera multiflora* and *F. macrostigma* showed the lowest asynchrony, while *K. affinis*, *M. tomentosa* and *C. solanifolius* showed the highest. The remaining species presented asynchronies between 9.3–16.8%. Similar Fourier cospectral analyses of rainfall patterns from WorldClim data across the two study sites detected a significant annual pattern in precipitation for both sites and a precipitation asynchrony of 19% between them (Table S3).

**Filtered genetic datasets**

After SNP calling and quality control using different filtering procedures, we obtained a mean of 2,174,885 SNP loci per species (± 834,061 SD; range: 1,071,520–3,370,979), with a mean coverage ranging from 12.6–22.7 read depth per loci across species (Table S4). After removing individuals with >50% missing data, final sample sizes of individuals per species per study site ranged from 7–12 (mean = 9 ± 1.5 SD), and the number of variant loci ranged from 1,082–7,624 (mean = 3,840 ± 2,199 SD) across species, with missing data across species ranging from 35–40% (mean = 38 ± 2.3 SD) (Table S5 and S6).

Gene diversity was similar across species, with He (expected heterozygosity) within sites ranging from 0.19–0.26 (mean = 0.24 ± 0.02). Additionally, all species showed statistically significant levels of inbreeding, as indicated by significant *Gis* values, when these values are pooled across sites for each species (mean = 0.51 ± 0.2 SD; Table S5) as well as when they are
analyzed separately by site for each species (mean = 0.51 ± 0.2 SD, Table S6).

Population genetic divergence

Population genetic divergence between sites was significant for all species (Fig. 3). Pairwise F<sub>ST</sub> values ranged from 0.09–0.30 (mean = 0.16 ± 0.09 SD), and Jost’s D values from 0.03–0.13 (mean = 0.06 ± 0.04 SD). Further inspection of genetic divergence based on clustering STRUCTURE analyses showed that K = 2 was the most common supported number of clusters within species for all of the species, with the exception of *D. tenuis* for which K = 3 was the most likely number (Fig. 3 and Fig. S2). These genetic clusters most frequently followed geography, with one genetic cluster assigned to each of the two study sites. For *B. tiliifolia* and *B. multiflora*, there was one admixed individual identified at each site based on STRUCTURE Q values, while *F. macrostigma* and *M. tomentosa* showed no evidence of admixture between clusters. *Centropogon solanifolius*, *G. quitensis* and *K. affinis* exhibited a directional pattern of admixture, with varying amounts of alleles from Santa Lucía in Golondrinas but not vice-versa. For *D. tenuis*, Santa Lucía was almost homogeneous in cluster assignment except for one admixed individual, while all three genetic clusters were present in Golondrinas. Lastly, *B. solanoides* was composed of two genetic clusters present in both study sites (Fig. 3). This unexpected result might indicate that *B. solanoides* is composed of two cryptic species which are present at both sites.
Flowering asynchrony and genetic divergence

We performed linear and phylogenetic regressions to evaluate the relationship across species between flowering asynchrony and genetic divergence (in terms of pairwise $F_{ST}$ and Jost’s $D$ values). Because genetic clustering results suggest that individuals of $B. solanoides$ may potentially represent two species, we repeated regressions either including or excluding $B. solanoides$ (Table 3).

Results demonstrate that flowering asynchrony is a significant predictor of pairwise $F_{ST}$ ($F(1, 7) = 39.1$, adjusted-$R^2 = 0.83$, $p = 0.0004$) and Jost’s $D$ ($F(1, 7) = 33.5$, adjusted-$R^2 = 0.80$, $p = 0.0007$) (Table 3). The same analyses without $B. solanoides$ yielded similar positive associations between flowering asynchrony and pairwise $F_{ST}$ ($F(1, 6) = 36.3$, adjusted-$R^2 = 0.83$, $p = 0.0009$) and Jost’s $D$ ($F(1, 6) = 29.2$, adjusted-$R^2 = 0.80$, $p = 0.002$) (Table 3 and Fig. 4A, B).

Phylogenetic regressions did not improve model fit and produced identical results. Similarly, Pagel’s $\lambda$ tests of phylogenetic signal on the error term of all linear regressions were non-significant (Table 3), consistent with a lack of phylogenetic autocorrelation in the data.

Discussion

Our results reveal a robust positive association between flowering asynchrony and population genetic divergence across our nine focal species of Andean angiosperms (Table 3, Fig. 4). Those species with greater shifts in flowering patterns across our two study sites had greater levels of genetic
divergence between their two subpopulations. Given that precipitation patterns were significantly different across these sites, these results support the idea that spatial variation in climatic seasonality may drive increased levels of genetic divergence, which in turn might be an important mechanism for the origin of new species of angiosperms.

Our study design controlled for many other factors that might impact population genetic divergence, increasing the probability that the association we found is in fact due directly to flowering asynchrony rather than a confounding variable. For instance, by choosing the same two study sites for all species, geographic distance could not influence differences in $F_{ST}$ values across species. Similarly, study species are likely all exposed to the same geographic barriers. They all occur in the understory of cloud forests on the same slope of the Andes, and both sites belong to the southern end of the Choco Andean corridor (Mordecai et al. 2009) and are presumably well-connected by a continuous corridor of forests due to the presence of the Cotacachi-Cayapas national park between them. Finally, differences in inbreeding levels do not seem to underlie the differences in population genetic divergence. Inbreeding can affect population genetic structure (Duminil et al. 2007), however we do not find such association in our dataset: the inbreeding coefficient ($G_{IS}$ in Table S5) does not predict $F_{ST}$ ($F(1, 7) = 0.19$, adjusted $R^2 = -0.11$, $p = 0.7$).

We note that six of our study species presented relatively high inbreeding coefficients (i.e., $F_{IS}$ values were $> 0.5$ in $B. tiliifolia$, $B. solanoides$, $C. solanaoides$, $D. tenuis$, $G. quitensis$, and $K. affinis$), which is generally associated
with selfing. This is striking given that five of the species are largely visited by
hummingbirds (Weinstein and Graham 2017), while only one (B. tiliifolia) is
presumably insect pollinated (pers. obs.). Studies of the pollination biology of B.
tiliifolia are lacking, but it is possible that this monoecious herb is self-compatible,
as are many other Begonia (Agren and Schemske 1993; Matolweni et al. 2000;
Waytt & Sazima 2011). Self-compatibility is also common among other species
related to our focal taxa, as has been shown in Besleria (Martin-Gajardo 1999),
Drymonia (Steiner 1985), and other neotropical species (Schatz 1990). However,
spontaneous self-pollination is unlikely due to monoecy in B. tiliifolia, and
protandry in the hummingbird pollinated species. It is likely that pollinators
promote geitonogamy and thus increase inbreeding within subpopulations,
especially for hummingbird pollinated species that produce multiple flowers
simultaneously (i.e., G. quitensis and K. affinis).

We also note that species with lower genetic divergence (e.g., B.
multiflora) showed a more constant production of flowers throughout the year,
while species with greater genetic divergence showed markedly interrupted
production of flowers, with periods of 0% production ranging from 1–4 months.
Specifically, in M. tomentosa zero-flowering periods were long and extended (~ 4
months, one valley per year, figure 2), while in C. solanifolius zero-flowering
periods were short and intermittent (~ 2 months or shorter, multiple valleys per
year, Fig. 2). Thus, some zero-flowering periods at a given site may be an
important contributor to cutting off gene flow between nearby sites.

The mode of gene dispersal between subpopulations could also affect
the importance of flowering asynchrony in population genetic divergence. If gene flow between nearby sites is mainly achieved via pollen dispersal, flowering asynchrony would be the primary mechanism for genetic divergence. However, if gene flow is also achieved via seed dispersal, flowering asynchrony might not be as important to promote genetic divergence. In the presence of seed dispersal, the association between flowering asynchrony and genetic divergence will largely depend on the fate of migrant seeds in a new site in combination with the underlying drivers of flowering time. If flowering time is a phenotypically plastic response to rainfall patterns (Levin 2009), adult migrants would flower at the same time as the local population, while if it is an evolved response to some other cue (Hall and Willis 2006), these migrants may remain out-of-synch with conspecifics in the new site. Common garden experiments (as in Fudickar et al. 2016), or reciprocal transplants (as in Hall and Willis 2006), would help to evaluate the role of phenotypic plasticity and environmental cues in determining flowering phenology.

If migrants remain out of synch with conspecifics in the new site, flowering asynchrony could arise within a site and prevent gene flow between sympatric individuals. Asynchrony in flowering time among sympatric individuals is often termed allochrony (Gaudinier and Blackman 2019) and has been proposed as a possible mechanism for reproductive isolation in sympatry (Hendry and Day 2005; Taylor and Friesen 2017). A model of speciation in sympatry proposes that reproductive isolation can quickly evolve within small populations exhibiting long population-level periods of flowering, but short
individual-level periods of flowering, as this will cluster individuals genetically according to their flowering time (Devaux and Lande 2008). However, whether or how frequently this occurs in nature remains unclear. Allochrony has also been proposed as a mechanism that strengthens boundaries between incipient species when ranges rejoin in secondary contact, with prominent empirical examples in nature (Briscoe Runquist et al. 2014; Hipperson et al. 2016; Spriggs et al. 2019). This evidence suggests that flowering asynchrony likely evolves in allopatri, in line with the ‘asynchrony of seasons hypothesis’ (ASH), and its persistence after secondary contact helps to reduce gene flow and maintain species boundaries.

Among our focal species, B. solanoides was the only taxon for which we detected two genetic clusters that did not correspond to the two study sites, but rather both occurred at both study sites. Interestingly, we note that one genetic cluster (in blue in figure 3) corresponds to early bloomers in both study sites, while the other (in orange) is composed of late bloomers in both study sites. Thus, these clusters might represent cryptic species separated by flowering time. This pattern suggests empirical support for the scenario discussed above, where shifts in flowering time evolved in allopatri (as per the ASH) and now maintain boundaries of these hypothetical cryptic species after one or both expanded their range into sympatry. Remarkably, the pairwise $F_{ST}$ between genetic clusters was 0.23 ($p<0.001$), greater than the pairwise $F_{ST}$ between sites (0.09, Table 3). A thorough taxonomic and demographic study including individuals across B. solanoides' range would help to evaluate this hypothesized scenario of cryptic
speciation after secondary contact driven by flowering asynchrony.

One important caveat to our study is that the relationship between flowering asynchrony and population genetic divergence between sites only establishes a correlation, not a causation. Greater asynchrony may drive increased genetic divergence, as we have argued above. However, it could also be that subpopulations in each study site first became genetically differentiated due to other factors, and this divergence then led to differences in flowering phenologies. In such a case, flowering asynchrony would further strengthen the existing genetic divergence between subpopulations. Nonetheless, whether shifts in flowering time cause or strengthen genetic divergence, our main finding supports flowering asynchrony as an important mechanism that limits gene flow between subpopulations.

Our study provides the first test to date of the ‘asynchrony of seasons hypothesis’ (Martin et al. 2009) in flowering plants. We found evidence for a central prediction of the ASH, namely that reproductive asynchrony between tropical sites with different seasonality is associated with increased population genetic divergence. Thus, reproductive asynchrony may accelerate rates of population differentiation, and ultimately speciation in tropical plants. Before our study, ASH had only been tested in birds (Moore et al. 2005; Quintero et al. 2014). We thus encourage more phenological studies, in flowering plants and other organisms, to broadly document patterns of reproductive asynchrony and how these relate to ‘isolation by time’ in allopatry. Future work should also examine whether reproductive asynchrony is more prevalent in tropical than in
temperate systems, as predicted by their increased seasonal asynchrony. If so, flowering asynchrony could represent a key explanation for the latitudinal diversity gradient observed in flowering plants.

Acknowledgements

Ben Weinstein and Holger Beck were instrumental for locating plants in Santa Lucía and provided useful preliminary phenological data for a number of species. Thanks to Nora Oleas and Paola Peña for help with the research permit in Ecuador (MAE-DNB-CM-2015-017). Robert Ricklefs, Christine Edwards, and Carmen Ulloa provided great advice in this study. Thanks to field assistants Hugo Quintanchala, Paola Peña, Nelly Muñoz, Justin Zweck, An Nguyen, and Carlos Imery, and to families at Santa Lucía cloud forest reserve and at Bosque Protector Golondrinas for their hospitality. Joel Swift offered useful guidance for 2b-RAD, and Isabel Loza for Fourier analyses. Discussions with members of the Muchhala lab at the University of Missouri-Saint Louis (UMSL) greatly improved a previous version of this manuscript. This study was funded with graduate student research grants from the Whitney R. Harris World Ecology Center at the University of Missouri at St. Louis, the Botanical Society of America, the Society of Systematic Biologists, and the American Philosophical Society to DG, and a grant from the Office of Research Administration at the University of Missouri at St. Louis to NM.
References


Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters


Steiner, K. E. 1985. The role of nectar and oil in the pollination of Drymonia serrulata (Gesneriaceae) by Epicharis bees (Anthophoridae) in Panama. Biotropica 17:217–229.


**Data accessibility statement**

Should the manuscript be accepted, the data and R scripts supporting the results will be archived in Dryad and their DOI will be included at the end of this article.

**Author contributions**

DG and NM planned and designed the research. DG collected and analyzed the data. AL performed STRUCTURE analyses. DG wrote the first draft of the manuscript. DG, AL, and NM contributed equally to substantial revisions of the manuscript.
**Fig. 1 (a)** Location of study sites in northwestern Ecuador, South America, with map color representing elevation over sea level in m. The grey circle is Bosque Protector Golondrinas and the black circle is Santa Lucía Cloud Forest Reserve. **(b)** Rainfall seasonality at study sites: the y-axis is the amount of monthly rainfall in mm. Boxplots show the distribution of rainfall data across the geographic extent of each reserve; black circles are monthly means, horizontal grey lines are medians, and the boxes’ lower and upper limits are 25th and 75th percentiles. Elevation and monthly rainfall data come from WorldClim raster layers at a projected resolution of 1 km².
Fig. 2 Flowering phenology of the nine studied species recorded for one year (July 2017 – June 2018). Flowering data is depicted in the y-axis as a monthly percent of peak flowering in the year. Grey lines correspond to flowering in Golondrinas, and black lines in Santa Lucía.
**Fig. 3** Identified genetic clusters and Bayesian admixture proportions depicted for individual plants of each species. For most species $K = 2$ was the best $K$-fit to the data, except for *D. tenuis* which best $K = 3$. The black vertical bar on each structure plot separates individuals from Santa Lucía to the left and Golondrinas to the right (clusters between species are independent). Measures of genetic divergence between sites are indicated with pairwise $F_{ST}$ values (fixation index) and Jost’s $D$ values (allelic differentiation). All statistics were significant ($p<0.005$) based on 1000 permutations.
**Fig. 4** The positive and significant ($p < 0.005$) association between flowering asynchrony and population genetic divergence across eight species of tropical angiosperms (excluding *B. solanoides*): **A** with pairwise $F_{ST}$ in the y-axis, and **B** with Jost's $D$ in the y-axis. The blue line represents the prediction based on linear models with associated error in grey shading.
### Table 1: Characteristics of studied species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Growth form</th>
<th>Pollinators</th>
<th>Fruit type</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Begonia tiliifolia</em></td>
<td>Herbaceous</td>
<td>Insects (pers. obs.)</td>
<td>Indehiscent capsule</td>
</tr>
<tr>
<td><em>Besleria solanoides</em></td>
<td>Shrub</td>
<td>Hummingbirds (1)</td>
<td>Berry</td>
</tr>
<tr>
<td><em>Burmeistera multiflora</em></td>
<td>Herbaceous</td>
<td>Bats (2)</td>
<td>Inflated berry</td>
</tr>
<tr>
<td><em>Centropogon solanifolius</em></td>
<td>Herbaceous</td>
<td>Hummingbirds (1)</td>
<td>Indehiscent capsule</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>Sub-shrub</td>
<td>Hummingbirds (1)</td>
<td>Berry</td>
</tr>
<tr>
<td><em>Fuchsia macrostigma</em></td>
<td>Herbaceous</td>
<td>Hummingbirds (1)</td>
<td>Indehiscent capsule</td>
</tr>
<tr>
<td><em>Gasteranthus quitensis</em></td>
<td>Sub-shrub</td>
<td>Hummingbirds (1)</td>
<td>Fleshy capsule</td>
</tr>
<tr>
<td><em>Kohleria affinis</em></td>
<td>Epiphyte</td>
<td>Hummingbirds (1)</td>
<td>Fleshy capsule</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>Shrub</td>
<td>Bats and hummingbirds (3)</td>
<td>Indehiscent capsule</td>
</tr>
</tbody>
</table>

(1) Weinstein and Graham 2017

(2) Muchhala 2006

(3) Dellinger et al. 2019
Table 2: Flowering seasonality and asynchrony of studied species. A significance test of Fourier spectral analyses indicated that periodicity (i.e. seasonality) was significant for all studied species and consistent between sites ($p < 0.05$). A Fourier cospectral analysis was used to quantify flowering asynchrony (% async) between sites.

<table>
<thead>
<tr>
<th>Species</th>
<th>Seasonality (sample size)</th>
<th>% async</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Begonia tiliifolia</em></td>
<td>annual (25)</td>
<td>10.5</td>
</tr>
<tr>
<td><em>Besleria solanoides</em></td>
<td>bi-annual (25)</td>
<td>11.1</td>
</tr>
<tr>
<td><em>Burmeistera multiflora</em></td>
<td>sub-annual (25)</td>
<td>2.7</td>
</tr>
<tr>
<td><em>Centropogon solanifolius</em></td>
<td>annual (25)</td>
<td>87.0</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>annual (25)</td>
<td>16.8</td>
</tr>
<tr>
<td><em>Fuchsia macrostigma</em></td>
<td>annual (25)</td>
<td>6.4</td>
</tr>
<tr>
<td><em>Gasteranthus quitensis</em></td>
<td>annual (25)</td>
<td>9.3</td>
</tr>
<tr>
<td><em>Kohleria affinis</em></td>
<td>annual (25)</td>
<td>26.3</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>annual (20)</td>
<td>64.6</td>
</tr>
</tbody>
</table>
**Table 3** Results of linear regressions of population genetic divergence as predicted by flowering asynchrony for four tests. Tests (1) and (2) include *B. solanoides*. Tests (3) and (4) exclude *B. solanoides*. Significance of linear models is denoted in bold. Pagel’s λ measures phylogenetic signal in the error term of each linear model. Phylogenetic regressions produced identical results.

<table>
<thead>
<tr>
<th>Test</th>
<th>Response</th>
<th>Estimate</th>
<th>Std. E</th>
<th>t-value</th>
<th>p-value</th>
<th>adj-R²</th>
<th>AIC</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>intercept</td>
<td>0.092</td>
<td>0.016</td>
<td>5.7</td>
<td>0.0007</td>
<td>0.83</td>
<td>-30.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>F&lt;sub&gt;ST&lt;/sub&gt;</td>
<td>0.003</td>
<td>0.0004</td>
<td>6.3</td>
<td>0.0004</td>
<td>0.83</td>
<td>-30.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>intercept</td>
<td>0.031</td>
<td>0.008</td>
<td>4.0</td>
<td>0.005</td>
<td>0.80</td>
<td>-44.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Jost’s D</td>
<td>0.001</td>
<td>0.0002</td>
<td>5.7</td>
<td>0.0007</td>
<td>0.80</td>
<td>-44.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td>intercept</td>
<td>0.100</td>
<td>0.02</td>
<td>5.8</td>
<td>0.001</td>
<td>0.83</td>
<td>-27.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>F&lt;sub&gt;ST&lt;/sub&gt;</td>
<td>0.003</td>
<td>0.0004</td>
<td>6.0</td>
<td>0.0009</td>
<td>0.83</td>
<td>-27.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>intercept</td>
<td>0.033</td>
<td>0.008</td>
<td>3.9</td>
<td>0.008</td>
<td>0.80</td>
<td>-38.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Jost’s D</td>
<td>0.001</td>
<td>0.0002</td>
<td>5.4</td>
<td>0.002</td>
<td>0.80</td>
<td>-38.7</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Additional supporting information that will appear in the expanded online version of this article:

**Fig. S1** Phylogeny of studied species extracted with V.PhyloMaker.

**Fig. S2** Summary of Delta K results for each species.

**Table S1** Total flower count per species at each survey date (Table S1.xlsx).

**Table S2** Percent of flowering peak data per species (TableS2.xlsx).

**Table S3** Results from Fourier spectral and cospectral analyses (TableS3.xlsx).

**Table S4** Unfiltered catalog of loci for studied species.

**Table S5** Genetic diversity of studied species across loci.

**Table S6** Genetic diversity of studied species within sites.
**Fig. S1** Phylogeny of studied species extracted from a backbone tree in V.PhyloMaker.

- *Fuchsia macrostigma*
- *Merania tomentosa*
- *Begonia tiliifolia*
  - *Gasteranthus quitensis*
  - *Besleria solanoides*
  - *Kohleria affinis*
  - *Drymonia tenuis*
- *Burmeistera multiflora*
- *Centropogon solanifolius*
**Fig. S2** Summary of Delta K results for each species based on the Evanno et al. (2005) method. $K = 2$ was the best-fit to the data for most species, except for *D. tenuis* where the best $K = 3$.

**Table S1** Total flower count per species at each survey date (in file TableS1.TotalFlowers.xlsx).

**Table S2** Percent of flowering peak data per species per site used for Fourier spectral and cospectral analyses (in file TableS2.FlowersRdata.xlsx)

**Table S3** Results of Fourier spectral and cospectral analyses, significance of seasonality tests and estimated flowering asynchrony (in TableS3.SpectralCospectralResults.xlsx)
**Table S4** Unfiltered catalog of loci recovered with the STACKS v2.3e pipeline for non-model organisms. N is the number of individuals. Coverage refers to the mean depth of reads used to build loci.

<table>
<thead>
<tr>
<th>Species</th>
<th>total N genotyped</th>
<th>total loci genotyped</th>
<th><strong>Effective depth of coverage</strong> (reads/loci)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>stdev</td>
<td>min</td>
</tr>
<tr>
<td><em>Begonia tiliifolia</em></td>
<td>25</td>
<td>2196507</td>
<td>15.6</td>
</tr>
<tr>
<td><em>Besleria solanoides</em></td>
<td>26</td>
<td>1751170</td>
<td>12.6</td>
</tr>
<tr>
<td><em>Burmeistera multiflora</em></td>
<td>21</td>
<td>3370979</td>
<td>17.7</td>
</tr>
<tr>
<td><em>Centropogon solanifolius</em></td>
<td>23</td>
<td>1469286</td>
<td>20.8</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>26</td>
<td>3266205</td>
<td>16</td>
</tr>
<tr>
<td><em>Fuchsia macrostigma</em></td>
<td>21</td>
<td>2439480</td>
<td>16.1</td>
</tr>
<tr>
<td><em>Gasteranthus quitensis</em></td>
<td>23</td>
<td>1326095</td>
<td>16.2</td>
</tr>
<tr>
<td><em>Kohleria affinis</em></td>
<td>22</td>
<td>1071520</td>
<td>15.2</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>27</td>
<td>2682725</td>
<td>22.7</td>
</tr>
</tbody>
</table>
Table S5 Genetic diversity of studied species estimated across filtered loci. N var loci: number of variant loci, N total a: total number of alleles, %md: percent missing data, N a: mean number of alleles per locus, Ne a: mean effective number of alleles per locus, Ho: observed heterozygosity, Hs: mean expected heterozygosity across subpopulations, Ht: total expected heterozygosity over all subpopulations, Gis: inbreeding coefficient. Standard deviations of statistics (in parentheses) were obtained through jackknifing over loci and significance (*p < 0.005*) through 1000 permutations (denoted in bold).

<table>
<thead>
<tr>
<th>Species</th>
<th>N var loci</th>
<th>N total a</th>
<th>% md</th>
<th>N a</th>
<th>Ne a</th>
<th>Ho</th>
<th>Hs</th>
<th>Ht</th>
<th>Gis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begonia tiliifolia</td>
<td>4608</td>
<td>9035</td>
<td>40</td>
<td>1.96</td>
<td>1.29</td>
<td>0.09</td>
<td>0.24</td>
<td>0.26</td>
<td>0.62</td>
</tr>
<tr>
<td>Besleria solanoides</td>
<td>1082</td>
<td>2144</td>
<td>35</td>
<td>1.98</td>
<td>1.31</td>
<td>0.11</td>
<td>0.25</td>
<td>0.26</td>
<td>0.55</td>
</tr>
<tr>
<td>Burmeistera multiflora</td>
<td>7624</td>
<td>14175</td>
<td>36</td>
<td>1.97</td>
<td>1.25</td>
<td>0.08</td>
<td>0.22</td>
<td>0.26</td>
<td>0.62</td>
</tr>
<tr>
<td>Centropogon solanifolius</td>
<td>3182</td>
<td>6281</td>
<td>40</td>
<td>1.88</td>
<td>1.28</td>
<td>0.11</td>
<td>0.24</td>
<td>0.25</td>
<td>0.53</td>
</tr>
<tr>
<td>Drymonia tenuis</td>
<td>2389</td>
<td>4708</td>
<td>39</td>
<td>1.95</td>
<td>1.31</td>
<td>0.17</td>
<td>0.25</td>
<td>0.27</td>
<td>0.32</td>
</tr>
<tr>
<td>Fuchsia macrostigma</td>
<td>6634</td>
<td>12908</td>
<td>37</td>
<td>1.90</td>
<td>1.27</td>
<td>0.06</td>
<td>0.24</td>
<td>0.27</td>
<td>0.77</td>
</tr>
<tr>
<td>Gasteranthus quitensis</td>
<td>3251</td>
<td>6179</td>
<td>41</td>
<td>1.92</td>
<td>1.26</td>
<td>0.08</td>
<td>0.22</td>
<td>0.24</td>
<td>0.66</td>
</tr>
<tr>
<td>Kohleria affinis</td>
<td>1457</td>
<td>2716</td>
<td>36</td>
<td>1.95</td>
<td>1.29</td>
<td>0.16</td>
<td>0.23</td>
<td>0.28</td>
<td>0.33</td>
</tr>
<tr>
<td>Meriania tomentosa</td>
<td>4224</td>
<td>8236</td>
<td>36</td>
<td>1.96</td>
<td>1.29</td>
<td>0.09</td>
<td>0.24</td>
<td>0.26</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Table S6 Genetic diversity of studied species within sites estimated from filtered loci. S: Santa Lucía, G: Golondrinas, N: number of individuals in the final genetic dataset, Ne: effective number of individuals, P a: number of private alleles, % P a: proportion of private to total alleles, N a: mean number of alleles per locus, Ne a: mean effective number of alleles per locus, Ho: observed heterozygosity, Hs: mean expected heterozygosity within site, Gis: inbreeding coefficient. Significance (p < 0.005) was obtained through 1000 permutations and is denoted in bold.

<table>
<thead>
<tr>
<th>Species</th>
<th>Site</th>
<th>N</th>
<th>Ne</th>
<th>P a</th>
<th>% P a</th>
<th>N a</th>
<th>Ne a</th>
<th>Ho</th>
<th>Hs</th>
<th>Gis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begonia tiliifolia</td>
<td>S</td>
<td>11</td>
<td>7.4</td>
<td>1809</td>
<td>0.20</td>
<td>1.74</td>
<td>1.36</td>
<td>0.09</td>
<td>0.26</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>9</td>
<td>6.3</td>
<td>945</td>
<td>0.10</td>
<td>1.54</td>
<td>1.29</td>
<td>0.09</td>
<td>0.22</td>
<td>0.57</td>
</tr>
<tr>
<td>Besleria solanoides</td>
<td>S</td>
<td>12</td>
<td>8.2</td>
<td>295</td>
<td>0.14</td>
<td>1.73</td>
<td>1.32</td>
<td>0.11</td>
<td>0.23</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>8</td>
<td>5.6</td>
<td>269</td>
<td>0.13</td>
<td>1.70</td>
<td>1.37</td>
<td>0.12</td>
<td>0.27</td>
<td>0.56</td>
</tr>
<tr>
<td>Burmeistera multiflora</td>
<td>S</td>
<td>7</td>
<td>5.9</td>
<td>2563</td>
<td>0.18</td>
<td>1.62</td>
<td>1.34</td>
<td>0.20</td>
<td>0.24</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>7</td>
<td>4.9</td>
<td>1721</td>
<td>0.12</td>
<td>1.60</td>
<td>1.33</td>
<td>0.18</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td>Centropogon solanifolius</td>
<td>G</td>
<td>10</td>
<td>6.2</td>
<td>1090</td>
<td>0.17</td>
<td>1.60</td>
<td>1.30</td>
<td>0.08</td>
<td>0.23</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>10</td>
<td>5.7</td>
<td>535</td>
<td>0.11</td>
<td>1.59</td>
<td>1.34</td>
<td>0.12</td>
<td>0.26</td>
<td>0.54</td>
</tr>
<tr>
<td>Drymonia tenuis</td>
<td>G</td>
<td>7</td>
<td>7.0</td>
<td>668</td>
<td>0.14</td>
<td>1.62</td>
<td>1.30</td>
<td>0.10</td>
<td>0.22</td>
<td>0.54</td>
</tr>
<tr>
<td>Fuchsia macrostigma</td>
<td>G</td>
<td>9</td>
<td>5.9</td>
<td>1701</td>
<td>0.13</td>
<td>1.66</td>
<td>1.35</td>
<td>0.17</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>Gasteranthus quitensis</td>
<td>G</td>
<td>9</td>
<td>6.4</td>
<td>1127</td>
<td>0.18</td>
<td>1.59</td>
<td>1.32</td>
<td>0.06</td>
<td>0.24</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>9</td>
<td>6.4</td>
<td>388</td>
<td>0.14</td>
<td>1.50</td>
<td>1.25</td>
<td>0.07</td>
<td>0.19</td>
<td>0.63</td>
</tr>
<tr>
<td>Kohleria affinis</td>
<td>G</td>
<td>9</td>
<td>6.2</td>
<td>628</td>
<td>0.23</td>
<td>1.65</td>
<td>1.35</td>
<td>0.08</td>
<td>0.26</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>10</td>
<td>7.9</td>
<td>1507</td>
<td>0.18</td>
<td>1.65</td>
<td>1.35</td>
<td>0.16</td>
<td>0.24</td>
<td>0.33</td>
</tr>
<tr>
<td>Meriania tomentosa</td>
<td>G</td>
<td>10</td>
<td>7.0</td>
<td>1082</td>
<td>0.13</td>
<td>1.60</td>
<td>1.33</td>
<td>0.15</td>
<td>0.23</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Chapter III: Pollination by hummingbirds strongly decreases genetic structure between and within plant populations relative to pollination by insects

Running title: Effects of animal pollination on plant gene flow in tropical plants

Diana Gamba, Department of Biology, University of Missouri at Saint Louis
dlgtk5@mail.umsl.edu

Nathan Muchhalan, Department of Biology, University of Missouri at Saint Louis
muchhalan@umsl.edu

Author for correspondence: Diana Gamba; One University Boulevard, 223R Research Hall, Saint Louis, MO 63121, USA; dlgtk5@mail.umsl.edu; +1 (314) 702-0326.

Target journal: Annals of Botany

Abstract: 205

Total word count (main text): 4302
Abstract

Animal pollinators have a direct effect on plant gene flow because they carry the pollen grains. Pollinators with restricted mobility are predicted to limit gene flow within and among populations, while pollinators that fly longer distances likely promote genetic cohesion. Such predictions, however, remain surprisingly poorly tested. Here, we examined population genetic structure and fine-scale spatial genetic structure (SGS) in six perennial understory angiosperms in Andean cloud forests of northwestern Ecuador. Species belong to three families and within each family we selected one insect-pollinated species and one hummingbird-pollinated species. Based on differences in foraging behavior and flying ability, we tested the predictions that species pollinated by insects should have greater population genetic differentiation among study sites (as quantified with the $F_{ST}$ statistic), and stronger SGS (as quantified with the $S_P$ statistic), than species pollinated by hummingbirds. We confirmed putative pollinators through a literature review and fieldwork, and inferred population genetic parameters with a genome-wide genotyping approach. Generalized linear mixed-effects models showed that insect pollination is significantly associated with both greater population genetic differentiation and stronger SGS than hummingbird pollination. Our results clearly show for the first time that pollination by insects significantly restricts the spatial scale of intraspecific gene flow relative to pollination by hummingbirds.

**Key words:** 2b-RAD sequencing, Andean cloud forest understory, fine-scale spatial genetic structure, animal pollination, population genetic structure.
Introduction

Understanding how plant mutualists influence spatial patterns of genetic diversity is central to plant biology, especially in the present scenario of biodiversity decline due to human-accelerated environmental change (Hardy et al. 2006; Dick et al. 2008; Aguilar et al. 2008, 2019). Animal pollinators directly affect gene flow within and among flowering plant populations because they are the carriers of pollen grains (Loveless and Hamrick 1984; Hamrick et al. 1992). Previous broad-scale studies on patterns of genetic structure in plants have lumped together all animals, and compared them to wind, thus overlooking the effect of different animals on gene flow dynamics within and among plant population (Hamrick and Godt 1996; Duminil et al. 2007). Findings from such studies reveal that wind tends to homogenize plant gene pools, while animal pollination is associated with higher population genetic differentiation as well as stronger fine-scale spatial genetic structure (i.e., the non-random spatial distribution of closely related individuals) (Dick et al. 2008; Gelmi-Candusso et al. 2017). Thus, in general, animal pollination may significantly disrupt gene flow relative to wind pollination within and among populations. Such patterns, however, should vary depending on the pollen dispersal ability of the pollinator, which will depend on foraging behavior and pollen carry-over capacity (Levin 1979). Pollinators with large foraging areas can carry pollen long distances, potentially enhancing gene flow within and among plant populations. In contrast, pollinators with local foraging behavior potentially reduce pollen dispersal, likely disrupting gene flow within and among plant populations. This potential trend has
been suggested in seminal reviews (Levin 1981; Loveless and Hamrick 1984), and in some empirical studies (Linhart et al. 1987; Linhart and Grant 1996; Kramer et al. 2011; Amico et al. 2014). However, no study to date has formally tested the prediction that pollinators with limited mobility should lead to stronger patterns of isolation by distance across individuals, potentially increasing population genetic differentiation across subpopulations, relative to pollinators that fly longer distances.

Vertebrate pollinators, such as nectarivorous bats and birds, generally fly longer distances during foraging bouts than insects, likely enhancing pollen flow among distantly spaced individuals and subpopulations, even across fragmented habitats (Levin 1979; Machado et al. 1998; Sahley 2001; Southerton et al. 2004; Byrne et al. 2007; Dick et al. 2008; Hadley and Betts 2009; McCulloch et al. 2013; Breed et al. 2015; Krauss et al. 2017; Solís-Hernández and Fuchs 2019). Thus, pollination by volant vertebrates potentially results in larger genetic plant neighborhoods (sensu Wright 1946; Webb 1984) than pollination by insects (Karron et al. 1995; Krauss 2000; Krauss et al. 2009; Bezemer et al. 2016). Although studies on the contrasting effects of pollination by volant vertebrates vs. insects on plant gene flow are remarkably lacking, this idea is supported by pollination studies on focal species. For example, studies in entomophilous plants show that small insects such as flies, solitary bees, and small beetles generally visit most flowers in a single plant, and then move to nearby plants restricting foraging to relatively small areas (Campbell 1985; Escaravage and Wagner 2004; Hasegawa et al. 2015). Furthermore, large insects such as large
bees and lepidoptera have larger foraging areas, frequently associated with traplining behavior (i.e., repeated sequence of floral visits over several locations) (Levin 1979; Schmitt 1980; Murawski and Gilbert 1986; Rhodes et al. 2017). Similarly, vertebrate pollinators such as non-territorial hummingbirds and bats also follow a traplining foraging behavior (Fleming 1982; Lemke 1984, 1985; Tello-Ramos et al. 2015), and potentially cover even larger areas than large insects (Linhart 1973; Webb and Bawa 1983; Melampy 1987; Campbell and Dooley 1992; Sahley 2001; Castellanos et al. 2003; Serrano-Serrano et al. 2017). Taken together, pollination by volant vertebrates should increase the spatial scale of intraspecific plant gene flow relative to pollination by insects.

In this study we aimed to test two predictions: (1) insect pollination is associated with greater genetic differentiation between plant populations than hummingbird pollination, and (2) insect pollination is associated with stronger fine-scale spatial genetic structure (SGS) within plant populations than hummingbird pollination. We focused on six perennial understory angiosperms in the Andean cloud forest of northwestern Ecuador, a highly diverse but threatened ecosystem. Species belong to three families and within each family we selected one insect-pollinated species (euglossine bees, or small buzzing bees, or hoverflies and wasps), and one hummingbird-pollinated species (traplining hummingbirds) (Renner 1989; Gamba and Almeda 2014; Weinstein and Graham 2017; Dellinger et al. 2019) (Table 1). All six focal species are likely very limited in their seed dispersal, as they are dispersed by gravity or by understory birds with sedentary lifestyles (Renner 1989; Loiselle and Blake 1993, 1999; Kessler-
Ríos and Kattan 2012; Theim et al. 2014). Thus, we expect that any trend of variation in population genetic differentiation and SGS across species will be due primarily to pollination mode. We confirmed putative pollinators through field work, and we used a genome-wide genotyping approach to obtain genetic data. We then tested whether animal pollination mode explained differences in population genetic differentiation, as well as in strength of SGS, across species.

**Materials and Methods**

**Study sites**

We performed this study in Santa Lucía (0.12 N, 78.6 W), El Pahuma (0.02 N, 78.6 W), Bellavista (0.01 S, 78.7 W), and Las Tángaras (0.08 S, 78.8 W), four private reserves located on the northwestern slope of the Andean cordillera of Ecuador, in the province of Pichincha around 40 km northwest of Quito. Sites are 5–23 km apart from each other and are composed of secondary and primary cloud forest ranging from 1800–2500 m in elevation. Because they are nearby and similar in elevation, they share many species, yet the distance between them potentially imposes a physical barrier for movement of pollinators, making them ideal for testing our predictions.

**Study species and pollinators**

To select our focal species, we began by compiling a list of species occurring at all sites using the Tropicos.org database of the Missouri Botanical Garden. Through fieldwork we further narrowed this list to six perennial
understory angiosperms from three families, with one insect-pollinated and one hummingbird-pollinated species per family, including *Drymonia brochidodroma* Wiehler and *Drymonia tenuis* (Benth.) J.L. Clark (Gesneriaceae), *Miconia rubescens* (Triana) Gamba & Almeda and *Meriania tomentosa* (Cogn.) Wurdack (Melastomataceae), and *Notopleura longipedunculoides* (C.M. Taylor) C.M. Taylor and *Palicourea demissa* Standl. (Rubiaceae; with the hummingbird-pollinated species listed second in each case). Among study species, *M. tomentosa* is also pollinated by nectarivorous bats (Muchhala and Jarrín-V 2002).

Pairing by family allowed us to control for phylogenetic autocorrelation in subsequent tests. Based on our observations in the field, the spatial distribution of all species appeared widespread and consistent within sites, with occasional clusters of individuals. Additionally, seed dispersal in selected species is mostly achieved by understory birds with sedentary lifestyles such as tanagers and manakins, as has been shown for fleshy berries in Rubiaceae (Loiselle and Blake 1993, Loiselle et al. 1995; Theim *et al.* 2014) and Melastomataceae (Renner 1989; Loiselle and Blake 1999; Kessler-Ríos and Kattan 2012), and for fleshy capsules (often referred as display-capsules) in understory Gesneriaceae (Clark *et al.* 2012). The dry indehiscent capsules of *M. tomentosa* are likely gravity dispersed, as are many understory Melastomataceae with the same type of fruit (Renner 1989).

We obtained information on pollination mode from peer-reviewed literature of studied species (Renner 1989; Muchhala and Jarrín-V 2002; Gamba and Almeda 2014; Weinstein and Graham 2017; Dellinger *et al.* 2019), and by
videotaping plants in the field (Table 1). Specifically, for species with little information on pollination mode (*D. brochidodroma* and *N. longipedunculoides*), we confirmed putative pollinators by videotaping flowers with four high definition Sony digital camcorders for four days at each site. Cameras simultaneously videotaped four individuals per day (one species per day, eight individuals per species per site). Flowers were videotaped in the morning (0630 to 1130) and in the afternoon (1330 to 1830) (Additional file 1).

**Genomic sampling, library preparation and sequencing**

For molecular work, we collected leaf tissue in silica gel from 20 individuals per species from each of the three study sites (see Table 1 for sampled sites per species). We largely followed available trails in the reserves, making sure sampled individuals were at least 20 m apart from each other, and taking geographic coordinates in decimal degrees for each of them (Additional file 2).

We extracted total genomic DNA from silica-dried leaf tissue following the CTAB protocol (Doyle and Doyle 1987), but incorporating two additional ethanol washes of the DNA pellet. We quantified DNA with a Qubit 2.0 Fluorometer (Invitrogen, Thermo Fisher Scientific), using the manufacturer’s protocol. For each of our samples with sufficient DNA, we obtained single nucleotide polymorphisms (SNPs) using 2b-RAD, a restriction site-associated DNA sequencing technique (Wang *et al.* 2012). We constructed 2b-RAD libraries for each individual following the available protocol (Wang *et al.* 2012). Five hundred
ng of total genomic DNA were digested with a type IIb endonuclease, BgII (New England Biolabs), which cuts DNA on both sides of a recognition site to obtain uniform 36-bp fragments distributed across the genome. Oligonucleotide Illumina sequences were ligated to these fragments with 12 double-stranded barcoded adapters, one per each column of a 96-sample plate. In order to increase sequence coverage per locus, we utilized reduced representation barcoded adapters which reduce the total number of loci sequenced. Samples with different barcoded adapters were pooled into 8 groups of 12 samples. Following initial pooling, Illumina RAD PCR primers (1–8) were incorporated into the fragments of each pool via 14 cycles of PCR amplification. Amplified pools were then purified via gel electrophoresis. Fragments of 75bp were size selected by excising target bands from the agarose gel. We then used a Min Elute Gel extraction kit (Qiagen) to purify target bands. Purified samples were quantified and pooled into a single library in equimolar concentrations. We generated three libraries, which together included ~ 15–20 individuals per species per study site. Libraries were sequenced on Illumina HiSeq 4000 (Duke University, NC) machines, to generate single end 50 bp reads.

**Building loci and genotyping individuals**

Reads were demultiplexed using a custom script (trim2bRAD) generated by the Matz lab at the University of Austin, TX (https://github.com/z0on/2bRAD_denovo). This script trims 2b-RAD fragments from barcodes to produce one fastq file per sample. The resulting files were
quality filtered with FastQC (Babraham Bioinformatics) and the FASTX-toolkit (Gordon and Hannon 2010). We discarded low quality reads and obtained sequences that were 36 bp in length, with a minimum of 90% bases having a Phred quality score of at least 20 and an input quality offset of 33 (fastq files will be available in the Dryad repository). We then used the Stacks v2.3e pipeline to genotype individuals and produce a catalog of loci for each species (Catchen et al. 2013). We ran Stacks using the default parameter settings for building loci, which we considered to be appropriate for the short size of the 2b-RAD fragments, including a maximum distance of 2 nucleotide differences allowed between reads, a minimum depth of coverage of 3 reads required to create a stack, and a maximum distance of 4 nucleotide differences allowed to align secondary reads to primary stacks. We also allowed one gap between stacks before merging into putative loci. We filtered loci with the program ‘populations’ on the same pipeline. We excluded loci that were genotyped in <40% of individuals within each species. To avoid using SNPs in high linkage disequilibrium, we used one random SNP per locus. To prevent potential low-frequency SNP miscalls, we discarded alleles that had a frequency <5% in any locus across all individuals per species. To avoid repetitive or paralogous loci, the maximum number of heterozygous individuals that may be present in any locus was set to 75%. Lastly, we used the program VCFtools v0.1.16 (Danecek et al. 2011) to identify individuals with >50% missing data relative to variant sites and removed these individuals from subsequent analyses. We removed a total of 51 individuals across all species, with an average of 9 individuals/species (± 4
Inference of population genetic parameters

We used the program GenoDive v3.0 (Meirmans and Van Tienderen 2004) to calculate genetic diversity statistics for each species. We assessed population genetic structure using the F-statistics derived from an Analysis of Molecular Variance or AMOVA (Excoffier et al. 1992). AMOVA determines the proportion of genetic variance partitioned within individuals, among individuals within subpopulations, and among subpopulations. Related F-statistics were obtained with an infinite allele model; thus, they are equivalent to G-statistics (Nei 1973; Nei and Chesser 1983). These include $F_{IT}$ (the mean reduction in heterozygosity of an individual relative to the total population), $F_{IS}$ (the inbreeding coefficient among individuals within sites), and $F_{ST}$ (the global genetic differentiation among sampled sites). The statistical significance of diversity statistics was assessed using 1000 random permutations of the data, while their standard deviations were obtained by jackknifing over loci.

Inference of fine-scale spatial genetic structure (SGS)

We evaluated SGS for each species via spatial autocorrelation analyses at the individual level (Vekemans and Hardy 2004) using the program SPAGeDi v. 1.3a (Hardy and Vekemans 2002). We first transformed individuals’ decimal degrees coordinates into the Universal Transverse Mercator coordinate system, which is compatible with the SPAGeDi version we used. We then assessed
genetic relatedness between all pairs of individuals i and j with Nason’s kinship coefficient, $F_{ij}$ (Loiselle et al. 1995). We specified 5 distance intervals for each species and allowed the program to define their maximal distance such that the number of pairwise comparisons within each interval was kept approximately constant. $F_{ij}$ values were regressed on the natural logarithm of the spatial distance separating pairs of individuals, $\ln(d_{ij})$, in order to quantify regression slopes, $b$. To test for SGS, spatial positions of individuals were permuted 1000 times to obtain a frequency distribution of $b$ under the null hypothesis that $F_{ij}$ and $\ln(d_{ij})$ are not correlated. We quantified the strength of SGS with the $S_P$ statistic (Vekemans and Hardy 2004), which is calculated as $-b/(1 - F_1)$, where $F_1$ is the mean $F_{ij}$ between all pairs of individuals in the first distance interval containing nearest neighbors (< ~1 km for all species). The $S_P$ statistic mainly depends on the slope of the kinship-distance curve, allowing direct comparisons of SGS among species (Vekemans and Hardy 2004). Standard errors of all SGS statistics were obtained by jackknifing over loci. To visualize SGS, we plotted the mean $F_{ij}$ at each distance interval over the five distance intervals for each species.

**Testing for the effect of animal pollinators on plant $F_{ST}$ and SGS**

We used generalized linear mixed-effects models in RStudio V 1.2.5019 (R Core Team 2018) to examine if insect pollination is associated with both higher genetic differentiation across subpopulations (i.e., higher $F_{ST}$ values) and stronger SGS across individuals (i.e., higher $S_P$ values) than hummingbird
pollination, across our study species. Given that the natural logarithm of $F_{ST}$ and $S_P$ values are normally distributed, we fitted models with the R function glmer() and the ‘lognormal’ distribution (family=gaussian, link=’log’) for the structure of the residuals, specifying taxonomic family as a random effect.

**Results**

**Pollinators**

We recorded a total of 10 individuals and 30 hours (i.e., ~3 hours/individual) for *Drymonia brochidodroma*, and 12 individuals and 35 hours (i.e., ~2.9 hours/individual) for *Notopleura longipedunculoides*. From these videos, we observed that *D. brochidodroma* was exclusively visited by Euglossine bees, with 5 bee visits lasting ~10 seconds each, while *N. longipedunculoides* was visited by wasps, hoverflies, and small bees. We recorded 18 wasp visits lasting ~60 seconds each, 10 hoverfly visits ~30 seconds each, and 5 bees visits ~15 seconds each.

**Filtered genetic datasets**

After SNP calling and quality control using different filtering procedures, we obtained a mean of 2,797,308 SNP loci per species (± 1,091,949 SD; range: 879,138–4,151,836), with a mean coverage ranging from 14–95.1 read depth per loci across species (Table S1). After removing individuals with >50% missing data, final sample sizes of individuals per species per study site ranged from 8–18 (mean = 13 ± 3 SD), and the number of variant loci ranged from 1,044–4,907
(mean = 2,699 ± 1,427 SD) across species, with missing data across species ranging from 24–38% (mean = 33% ± 5 SD) (Table S2 and S3).

Gene diversity was similar across species; total expected heterozygosity (HT) ranged from 0.21–0.25 (mean = 0.23 ± 0.02) across species (Table S2) and mean expected heterozygosity within sites (HS) ranged from 0.17–0.26 (mean = 0.22 ± 0.02). Additionally, all species showed statistically significant levels of inbreeding, as indicated by significant GIS values whether these are pooled across sites (mean = 0.30 ± 0.14 SD; Table S2) or analyzed separately by site (mean = 0.32 ± 0.16 SD, Table S3).

**Population genetic structure**

AMOVA results revealed that in all species most of the genetic diversity resides within individuals and among individuals within sites, while less genetic diversity resides among sites (Table S4). AMOVA FIT showed that for most species a large proportion of individuals across study sites were out of Hardy-Weinberg equilibrium, likely due to inbreeding among individuals. In fact, AMOVA FIS was significant for all species, congruent with our GIS estimates above, and confirming that there is substantial genetic inbreeding within sites across studied species. Furthermore, AMOVA FST was variable (range = 0.03–0.21, mean = 0.10 ± 0.06) but significant for all species, hence there is considerable genetic differentiation among study sites (Table 2).
Fine-scale spatial genetic structure (SGS)

SGS was significant for all studied species; regression slopes $b$ of pairwise kinship coefficients on the natural logarithm of spatial distance were significantly negative in all species (Table 3). Additionally, the extent of SGS as quantified with the S$_P$ statistic was quite variable across species, ranging from 0.009–0.089 (mean = 0.04 ± 0.03 SD). Such variation is evident in our SGS visualizations (Fig. 1, Tables S5–S10), which show that species pollinated by insects tend to have steeper average kinship-distance slopes (Fig. 1 a, c, e) than species pollinated by vertebrates (Fig. 1 b, d, f). Given that standard errors associated with each average $F_{ij}$ are vanishingly small (Tables S5–S10), they are not observable in Fig. 1.

Effect of insect vs. vertebrate pollination modes on plant $F_{ST}$ and SGS

We hypothesized that insect pollination results in both stronger SGS and higher population genetic differentiation than hummingbird pollination. On average, plants pollinated by insects had greater $F_{ST}$ values (0.14 ± 0.07 SD) than plants pollinated by hummingbirds (0.06 ± 0.04 SD) (Table 2). We observed a similar trend for S$_P$ values; 0.054 ± 0.03 SD for plants pollinated by insects vs. 0.017 ± 0.01 SD for plants pollinated by hummingbirds (Table 3). Results from a generalized linear mixed-effects model (GLMM), specifying taxonomic family as a grouping factor, supported our predictions: insect pollination is associated with both significantly higher $F_{ST}$ and significantly higher S$_P$ values than vertebrate pollination (Fig. 2, Table 4).
Discussion

The contrasting effect of different animal pollinators on plant gene flow has remained largely unexplored across plant species. Our study provides an important advance in this matter and our results supported our predictions: species pollinated by insects had significantly greater levels of population genetic differentiation and stronger fine-scale spatial genetic structure than species pollinated by hummingbirds (Table 4, Fig.1 and 2). Our findings support the idea that pollinator movement during foraging has strong effects on the spatial scale of intraspecific plant gene flow. The limited movement of insects restricts gene flow within and among populations, while the traplining behavior of hummingbirds promotes genetic cohesion.

Our chosen study species allowed us to control for other factors that might impact plant population genetic structure and SGS, increasing the probability that the association we found is in fact due directly to animal pollination mode rather than a confounding variable. For example, choosing species pairs with distinct animal pollination modes (insect vs. vertebrate), each pair in one plant family, allowed us to control for evolutionary relationships that could have resulted in phylogenetic autocorrelation in our dataset. Furthermore, all species belong to cloud forest understory sites inside the southern end of the Choco Andean corridor (Mordecai et al. 2009) that are relatively well-connected by a continuous corridor of forests. Thus, pollinator movement between sites for all species should be constrained by the same type of geographic barriers inherent to the landscape heterogeneity of the Andes. Likewise, seed dispersal across species
is likely limited; seeds either fall under mother plants or are dispersed by sedentary understory birds like tanagers and manakins (Loiselle and Blake 1993, 1999; Smith 2001; Gamba and Almeda 2014). Additionally, most species pairs have the same type of fruit: x and x of gesner havaex, x and x of x have x. The exception are the Melastomataceae pair, in which Miconia rubescens has fleshy berries and Meriania tomentosa has indehiscent capsules. We would expect indehiscent capsules to be more dispersed limited that fleshy berries, resulting in higher $F_{ST}$ and $S_P$ values. Our data instead found that M. tomentosa has smaller $F_{ST}$ and $S_P$ values than M. rubescens, suggesting vertebrate pollination in the former may override any dispersal limitation imposed by the indehiscent capsules. Overall, we expect that seed dispersal likely contributes little to gene flow. Finally, differences in inbreeding levels do not seem to underlie the differences in population genetic differentiation or strength of SGS. Inbreeding can affect population genetic structure and SGS (Vekemans and Hardy 2004; Duminil et al. 2007), however we do not find such association in our dataset: the inbreeding coefficient (AMOVA $F_{IS}$ in Table 2) does not predict $F_{ST}$ (GLMM, $p=0.9$) or $S_P$ values (GLMM, $p=0.5$).

We note that differences in $F_{ST}$ and $S_P$ values were more pronounced between the Rubiaceae species pairs (7 and 10-fold, respectively), followed by the Melastomataceae pairs (2.2 and 2.5-fold, respectively), and lastly by the Gesneriaceae pairs (almost equivalent values) (Table 2 and 3). Notopleura longipedunculoides is largely pollinated by tiny wasps and hoverflies that probe most flowers in the same individual and stay among nearby plants (pers. obs),
consistent with the greatest observed $F_{ST}$ and $S_P$ values. *Miconia rubescens* is pollinated by *Melipona* and *Trigona*, which are relatively small pollen collecting bees (Renner 1989), consistent with the intermediate $F_{ST}$ and $S_P$ values. Finally, *Drymonia brochidodroma* is pollinated by euglossine bees (pers. obs.), which are larger and have been reported to flight long distances (Janzen 1971; López-Uribe *et al.* 2008), which is in line with *D. brochidodroma* having the smallest $F_{ST}$ and $S_P$ values among our insect pollinated plants. Thus, differences between insect pollinators may explain this pattern. Among vertebrate pollinated plants, *Palicourea demissa* is visited by ~15 hummingbird species, *Meriania tomentosa* is visited by ~8 hummingbird species and by nectarivorous bats (Muchhala and Jarrín-V 2002), and *Drymonia tenuis* is visited by ~7 hummingbird species (Weinstein and Graham 2017), consistent with lower $F_{ST}$ and $S_P$ values. The fact that the two *Drymonia* species had such similar $F_{ST}$ and $S_P$ values suggests that euglossine bees and hummingbirds are similar in their pollen dispersal ability. Direct measures of pollen dispersal based on paternity analyses are in line with the patterns of genetic structure we found, in that bats and hummingbirds can transport pollen for several kilometers, large insects such as large bees (including euglossine bees) for over 600 meters, while most small insects (smaller than a honeybee) rarely transfer pollen more than 300 meters (Webb and Bawa 1983; Dick *et al.* 2008).

One important consideration of our study is that we categorized pollination systems fairly broadly as insects vs. vertebrates. But in the same way that insects can vary in pollen dispersal ability, as described above, different
vertebrates may also differ in pollen dispersal. For instance, traplining vs. territorial behavior among hummingbirds might strongly impact plant gene flow (Murawski and Gilbert 1986; Cuevas et al. 2018; Schmidt-Lebuhn et al. 2019), since territorial hummingbirds have been shown to move pollen much shorter distances than traplining hummingbirds (Ohashi and Thomson 2009; Wolowski et al. 2013; Betts et al. 2015). There also might be differences between hummingbirds and bats, as the latter have been found to carry pollen more efficiently (Muchhala and Thomson 2010) and to longer distances than hummingbirds (Lemke 1984, 1985; Tello-Ramos et al. 2015). Future work should look more in depth at how plant gene flow is affected by differences within pollinator guilds, including large vs. small insects, territorial vs. traplining hummingbirds, and nectarivorous bats vs. hummingbirds.

Our study provides new evidence on the contrasting effect that different animal pollinators can have on the spatial scale of intraspecific plant gene flow. We found that insect-pollinated plants have significantly higher population genetic differentiation and stronger fine-scale spatial genetic structure than hummingbird pollinated plants. Thus, the effect of animal pollinators on plant gene flow is significant at local (within populations) and regional (among populations) scales. Our results support the idea that plants pollinated by insects are likely very susceptible to habitat fragmentation (more so than vertebrate pollinated plants; e.g. Côrtes et al. 2013), because it can further isolate populations and result in loss of genetic variability due to increased genetic drift (Aguilar et al. 2008, 2019). Nevertheless, focal studies reveal that hummingbird
and bat pollinated plants can also experience detrimental effects due to habitat fragmentation (Wanderley et al. 2020). Increased deforestation results in significant declines of hummingbird species richness and thus of pollinator availability (Hadley and Betts 2009; Hadley et al. 2018). Furthermore, habitat destruction due to urbanization likely decreases areas of cross-pollination mediated by nectarivorous bats, because their habitat becomes restricted to few forest fragments inside large tropical cities (Nunes et al. 2017). Future studies should seek to compare how animal foraging behavior and its related effect on plant gene flow might be altered due to anthropogenic disturbance. In general, the current scenario of human-accelerated change should push conservation efforts to maintain connectivity between fragments that harbor many understory tropical species.

Acknowledgements

Thanks to Nora Oleas and Paola Peña for help with the research permit in Ecuador (MAE-DNB-CM-2015-017). Robert Ricklefs, Christine Edwards, and Carmen Ulloa provided advice in this study. We also thank field assistants An Nguyen, Carlos Imery, and Alexander Lascher-Posner for their valuable help. Thanks to families at Santa Lucía, Bellavista, El Pahuma, and Las Tángaras cloud forest reserves for their conservation efforts and hospitality. Finally, we thank Amanda Grusz for help with SPAGeDi analyses. This research was supported by two graduate research grants from the Whitney R. Harris World Ecology Center at UMSL and one research grant from the American Society of...
Plant Taxonomists to DG.

References


Loiselle BA, Blake JG. 1993. Spatial distribution of understory fruit-eating birds and fruiting plants in a neotropical lowland wet forest In: Fleming TH,


Data accessibility statement

Should the manuscript be accepted, the data and R scripts supporting the results will be archived in Dryad and their DOI will be included at the end of this article.

Author Contributions

DG and NM planned and designed the research. DG collected and analyzed the data. DG wrote the initial draft of the manuscript. DG and NM contributed equally to substantial revisions of the manuscript.
<table>
<thead>
<tr>
<th>Species</th>
<th>Growth form</th>
<th>Pollinators (source)</th>
<th>Fruit type</th>
<th>Sites †</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Drymonia brochidodroma</em></td>
<td>Herbaceous</td>
<td>Euglossine bees (pers. obs.)</td>
<td>Fleshy capsule</td>
<td>SL, T</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>Sub-shrub</td>
<td>Traplining hummingbirds (1)</td>
<td>Fleshy capsule</td>
<td>SL, P, B</td>
</tr>
<tr>
<td><em>Miconia rubescens</em></td>
<td>Shrub</td>
<td>Small-buzzing bees (2)</td>
<td>Berry</td>
<td>SL, P, B</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>Shrub</td>
<td>Traplining hummingbirds/bats (1, 3)</td>
<td>Dry capsule</td>
<td>SL, P, B</td>
</tr>
<tr>
<td><em>Notopleura longipedunculoides</em></td>
<td>Sub-shrub</td>
<td>Wasps/flies/bees (pers. obs.)</td>
<td>Berry</td>
<td>SL, P, B</td>
</tr>
<tr>
<td><em>Palicourea demissa</em></td>
<td>Shrub</td>
<td>Traplining hummingbirds (1)</td>
<td>Berry</td>
<td>SL, B</td>
</tr>
</tbody>
</table>

(1) Weinstein and Graham 2017
(2) Gamba and Almeda 2004
(3) Dellinger et al. 2019

† SL: Santa Lucía, T: Las Tángaras, P: El Pahuma, B: Bellavista.
Table 2: Estimates of population genetic structure for each studied species. N total, number of genotyped individuals in the final genetic dataset; N loci, number of variant loci in the final genetic dataset; AMOVA $F_{IT}$ represents the deviation from Hardy-Weinberg Equilibrium within individuals relative to the expected heterozygosity in the total population; AMOVA $F_{IS}$ represents the inbreeding coefficient among individuals within sites; AMOVA $F_{ST}$ represents the global genetic differentiation among sampled sites. Population genetic parameters ($F_{IS}$ and $F_{ST}$) were all statistically significant ($p = 0.001$ in bold) based on 1000 permutations of the data.

<table>
<thead>
<tr>
<th>Species</th>
<th>N total</th>
<th>N loci</th>
<th>AMOVA $F_{IT}$ (SE)</th>
<th>AMOVA $F_{IS}$ (SE)</th>
<th>AMOVA $F_{ST}$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Drymonia brochidodroma</em></td>
<td>35</td>
<td>4907</td>
<td>0.42 (0.007)</td>
<td>0.37 (0.007)</td>
<td>0.08 (0.004)</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>29</td>
<td>1044</td>
<td>0.56 (0.014)</td>
<td>0.51 (0.015)</td>
<td>0.10 (0.010)</td>
</tr>
<tr>
<td><em>Miconia rubescens</em></td>
<td>34</td>
<td>2171</td>
<td>0.50 (0.009)</td>
<td>0.43 (0.010)</td>
<td>0.13 (0.005)</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>32</td>
<td>3883</td>
<td>0.29 (0.008)</td>
<td>0.24 (0.008)</td>
<td>0.06 (0.003)</td>
</tr>
<tr>
<td><em>Notopleura longipedunculoides</em></td>
<td>41</td>
<td>1815</td>
<td>0.35 (0.013)</td>
<td>0.17 (0.014)</td>
<td>0.21 (0.008)</td>
</tr>
<tr>
<td><em>Palicourea demissa</em></td>
<td>30</td>
<td>2376</td>
<td>0.22 (0.012)</td>
<td>0.19 (0.012)</td>
<td>0.03 (0.003)</td>
</tr>
</tbody>
</table>
Table 3 Estimates of SGS parameters for each studied species. N pairs, number of comparisons between all pairs of conspecific individuals; $F_1$, kinship coefficient between individuals in the first distance interval (separated by <1 km); $b\ln(\text{distance})$, slope of the regression of kinship coefficients on the natural logarithm of spatial distance; $S_P$, intensity of SGS for each species. Standard errors (SE) were obtained through jackknifing over loci. SGS parameters ($F_1$ and $b$) were all statistically significant ($p < 0.01$ in bold) based on 1000 permutations of individual locations.

<table>
<thead>
<tr>
<th>Species</th>
<th>N pairs</th>
<th>$F_1$ (SE)</th>
<th>$b\ln(\text{distance})$</th>
<th>$S_P$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drymonia brochidodroma</td>
<td>595</td>
<td>0.053 (0.003)</td>
<td>−0.024</td>
<td>0.025 (0.001)</td>
</tr>
<tr>
<td>Drymonia tenuis</td>
<td>406</td>
<td>0.044 (0.007)</td>
<td>−0.021</td>
<td>0.022 (0.002)</td>
</tr>
<tr>
<td>Miconia rubescens</td>
<td>561</td>
<td>0.105 (0.005)</td>
<td>−0.043</td>
<td>0.048 (0.002)</td>
</tr>
<tr>
<td>Meriania tomentosa</td>
<td>496</td>
<td>0.051 (0.002)</td>
<td>−0.018</td>
<td>0.019 (0.001)</td>
</tr>
<tr>
<td>Notopleura longipedunculoides</td>
<td>820</td>
<td>0.180 (0.006)</td>
<td>−0.073</td>
<td>0.089 (0.003)</td>
</tr>
<tr>
<td>Palicourea demissa</td>
<td>435</td>
<td>0.018 (0.002)</td>
<td>−0.009</td>
<td>0.009 (0.001)</td>
</tr>
</tbody>
</table>
Table 4 Results from generalized linear mixed-effects models with taxonomic family specified as a grouping factor and pollination mode as a fixed effect on (1) $F_{ST}$ values and (2) $S_p$ values across six species of cloud forest understory angiosperms. Significant p-values (<0.05) are denoted in bold.

<table>
<thead>
<tr>
<th>TEST</th>
<th>Response</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept</td>
<td>-2.88</td>
<td>0.46</td>
<td>-6.19</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>$F_{ST}$</td>
<td>0.91</td>
<td>0.44</td>
<td>2.07</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Groups</td>
<td>Variance</td>
<td>Std. Dev.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Family</td>
<td>0.03</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>0.002</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 2    | Intercept| -4.63    | 0.61       | -7.65   | <0.0001 |
|      | $S_p$    | 1.49     | 0.43       | 3.45    | 0.0006  |
|      | Groups   | Variance | Std. Dev.  |         |         |
|      | Family   | 0.14     | 0.39       |         |         |
|      | Residual | 0.0002   | 0.014      |         |         |
Figure 1 Average kinship-distance curves of each studied species. Filled symbols represent significant (p < 0.05) average kinship coefficient values based on 1000 permutations of individual spatial locations among all individuals. For associated standard errors of average $F_{ij}$ at each distance interval refer to tables S5–S10. (a) Drymonia brochidodroma. (b) Drymonia tenuis. (c) Miconia rubescens. (d) Meriania tomentosa. (e) Notopleura longipedunculoides. (f) Palicourea demissa.
**Figure 2** Marginal effect of animal pollination mode on predicted (a) $F_{ST}$ values and (b) $S_P$ values in the GLMMs with taxonomic family specified as a random effect. Black dots are predicted means for each category and surrounded black bars correspond to ± one standard deviation. Vertebrate and insect pollination modes were significantly different on both models ($p<0.05$).
Additional supporting information that will appear in the expanded online version of this article:

**Table S1** Unfiltered catalog of loci recovered with STACKS v2.3e

**Table S2** Genetic diversity of studied species across filtered loci.

**Table S3** Genetic diversity of studied species within sites.

**Table S4** AMOVA results showing the percent of genetic variation partitioning.

**Tables S5–S10** Results of the spatial genetic structure (SGS) analysis.
Table S1: Unfiltered catalog of loci recovered with the STACKS v2.3e pipeline for non-model organisms. N is the number of individuals. Coverage refers to the mean depth of reads used to build loci.

<table>
<thead>
<tr>
<th>Species</th>
<th>total N genotyped</th>
<th>total loci genotyped</th>
<th>Effective depth of coverage (reads/loci)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>stdev</td>
<td>min</td>
</tr>
<tr>
<td>Drymonia brochidodroma</td>
<td>37</td>
<td>4151836</td>
<td>14</td>
</tr>
<tr>
<td>Drymonia tenuis</td>
<td>42</td>
<td>3215520</td>
<td>16</td>
</tr>
<tr>
<td>Miconia rubesens</td>
<td>41</td>
<td>2923808</td>
<td>22.5</td>
</tr>
<tr>
<td>Meriania tomentosa</td>
<td>44</td>
<td>3164936</td>
<td>23.3</td>
</tr>
<tr>
<td>Notopleura longipedunculoides</td>
<td>50</td>
<td>879138</td>
<td>95.1</td>
</tr>
<tr>
<td>Palicourea demissa</td>
<td>38</td>
<td>2448608</td>
<td>18.9</td>
</tr>
</tbody>
</table>
Table S2 Genetic diversity of studied species estimated across filtered loci. N ind: number of genotyped individuals in the final genetic dataset, N var loci: number of variant loci, N total a: total number of alleles, %md: percent missing data, N a: mean number of alleles per locus, Ne a: mean effective number of alleles per locus, H0: observed heterozygosity, Hs: mean expected heterozygosity across subpopulations, Ht: total expected heterozygosity over all subpopulations, Gis: inbreeding coefficient. Standard deviations of statistics (in parentheses) were obtained through jackknifing over loci and significance of Gis (p < 0.005) through 1000 permutations (all were statistically significant).

<table>
<thead>
<tr>
<th>Species</th>
<th>N ind</th>
<th>N var loci</th>
<th>N total a</th>
<th>% md</th>
<th>N a</th>
<th>Ne a</th>
<th>H0</th>
<th>Hs</th>
<th>Ht</th>
<th>Gis</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Drymonia brochidodroma</em></td>
<td>35</td>
<td>4907</td>
<td>9800</td>
<td>37</td>
<td>2.00</td>
<td>1.29</td>
<td>0.14</td>
<td>0.22</td>
<td>0.23</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.01)</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>29</td>
<td>1044</td>
<td>2018</td>
<td>38</td>
<td>1.93</td>
<td>1.22</td>
<td>0.09</td>
<td>0.19</td>
<td>0.21</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.02)</td>
</tr>
<tr>
<td><em>Miconia rubescens</em></td>
<td>34</td>
<td>2171</td>
<td>4314</td>
<td>34</td>
<td>1.99</td>
<td>1.29</td>
<td>1.13</td>
<td>0.23</td>
<td>0.25</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.01)</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>32</td>
<td>3883</td>
<td>7727</td>
<td>30</td>
<td>1.99</td>
<td>1.31</td>
<td>0.18</td>
<td>0.23</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.01)</td>
</tr>
<tr>
<td><em>Notopleura longipedunculoides</em></td>
<td>41</td>
<td>1815</td>
<td>3621</td>
<td>24</td>
<td>2.00</td>
<td>1.28</td>
<td>0.18</td>
<td>0.21</td>
<td>0.25</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.01)</td>
</tr>
<tr>
<td><em>Palicourea demissa</em></td>
<td>30</td>
<td>2376</td>
<td>4691</td>
<td>32</td>
<td>1.97</td>
<td>1.29</td>
<td>0.18</td>
<td>0.22</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>
Table S3 Genetic diversity of studied species within sites estimated from filtered loci. B: Bellavista, T: Las Tángaras, P: Pahuma, S: Santa Lucía. N: number of individuals in the final genetic dataset, Ne: effective number of individuals, P a: number of private alleles, % P a: proportion of private to total alleles, N a: mean number of alleles per locus, Ne a: mean effective number of alleles per locus, H\textsubscript{O}: observed heterozygosity, H\textsubscript{S}: mean expected heterozygosity within site, G\textsubscript{IS}: inbreeding coefficient. Significance (p < 0.005) was obtained through 1000 permutations (all were statistically significant).

<table>
<thead>
<tr>
<th>Species</th>
<th>Site</th>
<th>N</th>
<th>Ne</th>
<th>P a</th>
<th>% P a</th>
<th>N a</th>
<th>Ne a</th>
<th>H\textsubscript{O}</th>
<th>H\textsubscript{S}</th>
<th>G\textsubscript{IS}</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Drymonia brochidodroma</em></td>
<td>T</td>
<td>18</td>
<td>12</td>
<td>930</td>
<td>0.19</td>
<td>1.83</td>
<td>1.31</td>
<td>0.14</td>
<td>0.22</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>17</td>
<td>11</td>
<td>737</td>
<td>0.15</td>
<td>1.81</td>
<td>1.30</td>
<td>0.14</td>
<td>0.22</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>9</td>
<td>7</td>
<td>74</td>
<td>0.07</td>
<td>1.51</td>
<td>1.24</td>
<td>0.10</td>
<td>0.17</td>
<td>0.44</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>P</td>
<td>11</td>
<td>8</td>
<td>80</td>
<td>0.08</td>
<td>1.65</td>
<td>1.28</td>
<td>0.09</td>
<td>0.21</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>9</td>
<td>8</td>
<td>114</td>
<td>0.11</td>
<td>1.52</td>
<td>1.26</td>
<td>0.09</td>
<td>0.19</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>8</td>
<td>7</td>
<td>183</td>
<td>0.08</td>
<td>1.57</td>
<td>1.34</td>
<td>0.10</td>
<td>0.24</td>
<td>0.58</td>
</tr>
<tr>
<td><em>Miconia rubescens</em></td>
<td>P</td>
<td>13</td>
<td>9</td>
<td>80</td>
<td>0.04</td>
<td>1.71</td>
<td>1.33</td>
<td>0.15</td>
<td>0.23</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>13</td>
<td>9</td>
<td>164</td>
<td>0.08</td>
<td>1.73</td>
<td>1.33</td>
<td>0.14</td>
<td>0.23</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>11</td>
<td>9</td>
<td>136</td>
<td>0.04</td>
<td>1.73</td>
<td>1.34</td>
<td>0.18</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>P</td>
<td>11</td>
<td>9</td>
<td>162</td>
<td>0.04</td>
<td>1.65</td>
<td>1.33</td>
<td>0.17</td>
<td>0.22</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>10</td>
<td>8</td>
<td>385</td>
<td>0.10</td>
<td>1.73</td>
<td>1.37</td>
<td>0.17</td>
<td>0.26</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>14</td>
<td>11</td>
<td>57</td>
<td>0.03</td>
<td>1.56</td>
<td>1.29</td>
<td>0.18</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td><em>Notopleura longipedunculoides</em></td>
<td>P</td>
<td>14</td>
<td>11</td>
<td>339</td>
<td>0.19</td>
<td>1.77</td>
<td>1.38</td>
<td>0.17</td>
<td>0.25</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>13</td>
<td>11</td>
<td>202</td>
<td>0.11</td>
<td>1.59</td>
<td>1.32</td>
<td>0.18</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>14</td>
<td>10</td>
<td>235</td>
<td>0.10</td>
<td>1.77</td>
<td>1.31</td>
<td>0.18</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td><em>Palicourea demissa</em></td>
<td>S</td>
<td>16</td>
<td>13</td>
<td>479</td>
<td>0.20</td>
<td>1.84</td>
<td>1.31</td>
<td>0.18</td>
<td>0.22</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Table S4. AMOVA results showing the percent of genetic variation partitioned within individuals, among individuals within sites, and among sites for all studied species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Within individuals</th>
<th>Among individuals within sites</th>
<th>Among sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Drymonia brochidodroma</em></td>
<td>0.58</td>
<td>0.34</td>
<td>0.08</td>
</tr>
<tr>
<td><em>Drymonia tenuis</em></td>
<td>0.44</td>
<td>0.46</td>
<td>0.10</td>
</tr>
<tr>
<td><em>Miconia rubescens</em></td>
<td>0.50</td>
<td>0.37</td>
<td>0.13</td>
</tr>
<tr>
<td><em>Meriania tomentosa</em></td>
<td>0.71</td>
<td>0.23</td>
<td>0.06</td>
</tr>
<tr>
<td><em>Notopleura longipedunculoides</em></td>
<td>0.65</td>
<td>0.14</td>
<td>0.21</td>
</tr>
<tr>
<td><em>Palicourea demissa</em></td>
<td>0.78</td>
<td>0.19</td>
<td>0.03</td>
</tr>
</tbody>
</table>
**Legend for Tables S5–S10:** Results of the spatial genetic structure (SGS) analysis based on all pairs of individuals within six studied species. Maximum distance: the upper limit of each distance interval. Mean distance: the average distance separating pairs of individuals within each interval. Mean ln(distance): the average natural logarithm of the distance separating pairs of individuals within each interval. Number of pairs: the number of pairs of individuals separated by the given distance interval. % partic: the percentage of individuals participating at least once in a pairwise comparison within each interval. CV partic: the coefficient of variation (i.e. the ratio of the standard deviation over the average) of the number of times each individual participates in pairwise comparisons within each interval. Kinship coefficients ($F_{ij}$) were calculated according to Loiselle et al. (1995). Respective standard errors (SE) were obtained through jackknifing over loci. Significance tests ($p < 0.05$ is denoted in bold) are based on the comparison of the observed $F_{ij}$ values with the corresponding frequency distributions of 1000 random permutations of individual spatial locations among all individuals (next 2 pages).
Table S5  SGS analysis for *Drymonia brochidodroma* based on 35 individuals and 4907 loci.

<table>
<thead>
<tr>
<th>Distance interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum distance (km)</td>
<td>0.47</td>
<td>2.06</td>
<td>17.40</td>
<td>17.93</td>
<td>23.43</td>
</tr>
<tr>
<td>Mean distance (km)</td>
<td>0.18</td>
<td>1.08</td>
<td>10.80</td>
<td>17.68</td>
<td>19.05</td>
</tr>
<tr>
<td>Mean ln(distance)</td>
<td>-2.04</td>
<td>0.01</td>
<td>2.14</td>
<td>2.87</td>
<td>2.94</td>
</tr>
<tr>
<td>Number of pairs</td>
<td>119</td>
<td>119</td>
<td>119</td>
<td>119</td>
<td>119</td>
</tr>
<tr>
<td>% partic</td>
<td>94</td>
<td>94</td>
<td>100</td>
<td>86</td>
<td>97</td>
</tr>
<tr>
<td>CV partic</td>
<td>0.43</td>
<td>0.45</td>
<td>0.95</td>
<td>0.47</td>
<td>0.76</td>
</tr>
<tr>
<td><em>F</em>&lt;sub&gt;i&lt;/sub&gt; (Loiselle)</td>
<td><strong>0.053</strong></td>
<td><strong>0.040</strong></td>
<td>-0.013</td>
<td><strong>-0.057</strong></td>
<td><strong>-0.069</strong></td>
</tr>
<tr>
<td>SE</td>
<td>0.003</td>
<td>0.003</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Table S6  SGS analysis for *Drymonia tenuis* based on 29 individuals and 1044 loci.

<table>
<thead>
<tr>
<th>Distance interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum distance (km)</td>
<td>0.62</td>
<td>5.50</td>
<td>10.32</td>
<td>13.26</td>
<td>15.57</td>
</tr>
<tr>
<td>Mean distance (km)</td>
<td>0.18</td>
<td>3.03</td>
<td>6.68</td>
<td>11.00</td>
<td>13.89</td>
</tr>
<tr>
<td>Mean ln(distance)</td>
<td>-2.26</td>
<td>0.87</td>
<td>1.87</td>
<td>2.40</td>
<td>2.63</td>
</tr>
<tr>
<td>Number of pairs</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>82</td>
</tr>
<tr>
<td>% partic</td>
<td>93</td>
<td>100</td>
<td>76</td>
<td>90</td>
<td>66</td>
</tr>
<tr>
<td>CV partic</td>
<td>0.58</td>
<td>0.66</td>
<td>0.74</td>
<td>0.68</td>
<td>0.78</td>
</tr>
<tr>
<td><em>F</em>&lt;sub&gt;i&lt;/sub&gt;</td>
<td><strong>0.044</strong></td>
<td><strong>0.005</strong></td>
<td>-0.032</td>
<td><strong>-0.021</strong></td>
<td><strong>-0.079</strong></td>
</tr>
<tr>
<td>SE</td>
<td>0.007</td>
<td>0.006</td>
<td>0.005</td>
<td>0.006</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Table S7  SGS analysis for *Miconia rubescens* based on 34 individuals and 2171 loci.

<table>
<thead>
<tr>
<th>Distance interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum distance (km)</td>
<td>0.51</td>
<td>5.56</td>
<td>10.42</td>
<td>11.12</td>
<td>15.29</td>
</tr>
<tr>
<td>Mean distance (km)</td>
<td>0.26</td>
<td>2.47</td>
<td>7.74</td>
<td>10.78</td>
<td>13.83</td>
</tr>
<tr>
<td>Mean ln(distance)</td>
<td>-1.59</td>
<td>0.45</td>
<td>2.01</td>
<td>2.38</td>
<td>2.62</td>
</tr>
<tr>
<td>Number of pairs</td>
<td>112</td>
<td>112</td>
<td>112</td>
<td>112</td>
<td>113</td>
</tr>
<tr>
<td>% partic</td>
<td>97</td>
<td>100</td>
<td>97</td>
<td>74</td>
<td>85</td>
</tr>
<tr>
<td>CV partic</td>
<td>0.52</td>
<td>0.54</td>
<td>0.76</td>
<td>0.72</td>
<td>0.79</td>
</tr>
<tr>
<td><em>F</em>&lt;sub&gt;i&lt;/sub&gt;</td>
<td><strong>0.105</strong></td>
<td><strong>0.007</strong></td>
<td><strong>0.013</strong></td>
<td><strong>-0.079</strong></td>
<td><strong>-0.104</strong></td>
</tr>
<tr>
<td>SE</td>
<td>0.005</td>
<td>0.005</td>
<td>0.003</td>
<td>0.005</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Table S8 SGS analysis for *Meriania tomentosa* based on 32 individuals and 3883 loci.

<table>
<thead>
<tr>
<th>Distance interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum distance (km)</td>
<td>0.36</td>
<td>5.65</td>
<td>11.94</td>
<td>15.40</td>
<td>16.56</td>
</tr>
<tr>
<td>Mean distance (km)</td>
<td>0.20</td>
<td>3.19</td>
<td>8.31</td>
<td>12.80</td>
<td>15.82</td>
</tr>
<tr>
<td>Mean ln(distance)</td>
<td>-1.88</td>
<td>0.65</td>
<td>2.06</td>
<td>2.55</td>
<td>2.76</td>
</tr>
<tr>
<td>Number of pairs</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>% partic</td>
<td>91</td>
<td>91</td>
<td>94</td>
<td>88</td>
<td>100</td>
</tr>
<tr>
<td>CV partic</td>
<td>0.55</td>
<td>0.81</td>
<td>0.63</td>
<td>0.63</td>
<td>0.77</td>
</tr>
<tr>
<td>$F_{ij}$</td>
<td>0.051</td>
<td>0.005</td>
<td>-0.013</td>
<td>-0.048</td>
<td>-0.028</td>
</tr>
<tr>
<td>SE</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table S9 SGS analysis for *Notopleura longipedunculoides* based on 41 individuals and 1815 loci.

<table>
<thead>
<tr>
<th>Distance interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum distance (km)</td>
<td>0.55</td>
<td>5.33</td>
<td>10.42</td>
<td>13.46</td>
<td>16.28</td>
</tr>
<tr>
<td>Mean distance (km)</td>
<td>0.30</td>
<td>2.90</td>
<td>6.68</td>
<td>11.60</td>
<td>14.45</td>
</tr>
<tr>
<td>Mean ln(distance)</td>
<td>-1.43</td>
<td>0.78</td>
<td>1.87</td>
<td>2.45</td>
<td>2.67</td>
</tr>
<tr>
<td>Number of pairs</td>
<td>164</td>
<td>164</td>
<td>164</td>
<td>164</td>
<td>164</td>
</tr>
<tr>
<td>% partic</td>
<td>100</td>
<td>100</td>
<td>81</td>
<td>83</td>
<td>68</td>
</tr>
<tr>
<td>CV partic</td>
<td>0.41</td>
<td>0.62</td>
<td>0.75</td>
<td>0.66</td>
<td>0.75</td>
</tr>
<tr>
<td>$F_{ij}$</td>
<td>0.180</td>
<td>0.046</td>
<td>-0.006</td>
<td>-0.118</td>
<td>-0.144</td>
</tr>
<tr>
<td>SE</td>
<td>0.006</td>
<td>0.007</td>
<td>0.003</td>
<td>0.006</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Table S10 SGS analysis for *Palicourea demissa* based on 30 individuals and 2376 loci.

<table>
<thead>
<tr>
<th>Distance interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum distance (km)</td>
<td>0.36</td>
<td>1.20</td>
<td>14.02</td>
<td>14.44</td>
<td>16.47</td>
</tr>
<tr>
<td>Mean distance (km)</td>
<td>0.20</td>
<td>0.74</td>
<td>8.61</td>
<td>14.22</td>
<td>15.24</td>
</tr>
<tr>
<td>Mean ln(distance)</td>
<td>-1.82</td>
<td>-0.38</td>
<td>1.71</td>
<td>2.65</td>
<td>2.72</td>
</tr>
<tr>
<td>Number of pairs</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>% partic</td>
<td>87</td>
<td>100</td>
<td>100</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>CV partic</td>
<td>0.67</td>
<td>0.49</td>
<td>0.83</td>
<td>0.66</td>
<td>0.81</td>
</tr>
<tr>
<td>$F_{ij}$</td>
<td>0.018</td>
<td>0.018</td>
<td>-0.019</td>
<td>-0.028</td>
<td>-0.014</td>
</tr>
<tr>
<td>SE</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Chapter IV: Impact of animal pollinators and latitudinal regions on the spatial genetic structure of plants: a global test

Running title: Drivers of spatial genetic structure in plants

Diana Gamba, Department of Biology, University of Missouri at Saint Louis
dlgtk5@mail.umsl.edu

Nathan Muchhala, Department of Biology, University of Missouri at Saint Louis
muchhalan@umsl.edu

Author for correspondence: Diana Gamba; One University Boulevard, 223R Research Hall, Saint Louis, MO 63121, USA; dlgtk5@mail.umsl.edu; +1 (314) 702-0326.

Target journal: Proceedings of the Royal Society B: Biological Sciences

Abstract: 200 words

Total word count (main text): 5178
Abstract

Spatial genetic structure (SGS) in plants results from the nonrandom distribution of genotypes within populations, which is influenced by life-history traits including mating system, growth form, and seed dispersal mode. However, the effect of animal pollination and latitudinal region remain largely unknown. Based on their lower flying ability compared to other animals, we predict that SGS should be stronger in plants pollinated by small insects relative to plants pollinated by large insects and vertebrates. Likewise, we predict that plant SGS should be stronger in the tropics than in temperate zones, because higher spatial heterogeneity at local scales, lower population densities and higher species richness in the tropics may restrict plant gene flow. To test our predictions, we performed a literature review and assembled a 147-species global dataset of animal-pollinated plants with data on SGS intensity, as quantified with the $S_p$ statistic. Generalized linear models demonstrated that pollination mode, latitudinal region, and growth form were all significant predictors of $S_p$ values, while mating system and seed dispersal mode were not significant. Our findings strongly supported our predictions, particularly in non-woody plants and shrubs, highlighting differences among latitudinal regions, and the importance of animal pollination mode in shaping patterns of plant SGS.

**Key words:** animal pollination, flowering plants, fine-scale spatial genetic structure, latitudinal region, SGS, $S_p$ statistic.
**Introduction**

Fine-scale spatial genetic structure (SGS) in plants results from the non-random distribution of closely related individuals in space and represents the spatial scale of intraspecific gene flow within populations [1]. Understanding the factors that affect plant SGS is critical for analyzing demographic patterns such as the extent of genetic cohesion, or ‘neighborhood size’ [2,3], within natural and fragmented populations. Likewise, factors that influence plant SGS can strongly affect evolutionary processes within populations, such as local adaptation [4], and the maintenance of genetic diversity [5]. Plant life-history traits such as mating system, growth form, pollination mode and seed dispersal mode can influence patterns of SGS because they are directly involved in gene dispersal. In general, selfing herbs have significantly greater SGS than outcrossing trees [6], and animal-pollinated plants have greater SGS than wind-pollinated ones [5]. Additionally, SGS is greater in species with short-distance dispersers, lower in species dispersed by birds, and highly variable in species dispersed by active or passive seed accumulators [4,5], suggesting that dispersal limitation leads to high SGS. In fact, seed dispersal is often assumed to be the main determinant of SGS [7]. However, this relationship will ultimately depend on how successfully seeds establish and become adult plants. If most seeds fall under a mother plant—a common sign of dispersal limitation—but do not survive, then other factors that affect plant gene flow, such as pollination mode and landscape heterogeneity in a given region, should become important determinants of plant SGS. The effect of different animal pollinators on broad-scale patterns of plant
SGS, however, remains largely understudied.

Different pollinators can differ substantially in their flying ability and pollen carry-over capacity. Volant vertebrates and large insects, for example, generally fly longer distances during foraging bouts than small insects [5,8–12]. Studies on pollen carry-over in entomophilous plants reveal that small insects such as flies, solitary bees, and small beetles generally visit most flowers in a single plant, and then usually stay among nearby plants in the same patch [10–13]. In contrast to this, bumblebees are generally associated with significantly greater pollen carry-over and pollen dispersal distances [15]. For example, enclosed experiments and studies in natural populations show that although bumblebees deposit most pollen in nearby plants, significant amounts of pollen are transported to more distant flowers even after grooming [14,16,17]. Similarly, honeybees deposit pollen across distances three times larger than predicted by common exponential functions that evaluate pollen deposition, fitting a leptokurtic distribution comparable to that of bumblebees [18,19]. Furthermore, bumblebees and butterflies are highly directional in their flight while foraging, suggesting they can increase pollen flow distances when pollen carry-over is successful [8,20]. Studies of pollinator movement show that euglossine bees, hawkmoths, birds and bats can all travel quite far, even across fragmented habitats, potentially connecting individual plants across large distances [21–29]. In support of this, direct measures of pollen dispersal reveal that bats can transport pollen for several kilometers, large insects such as honeybees can transport pollen for >600 meters, while pollen transfer by most small insects (smaller than a
honeybee) rarely reaches 300 meters (reviewed in [5]). Based on these differences in the extent of pollen dispersal among animal pollinators, we predict that plants pollinated by small insects (smaller than a honeybee) should have stronger SGS than plants pollinated by large insects (honeybee or larger) or volant vertebrates (nectar-feeding birds and bats).

Furthermore, the influence of different latitudinal regions (i.e., temperate, tropical, subtropical), which differ substantially in landscape heterogeneity, is poorly understood. Across broader latitudinal scales, there are important environmental differences that may result in distinct patterns of SGS between plants in different latitudinal regions. For example, tropical regions have substantial habitat heterogeneity at a local scale, resulting in contrasting microclimates that could restrict plant demographic-range expansion at a given site [30–32]. Such restriction could limit gene flow within plant populations, and in turn potentially increase plant SGS in tropical plants relative to temperate ones. Subtropical forests similarly show considerable heterogeneity at a local scale compared to temperate ones [33], which could also result in higher plant SGS in subtropical than in temperate regions. Moreover, population densities tend to be significantly lower in tropical regions than temperate zones, which is usually associated with higher species diversity [5]. For instance, in a study of Ardisia crenata populations in subtropical China, sites with low population density and high species diversity were associated with greater SGS, relative to sites with high population density and low species diversity [34]. Given all of the above, we predict that species in tropical and subtropical regions should associate with
stronger SGS than species in temperate regions.

The strength of SGS can be quantified with the $S_P$ statistic [6], which is based on a model of isolation by distance at migration–drift equilibrium [2,3]. This model describes the degree to which genetic relatedness between individuals, as quantified with the kinship coefficient $F_{ij}$ [1], decreases with increasing geographic distance. $S_P$ is defined as $-b/(1 - F_1)$, where $b$ is the regression slope of genetic relatedness ($F_{ij}$) on geographic distance ($d_{ij}$) between individuals $i$ and $j$, and $F_1$ is the mean $F_{ij}$ [1] between all pairs of individuals in the first distance interval containing nearest neighbors. Because $S_P$ mainly depends on the regression slope $b$, it is not affected by an arbitrary choice of distance intervals defined in a given study, making it comparable across species and thus ideal for investigating the factors that affect the strength of plant SGS globally. Additionally, studies that use the $S_P$ statistic to characterize plant SGS frequently work at intermediate spatial scales (typically tens to hundreds of kilometers) at which both pollen and seed dispersal patterns have important effects on genetic diversity and population structure [5]. This is because the majority of seed dispersal often occurs at a small scale (i.e., $<0.1$ km), at which its effect is expected to determine plant SGS. At larger scales, i.e., beyond the bulk of seed dispersal, pollen dispersal can become equally or more important [5,35]. Thus, studies that report $S_P$ values allow investigation of the effects of pollen dispersal mode across zoophilous species.

While the effects of animal pollination mode and latitudinal region have been largely overlooked in previous reviews on plant SGS variation [4–6,35],
they were evaluated in a recent review on global patterns of population genetic differentiation in seed plants based on $F_{ST}$ values (D. Gamba and N. Muchhala, *in review*). Results of that study showed that tropical and subtropical mixed-mating non-woody plants pollinated by small insects were associated with higher $F_{ST}$ values relative to temperate outcrossing trees and to plants pollinated by large insects and vertebrates. $F_{ST}$ represents the proportion of genetic diversity partitioned among subpopulations, relative to the total population, and is usually taken at larger geographic scales than SGS studies (typically hundreds to thousands of kilometers). Thus, the $S_P$ statistic describes isolation by distance among conspecific individuals, while the $F_{ST}$ statistic may be used to examine isolation by distance among conspecific subpopulations [2,36,37]. Although $S_P$ and $F_{ST}$ values describe the arrangement of genetic diversity at different spatial scales, i.e., within (fine-scale) and among (large-scale) populations, respectively, the same processes, namely genetic drift, gene flow, and selection, underlie their patterns of variation. Thus, we expect that the same factors that affect $F_{ST}$ also affect $S_P$, in line with our predictions. To our knowledge, however, no study to date has tried to connect patterns of $S_P$ and $F_{ST}$ variation. Furthermore, because seed dispersal is generally considered to be more important locally [4,5], it likely affects plant $S_P$ values more than plant $F_{ST}$ values. On the other hand, because pollen dispersal can generally reach longer distances [5,35], it likely affects plant $S_P$ values as much as plant $F_{ST}$ values.

Here, we took advantage of the wealth of publications that report $S_P$ values and assembled a 147-species dataset of animal-pollinated plants at a
global scale. To the best of our knowledge, ours is the largest plant SGS dataset to be analyzed to date. We aimed to evaluate the effect of animal pollination mode and latitudinal region on $S_P$ values, while also accounting for other factors that have been shown to affect $S_P$, namely mating system, growth form, seed dispersal mode, and genetic marker choice. Using multiple regressions, we tested two predictions: (1) that species pollinated by small insects (smaller than a honeybee) have on average greater $S_P$ values that species pollinated by large insects (honeybees or larger) and vertebrates (hummingbirds and bats), and (2) that species from regions at tropical and subtropical latitudes have on average greater $S_P$ values that species from regions at temperate latitudes. We also examined the relative contributions of factors to explaining variation in $S_P$ values, in order to identify the most important factor affecting plant SGS.

**Materials and Methods**

**Dataset compilation**

We constructed an $S_P$ dataset by conducting a systematic literature search in Google Scholar (key words: “fine-scale spatial genetic structure” OR “SGS” OR “spatial genetic structure” OR “$S_P$ statistic”) focused on articles published through June 2018. This search yielded 254 peer-reviewed publications on seed plants for which $S_P$ values based on nuclear markers were available. We also included 6 more species from a recent unpublished study (D. Gamba & N. Muchhala, *in prep.*). Because we were mainly interested in animal-pollinated plants, we did not include wind-pollinated or selfing species in the
database. Furthermore, we only considered studies of adult plants, rather than on seedlings or saplings, given that adults should better represent the long-term effects of animal pollinators on SGS. Based on these criteria, our final dataset included mean S_P values and metadata for 147 species (Table S1, Appendix S1). When a single study reported S_P values for multiple populations of the same species, we calculated the mean S_P value for all populations surveyed. When multiple studies reported S_P values for the same species, we calculated the mean S_P value for all populations across studies. For clonal species (Asclepias syriaca and Piper sp.), we used the published S_P value based on genets (excluding clones).

Previous studies suggest that the S_P statistic can be unduly influenced by the genetic marker chosen to infer SGS parameters [4,38,39]. Thus, we also scored the genotyping technique used for each species (microsatellites; allozymes; AFLP: amplified fragment length polymorphism; SNP: single-nucleotide polymorphisms). When a single species was analyzed with multiple markers, we used the marker with the greatest sample size of individuals per population. We did not include studies based on RAPD (randomly amplified polymorphic DNA) markers, because these were scarce (N = 3) and we wanted to minimize potential bias on S_P estimates due to marker type.

**Species traits**

We extracted information on species traits directly from the source publications, including pollination mode (small insects; large insects;
vertebrates), latitudinal region (tropics; subtropics; temperate), growth form (non-woody; shrub; tree), mating system (mixed-mating; outcrossing), and seed dispersal mode (animals; gravity; wind). Below, we explain how we coded factors in more detail.

*Pollination mode*— Small insect pollinators of species in our dataset included small Hymenoptera (*Trigona* and *Melipona* bees and wasps), Diptera (hoverflies and gnats), Coleoptera (small curculionids), Hemiptera (Anthocoridae and Miridae), and Thysanoptera (i.e., thrips). Large insects included large bees (honeybees, bumblebees, carpenter bees, euglossine bees) and Lepidoptera (hawk moths and yucca moths, monarch butterflies). Vertebrates included bats, hummingbirds, and other nectarivorous birds such as honeyeaters and sunbirds.

*Latitudinal region*— Tropical regions included sites between the Tropic of Cancer and Tropic of Capricorn (23.5° north and south of the equator, respectively), sub-tropical regions included latitudes from 23.5° to 35° (north and south of the equator), and temperate regions included latitudes greater than 35° (north and south of the equator).

*Growth form*— Trees included woody plants >10 m tall, typically with a single trunk coming from the base. Shrubs included upright woody plants <10 m tall, typically with one or several trunks coming from the base. Hemi-epiphytes (*Ficus citrifolia* and *F. obtusifolia*) and woody climbers (*Ancistrocladus korupensis*) were included in the shrub category, while epiphytes (*Aechmea nudicaulis*) and non-woody climbers (*Borderea pyrenaica*, *Dioscorea japonica*, and *Haumania danckelmaniana*) were included in the non-woody category.
**Mating system**—Mixed-mating species included those that undergo both outcrossing and selfing to some extent, through either autogamy or geitonogamy. Outcrossing species included plants that are self-incompatible, unisexual (i.e. monoecious or dioecious), or dichogamous hermaphrodites—i.e. either having the male reproductive organs come to maturity before the female organs (protandry), or vice versa (protogyny).

**Seed dispersal mode**—Plants that presented fruits or seeds that were particularly light and/or winged were coded as wind dispersed. Plants with no adaptations for vector-mediated seed dispersal were coded as gravity dispersed. Publications often did not include disperser identities for animal-dispersed species, and some species were dispersed by many taxonomic groups, making animal dispersal difficult to characterize. Thus, we maintained a broad animal dispersal category including all zoochorous plants (effects of zoochory on plant SGS are reviewed in [4]).

**Statistical analyses**

We used multiple regression models to examine the influence of different animal pollinators and latitudinal regions on plant SGS intensity, while accounting for other potentially significant predictors (growth form, mating system, seed dispersal mode, and genetic marker). Given that natural logarithm-transformed $S_p$ values are normally distributed, we fitted generalized linear models (GLMs) with the ‘glm’ function in RStudio V 1.2.5019 [40] under a lognormal distribution structure for the residuals (family = ‘Gaussian’, link = ‘log’). First, we built a GLM
that included all variables to estimate multicollinearity between predictors with the
generalized variance inflation factor (GVIF) [41] calculated using the ‘vif’ R
function. All GVIF values were >1 and <3.05 (Table S2), indicating the presence
of some correlations among predictors, but that these were not sufficiently
problematic to create multicollinearity issues negatively influencing a multiple
regression [42]. Then, we examined our most inclusive model and sequentially
removed factors that did not significantly contribute to the explained variation in
$S_P$ values in order to find the best-fit model to the data. We compared the fit of
GLMs using model selection based on the Akaike Information Criterion (AIC)
[43,44]. Finally, we tested for two-way interactions of pollination mode and
latitudinal region with other factors in the best-fit model.

In order to measure and account for potential autocorrelations among the
data due to evolutionary relationships, we calculated phylogenetic signal in the
residual error of all models simultaneously with the regression parameters,
following recommendations by Revell [45]. We extracted a species-level
phylogeny containing our focal taxa (Fig. 1) from the angiosperm mega-tree [46]
available in the V.PhyloMaker R package [47]. Branch lengths were inferred
using the branch length adjuster algorithm in V.PhyloMaker [48]. Phylogenetic
signal was measured with Pagel’s $\lambda$ [49] as implemented in the ‘phylosig’ R
function in phytools [50]. We consistently obtained $\lambda < 0.001 (p = 1)$, indicating a
lack of phylogenetic autocorrelation in the residuals of our GLMs; thus, we only
present and interpret results from non-phylogenetic GLMs.

After finding the best-fit model, we used the rr2 R package [51] and the
‘R2.lik’ function to obtain the unique contribution of each factor, in terms of the amount of S_P variance explained, by comparing the best-fit model with a reduced model not including the factor of interest. We also obtained the partial R^2 for each interaction term found to be significant. We visualized the marginal effect of each factor on S_P values in the best-fit model using the R packages sjPlot and ggplot2 [52,53] and the function ‘plot_model’ (with type = ‘eff’). For conditional effects among factors (i.e., interactions), we set the plot_model type to ‘int’.

**Results**

**Taxonomic scope and phylogeny**

The 147 animal-pollinated species were distributed in 113 genera, representing 54 families in 28 orders. The majority of species (118) belonged to the Eudicots, followed by 20 Monocots, 8 Magnoliids, and one Gymnosperm (Zamia fairchildiana). The families Fabaceae and Moraceae (mostly Ficus; 9 species) were the most well represented in the dataset, with 16 and 10 species, respectively (Table S1). The resulting phylogeny had 147 tips and 138 internal nodes (Fig. 1), indicating that 94% of the phylogeny was resolved, and only 9 tips (6%) belonged to polytomies. These polytomies were located within clades for which phylogenetic information remains scarce or unclear [54]: Alcantarea (Bromeliaceae) and Psychotria (Rubiaceae).

**Best-fit model explaining variation in SGS intensity**

Among the predictors we tested, pollination mode, latitudinal region and
life form had significant effects on \( S_P \) values, while the effect of mating system was only marginally significant (Table 1). Seed dispersal mode and genetic marker did not enter the best-fit model. Although animal-dispersed plants, and plants for which \( S_P \) was obtained with AFLP markers, tended to have slightly higher mean \( S_P \) values than the other groups (Fig. S1), these differences were not statistically significant \((p > 0.05)\). In fact, removing these factors from the most-inclusive model (Table S3) greatly increased model fit to the data \((\Delta AIC = 5.95)\).

Our estimation of the relative contribution of each factor to the explained variance of \( S_P \) values showed that growth form was the most important predictor in the best-fit model, with a partial \( R^2 \) of 0.20. Latitudinal region was second in importance with a partial \( R^2 \) of 0.13, followed by pollination mode (partial \( R^2 = 0.05 \)), and lastly by mating system (partial \( R^2 = 0.02 \)).

**Patterns of \( S_P \) variation**

Our results reveal that species pollinated by small insects are associated with significantly greater \( S_P \) values than species pollinated by vertebrates and large insects, while the latter two animal pollination modes did not differ from each other (Fig. 2a). We also found that species in tropical regions have significantly greater \( S_P \) values than species in subtropical and temperate regions, while the latter two regions did not differ from each other (Fig. 2b). Consistent with initial expectations, we confirm that trees have significantly lower \( S_P \) values relative to non-woody plants and shrubs. The three types of growth form were
also significantly different from each other, with mean $S_P$ values increasing from trees to shrubs to non-woody plants (Fig. 2c). Lastly, mixed-mating plant species were associated with marginally higher $S_P$ values than outcrossing species (Fig. 2d).

Because we were mostly interested in examining the effect of different animal pollinators and latitudinal regions on $S_P$ values, we tested for interactions between pollination mode and latitudinal region with the other factors in our best-fit model, respectively. First, we found that differences between animal pollinators were significantly conditional on growth form ($p = 0.03$). Pollination by small insects is associated with higher mean $S_P$ values relative to vertebrate and large insect pollination in non-woody plants and shrubs, but not in trees. Rather, vertebrate pollination tends to increase mean $S_P$ in trees relative to large insects (Fig. 3a). The amount of variance explained by the model with this interaction was $R^2 = 0.26$, and this interaction had a partial $R^2 = 0.04$. Including it in the best-fit model, however, decreased model fit to the data (model with interaction AIC = $-721.57$, $\Delta$AIC = 2.58). Second, we found that differences between latitudinal regions are marginally conditional on growth form ($p = 0.08$). Tropical regions tend to be associated with higher $S_P$ values relative to subtropical and temperate zones in non-woody plants, but not in shrubs and trees. In shrubs, tropical regions seem related with higher $S_P$ values relative to subtropical regions, while values from temperate regions were highly variable and appeared not different from other regions. Trees, on the other hand, did not seem to differ in $S_P$ values among latitudinal regions (Fig. 3b). The amount of variance explained by the
model with this interaction was $R^2 = 0.26$, and this interaction had a partial $R^2 = 0.03$. Including this interaction in the best model, however, decreased model fit to the data (model with interaction AIC = $-720.11$, $\Delta AIC = 4.04$).

**Discussion**

Here, we analyzed for the first time the effects of animal pollination mode and latitudinal region on plant SGS using a comprehensive global dataset of $S_P$ values. Our results revealed a number of interesting patterns. Strikingly, we found that small insect pollination significantly increases $S_P$ values relative to large insect and vertebrate pollination, particularly in non-woody plants and shrubs (Fig. 2a, 3a). Likewise, species from tropical regions are associated with higher $S_P$ values relative to those from subtropical and temperate regions, especially for non-woody plants (Fig. 2b, 3b). Growth form was the most important predictor of $S_P$ values relative to the other factors, followed by latitudinal region and pollination mode, while mating system was the least important and only marginally significant. Seed dispersal mode and genetic marker were not significant predictors of $S_P$. Before discussing the roles of these different factors in influencing SGS in more detail, below we compare our results to those from a review on global patterns of population genetic differentiation (as quantified with the $F_{ST}$ statistic) in seed plants (D. Gamba & N. Muchhala, in review).

Our results are largely concordant with general patterns of variation in $F_{ST}$ values, particularly with our predictions in respect to animal pollination mode and
latitudinal region. In general, small insect pollination is associated with higher $F_{ST}$ and $S_P$ values compared to both large insect and vertebrate pollination. Similarly, species from tropical regions have significantly higher $F_{ST}$ and $S_P$ values compared to species from temperate regions. Additionally, trees have significantly lower $F_{ST}$ and $S_P$ values relative to non-woody plants. These patterns of variation suggest that the same factors affect the arrangement of genetic diversity at different spatial scales: from fine-scale spatial structure within populations to broad-scale spatial structure among populations. Although this is expected given that any structuring of genetic diversity ultimately depends on the fundamental processes of gene flow, genetic drift and selection, ours is the first study we are aware of to link patterns of $F_{ST}$ and $S_P$ variation at a broad scale. Furthermore, seed dispersal mode was also not significant for explaining variation in $F_{ST}$ or $S_P$ values. Because seed dispersal is generally considered to be more important at local scales [1,4–7,60], we expected that it would have an effect on $S_P$ values, particularly when comparing gravity vs. other modes of seed dispersal. We think that unrecorded secondary movement of seeds that fall under mother plants potentially precluded us from finding such difference. Finally, one difference between patterns of variation of $F_{ST}$ and $S_P$ values was the effect of mating system. It was a significant predictor for $F_{ST}$ values, but only marginally significant for $S_P$ values, with mixed-mating species generally associated with higher values. This was somewhat unexpected, given that mating system affects inbreeding, which lowers within-population variation, inflating between-population differentiation. Thus mixed-mating should increase both $F_{ST}$ and $S_P$ values due to
increased local genetic drift. Our result could simply be due to considerable amounts of outcrossing among the mixed-mating species in our \( S_P \) dataset, counteracting local genetic drift.

**Influence of pollination mode on \( S_P \)**

The strength of SGS was higher in species pollinated by small insects than in species pollinated by large insects and vertebrates (Fig. 2a). This is in line with differences in foraging behavior, pollen carry-over capacity, and flying ability among animal pollinators, which indicate that pollen dispersal by small insects is more limited compared to large insects and vertebrates [5,8,15]. Direct measures of pollen dispersal based on paternity analyses also support the limited distance covered by small insects in trees, as they reach maximum 300 meters [5]. This idea is also supported by indirect measures of pollen dispersal—i.e., obtained from observed SGS values derived from an isolation-by-distance process at equilibrium combined with estimates of the effective population density—which suggest they rarely surpass 20 meters in non-woody plants and shrubs [6,11,34], and 265 m in trees [5]. A remarkable exception is the pollen dispersal of fig trees by tiny agaonid wasps, which with the help of wind can achieve cross-pollination between trees separated by several kilometers [55]. Our dataset included 5 *Ficus* trees classified as pollinated by small insects. The mean \( S_P \) value for such *Ficus* was 0.017 (± 0.015 SD), which was not lower than expected compared to the mean \( S_P \) value of other tree species pollinated by small insects (0.013 ± 0.01 SD). However, the mean \( S_P \) value for all trees
pollinated by small insects (0.014 ± 0.01 SD) was considerably lower than that of non-woody plants and shrubs pollinated by small insects (0.032 ± 0.03 SD). This difference between trees vs. non-tree species in our dataset suggests that small insect pollination does not result in larger \( S_P \) values in trees. In fact, we also found that differences between animal pollinators in their effect on plant \( S_P \) values are rather restricted to non-woody plants and shrubs (Fig. 3a). Although it is not clear why this is the case, we propose that, as in agaonid wasps, other small insects that pollinate trees in our dataset could also be transported by wind when they reach the canopy. This would result in large breeding areas for many small insect pollinated trees, corresponding to their observed small \( S_P \) values.

**Influence of latitudinal region on \( S_P \)**

We predicted that species from tropical and subtropical regions should have stronger SGS than species from temperate regions. We did in fact find that tropical species had greater \( S_P \) values than temperate species, however subtropical and temperate species did not differ from each other (Fig. 2b). In general, tropical regions have greater species richness and higher habitat heterogeneity at local scales [30,56], and this combination could be underlying the pattern of \( S_P \) variation we found. This is because such combination likely makes gene dispersal less effective at local scales, decreasing the spatial scale of intraspecific gene flow and thus increasing \( S_P \) values. For example, high species richness implies that conspecific individuals are potentially separated by interspecific ones [57], making cross-pollination and thus intraspecific gene flow
harder to achieve across long distances in the tropics. Furthermore, high habitat
dergogeneity at local scales in the tropics may result in tropical species and their
mutualists to be highly restricted to certain microclimates due to local adaptation
[58]. Such fine-scale narrow niches suggest that conspecific individuals should
become rapidly genetically isolated with increasing geographic distance,
associating with high $S_P$ values.

Differences among latitudinal regions, however, tend to be restricted to
non-woody plants, to a lesser extent to shrubs, and not apparent in trees (Fig.
3b). A similar pattern was reported in Dick et al. [5], where $S_P$ values were not
different between temperate and tropical trees. This result is in line with findings
showing that trees worldwide can have extensive breeding areas, thus high gene
flow among distant individuals, even in tropical regions where inbreeding has
been hypothesized to be prevalent [5,55,59]. Even if trees are very good at
dispersing their genes, either via pollen or seed, it is not clear why differences
between latitudinal regions affect other types of growth forms but not trees. The
mode of zoochory might be a more important determinant of SGS strength in
trees (see [4,5,60]), which we were not able to analyze in our dataset, precluding
us from finding a pattern of $S_P$ variation among trees.

**Influence of growth form on $S_P$**

Growth form in animal-pollinated plants was by far the most important
predictor of $S_P$ variation in our best-fit model, with $S_P$ values increasing from
trees to shrubs to non-woody plants (Fig. 2c). A similar pattern was reported by
Vekemans and Hardy [6], although they did not provide an explanation for it. This pattern may reflect the fact that larger plants will be higher in the canopy and thus better at dispersing genes, whether via pollen or seeds. The pattern may also simply reflect scale: smaller plants show more fine-grained dispersal and thus will have more fine-grained genetic structure. Furthermore, growth form is frequently tightly linked to habitat, in that non-woody plants and shrubs live in the understory while many trees reach the canopy. The understory may restrict gene flow more than the canopy, due to the lower dispersal propensity and the sedentary lifestyle of animal mutualists in the understory [61–63].

**Factors that did not influence $S_P$**

We did not find a significant effect of mating system on $S_P$ values in the animal-pollinated plant species included in our study. Mixed-mating plants tend to have higher $S_P$ values than outcrossing plants (Fig. S1d, 2d), but the difference between them was only marginally significant (Table 1). Selfing increases local genetic drift by reducing the effective number of reproductive individuals, which associates with higher $S_P$ values than outcrossing [6]. Moreover, gene dispersal in outcrossing plants occurs via pollen and seed dispersal, whereas gene dispersal in selfing plants is solely determined by seed dispersal, increasing $S_P$ values in selfing plants. We note that we did not include solely-selfing species in our analysis, thus the amounts of outcrossing in the mixed-mating species may have led to the only marginally significant effects of mating system that we detected.
We also failed to find an effect of seed dispersal mode on $S_P$ values either (Table 1, Fig S1e). However, we note that our classification of dispersal mode was somewhat coarse, in that we lumped together all zoochorous plants. Indeed, differences in foraging behavior among seed dispersing animals have previously been found to affect plant $S_P$: species with short-distance dispersers have greater Sp values than those dispersed by birds, while Sp values are highly variable in species dispersed by scatter-hoarding animals [4,60]. Our dataset included gravity dispersed plants, which should be the most dispersal limited, but surprisingly they were not associated with higher $S_P$ values. This is probably due to some animals (like ants and rodents) creating equally restricted seed dispersal patterns, and because some gravity-dispersed species might have unrecorded secondary seed vectors. Similarly, $S_P$ values for wind dispersal were highly variable in our study. Previous studies suggest that wind dispersal is often restricted [5,60], but our results suggest that wind does not have a predictable effect on gene dispersal.

Conclusions

Our results have important implications for understanding the origin and maintenance of biodiversity and can inform conservation strategies. For example, we found a general pattern in which genetic relatedness rapidly decreases with increasing geographic distance (i.e., high $S_P$ values) among tropical non-woody plants and shrubs pollinated by small insects. This suggests that such plants likely have more genetically isolated subpopulations than other animal-pollinated
plants. A recent review on global patterns of population genetic differentiation in seed plants supports this idea. Non-woody tropical species pollinated by small insects were associated with greater $F_{ST}$ values than other plants (D. Gamba & N. Muchhala, in review). Such genetic isolation at small to large spatial scales (i.e., within and among populations) could result in nearby subpopulations that harbor unique genetic diversity. This in turn, could increase the probability for local adaptation and reproductive isolation if divergent selection between close-by sites is strong and seed-mediated gene flow is ineffective. Non-woody/shrubby tropical species pollinated by small insects, nonetheless, are likely very susceptible to non-random habitat fragmentation (more so than vertebrate pollinated plants; e.g. [64]), which can further isolate populations and result in loss of genetic variability due to increased genetic drift [65,66]. The current scenario of human-accelerated change should thus push conservation efforts to maintain connectivity between fragments that harbor many understory tropical species pollinated by small insects.

**Acknowledgements**

We thank the researchers whose published data we used in this paper. We also thank Robert Ricklefs, Christine Edwards, and Carmen Ulloa for advice on study design. Many thanks to members of the Muchhala lab at the University of Missouri at Saint Louis for constructive discussions on a previous version of this manuscript. This research was supported by funds from the Whitney Harris World Ecology Center at the University of Missouri–Saint Louis.
References


54. Stevens PF. 2001 Angiosperm Phylogeny Website.


**Data accessibility statement**

Should the manuscript be accepted, the data and R scripts supporting the results will be archived in Dryad and their DOI will be included at the end of this article.

**Author Contributions**

DG and NM planned and designed the research. DG collected and analyzed the data. DG wrote the initial draft of the manuscript. DG and NM contributed equally to substantial revisions of the manuscript.
**Table 1** Details of the best-fit model explaining variation in $S_P$ values. Variables in bold indicate the reference level for each categorical factor. $N$ indicates the sample size of each group. Significant $p$-values are in bold. Model $R^2 = 0.24$, Model $AIC = -724.15$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>$t$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-2.58</td>
<td>0.21</td>
<td>-12.302</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Pollination mode</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Small insects</strong></td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large insects</td>
<td>38</td>
<td>-0.38</td>
<td>0.19</td>
<td>-1.97</td>
<td>0.05</td>
</tr>
<tr>
<td>Vertebrates</td>
<td>27</td>
<td>-0.50</td>
<td>0.21</td>
<td>-2.42</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Latitudinal region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tropics</strong></td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtropics</td>
<td>17</td>
<td>-0.70</td>
<td>0.27</td>
<td>-2.60</td>
<td>0.01</td>
</tr>
<tr>
<td>Temperate</td>
<td>33</td>
<td>-1.01</td>
<td>0.25</td>
<td>-4.12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Growth form</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-woody</strong></td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shrub</td>
<td>37</td>
<td>-0.45</td>
<td>0.19</td>
<td>-2.39</td>
<td>0.02</td>
</tr>
<tr>
<td>Tree</td>
<td>67</td>
<td>-1.26</td>
<td>0.23</td>
<td>-5.48</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Mating system</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mixed-mating</strong></td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcrossing</td>
<td>113</td>
<td>-0.26</td>
<td>0.21</td>
<td>-1.89</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Figure 1 Phylogeny of studied species showing the taxonomic extent of this study with plotted $S_P$ values in a logathmic scale, revealing their general lability across the phylogenetic tree. Plotting of $S_P$ values was achieved with the R package ‘phytools’ and the function ‘contMap’.
Figure 2 Marginal effects of factors on predicted $S_P$ values in the best-fit model: (a) pollination mode, (b) latitudinal region, (c) growth form, (d) mating system. Black dots are predicted $S_P$ means and surrounding bars correspond to ± one standard deviation. Significant differences between groups are depicted by letters on top of bar.
**Figure 3** Marginal effects conditional on growth form of predicted $S_P$ values for (a) animal pollination mode and (b) latitudinal region. Colors correspond to grouping categories (animal pollination modes or latitudinal regions). Each interaction was estimated as an additional term in the best-fit model. Dots in the plot are predicted $S_P$ means and surrounding bars correspond to ± one standard deviation.
Additional supporting information that will appear in the expanded online version of this article:

**Appendix S1.** References of publication with $S_p$ data and species traits used in this study.

**Fig. S1** Violin plots of $S_p$ values as a function of factors tested in this study.

**Table S1** Dataset used in this study (in file Table S1.xlsx).

**Table S2** Estimates of the generalized variance inflation factor on predictors.

**Table S3** Details of the most-inclusive model explaining variation in $S_p$ values.
Appendix S1. References of publication with $S_p$ data and species traits used in this study.


Melo, A. T. de O., and E. V. Franceschinelli. 2016. Gene flow and fine-scale spatial genetic structure in *Cabralea canjerana* (Meliaceae), a common


**Figure S1** Violin plots of $S_p$ values as a function of (a) pollination mode, (b) latitudinal region, (c) growth form, (d) mating system, (e) seed dispersal mode, and (f) genetic marker. Central black dots indicate the mean $S_p$ for each group, surrounding black dots are all observations. Thick horizontal grey lines are median values, boxes indicate 25% and 75% quartiles, and grey bars are minimum and maximum values. (Abbreviations: S-ins: small insects, L-ins: large insects, verts: vertebrates, mixed-m: mixed-mating, allo: allozymes, SSR: microsatellites) (next page).
**Table S1** Dataset used in this study (in file Table S1.xlsx).

**Table S2** Estimates of the generalized variance inflation factor (GVIF), and its adjusted value accounting for the degrees of freedom (GVIF^{1/(2*Df)}) for each factor in the most-inclusive model explaining variation in $S_P$ values.

<table>
<thead>
<tr>
<th>Variable</th>
<th>GVIF</th>
<th>Df</th>
<th>GVIF^{1/(2*Df)}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker</td>
<td>3.05</td>
<td>3</td>
<td>1.20</td>
</tr>
<tr>
<td>Mating system</td>
<td>1.79</td>
<td>2</td>
<td>1.16</td>
</tr>
<tr>
<td>Growth form</td>
<td>1.27</td>
<td>1</td>
<td>1.13</td>
</tr>
<tr>
<td>Animal pollination mode</td>
<td>2.88</td>
<td>2</td>
<td>1.30</td>
</tr>
<tr>
<td>Seed dispersal mode</td>
<td>2.28</td>
<td>2</td>
<td>1.23</td>
</tr>
<tr>
<td>Latitudinal region</td>
<td>1.94</td>
<td>2</td>
<td>1.18</td>
</tr>
</tbody>
</table>
Table S3 Details of the most-inclusive model explaining variation in $S_P$ values.

Variables in bold indicate the reference level for each categorical factor. N indicates the sample size of each group. Significant $p$-values are in bold. Model $R^2 = 0.26$, Model AIC = $-718.2$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-2.39</td>
<td>0.27</td>
<td>-8.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Genetic marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AFLP</strong></td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allozymes</td>
<td>32</td>
<td>-0.02</td>
<td>0.23</td>
<td>-0.08</td>
<td>0.94</td>
</tr>
<tr>
<td>Microsatellites</td>
<td>89</td>
<td>-0.29</td>
<td>0.21</td>
<td>-1.37</td>
<td>0.17</td>
</tr>
<tr>
<td>SNPs</td>
<td>7</td>
<td>-0.33</td>
<td>0.28</td>
<td>-1.21</td>
<td>0.23</td>
</tr>
<tr>
<td>Seed dispersal mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Animals</strong></td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gravity</td>
<td>24</td>
<td>-0.15</td>
<td>0.26</td>
<td>-0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>Wind</td>
<td>39</td>
<td>-0.08</td>
<td>0.19</td>
<td>-0.41</td>
<td>0.68</td>
</tr>
<tr>
<td>Pollination mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Small insects</strong></td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large insects</td>
<td>38</td>
<td>-0.34</td>
<td>0.21</td>
<td>-1.66</td>
<td>0.10</td>
</tr>
<tr>
<td>Vertebrates</td>
<td>27</td>
<td>-0.40</td>
<td>0.22</td>
<td>-1.82</td>
<td>0.07</td>
</tr>
<tr>
<td>Latitudinal region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tropics</strong></td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtropics</td>
<td>17</td>
<td>-0.72</td>
<td>0.28</td>
<td>-2.54</td>
<td>0.01</td>
</tr>
<tr>
<td>Temperate</td>
<td>33</td>
<td>-1.08</td>
<td>0.29</td>
<td>-3.76</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Growth form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-woody</strong></td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shrub</td>
<td>37</td>
<td>-0.54</td>
<td>0.22</td>
<td>-2.50</td>
<td>0.01</td>
</tr>
<tr>
<td>Tree</td>
<td>67</td>
<td>-1.25</td>
<td>0.25</td>
<td>-5.01</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mating system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mixed-mating</strong></td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcrossing</td>
<td>113</td>
<td>-0.19</td>
<td>0.15</td>
<td>-1.27</td>
<td>0.21</td>
</tr>
</tbody>
</table>