



Phylogeography of the Neotropical epiphytic orchid, *Brassavola nodosa*: evidence for a secondary contact zone in northwestern Costa Rica

Dorset W. Trapnell ¹ · J. L. Hamrick¹ · Patrick A. Smallwood¹ · Tyler R. Kartzinel ^{2,3} · Caitlin D. Ishibashi¹ · Charlotte T. C. Quigley¹

Received: 19 July 2018 / Revised: 13 February 2019 / Accepted: 10 March 2019 / Published online: 23 April 2019
© The Genetics Society 2019

Abstract

Spatial patterns of genetic variation can reveal otherwise cryptic evolutionary and landscape processes. In northwestern Costa Rica, an approximately concordant genetic discontinuity occurs among populations of several plant species. We conducted phylogeographic analyses of an epiphytic orchid, *Brassavola nodosa*, to test for genetic discontinuity and to explore its underlying causes. We genotyped 18 populations with 19 nuclear loci and two non-coding chloroplast sequence regions. We estimated genetic diversity and structure, relative importance of pollen and seed dispersal, and divergence time to understand how genetic diversity was spatially partitioned. Nuclear genetic diversity was high with little differentiation among populations ($G_{STn} = 0.065$). In contrast, chloroplast haplotypes were highly structured ($G_{STc} = 0.570$) and reveal a discontinuity between northwestern and southeastern populations within Costa Rica. Haplotype differences suggest two formerly isolated lineages that diverged ~10,000–100,000 YBP. Haplotype mixing and greater genetic diversity occur in an intermediate transition zone. Patterns of nuclear and chloroplast data were consistent. Different levels of genetic differentiation for the two genomes reflect the relative effectiveness of biotic versus abiotic dispersers of pollen and seeds, respectively. Isolation of the two lineages likely resulted from the complex environmental and geophysical history of the region. Our results suggest a recent cryptic seed dispersal barrier and/or zone of secondary contact. We hypothesize that powerful northeasterly trade winds hinder movement of wind-borne seeds between the two regions, while the multi-directional dispersal of pollen by strong-flying sphinx moths resulted in lower differentiation of nuclear loci.

Introduction

Historical factors shape contemporary spatial genetic patterns within species, and may give rise to secondary contact of previously isolated conspecific lineages. In northwestern

Costa Rica several species of seasonal dry forest plant species, including two epiphytic orchids, have an intriguing intraspecific genetic discontinuity in a region devoid of obvious ecological or landscape barriers to gene exchange (Cavers et al. 2003; Trapnell and Hamrick 2004; Fuchs and Hamrick 2010; Kartzinel et al. 2013; Poelchau and Hamrick 2013). Congruent genetic discontinuities in multiple plant taxa suggest a shared response to historical environmental factors. Comparative phylogeographic analyses often provide insights into the influence of historical processes on regional biodiversity. Genetic discontinuities in orchids (e.g., *Laelia rubescens* [Trapnell and Hamrick 2004] and *Epidendrum firmum* [Kartzinel et al. 2013]) are surprising considering the seed dispersal potential of epiphytic orchids and genetic evidence for long-distance seed movement in *L. rubescens*. An unusual characteristic of orchids is their production of up to six million dust-like seeds/fruit (Arditti and Ghani 2000) that are potentially wind-dispersed over considerable distances (e.g., Docters van Leeuwen 1936).

Supplementary information The online version of this article (<https://doi.org/10.1038/s41437-019-0218-y>) contains supplementary material, which is available to authorized users.

✉ Dorset W. Trapnell
dorset@uga.edu

¹ Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

² Odum School of Ecology, University of Georgia, Athens, GA 30602, USA

³ Present address: Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, USA

Arboreal orchids, have the advantage of releasing their seeds higher in the air column where they can be picked up by wind currents (DWT, pers. obs.), increasing the probability of long-distance gene movement. Furthermore, *L. rubescens* and *E. firmum* release their seeds during the dry season when powerful northeasterly trade winds prevail (DWT, pers. obs.; Kartzinell et al. 2013). Thus, seed-mediated gene flow has the potential to genetically homogenize populations across landscapes and to erase genetic footprints of colonization and historical isolation.

Brassavola nodosa (Linnaeus) Lindley (Orchidaceae), is an epiphytic orchid that often grows sympatrically with *L. rubescens* in lowland seasonal dry forests of northwestern Costa Rica. An important difference between these species is that *L. rubescens* is hummingbird pollinated (Trochilidae; Trapnell and Hamrick 2005) while *B. nodosa* is pollinated by sphinx moths that can travel much greater distances (Sphingidae; Gregory 1964; Dressler 1981; Janzen 1983; Haber and Frankie 1989; Chapman et al. 2010). Pollen dispersal distance is, therefore, influenced by the pollinator's foraging range and/or territorial boundaries, while maternally inherited plastid genes (Mogensen 1996) are dispersed by wind-borne seeds that are potentially dispersed more broadly (e.g., Hamrick and Trapnell 2011; Trapnell and Hamrick 2004, 2005). As a result, neutral plastid and nuclear markers allow detection of phylogeographic patterns within contemporary populations and allow inference of historical events such as dispersal, colonization, isolation, and range expansion.

Our objectives are to (i) estimate levels of nuclear and chloroplast genetic diversity, and the relative importance of pollen and seed dispersal for gene flow at multiple spatial scales; (ii) investigate the spatial distribution of nuclear and plastid genetic variation in *B. nodosa* and test whether this species, in which both pollen and seeds can potentially be dispersed long distances, experiences the genetic discontinuity seen in other plant taxa (e.g., Cavers et al. 2003; Trapnell and Hamrick 2004; Fuchs and Hamrick 2010; Poelchau and Hamrick 2013; Kartzinell et al. 2013); and (iii) use nuclear and chloroplast genetic data to develop hypotheses regarding the underlying cause of this genetic discontinuity in the absence of obvious dispersal barriers.

Materials and methods

Study species and sampling

Brassavola nodosa is an orchid in subfamily Epidendroideae. This long-lived perennial occurs at 0–600 m above sea level (asl) in seasonal dry forest, as well as in the salt spray zone of coastal mangrove forests (Janzen 1983; DWT, pers. obs.). *Brassavola nodosa* occurs on many tree species and can be

locally common (DWT, pers. obs.). Although *B. nodosa* is self-compatible (Schemske 1980), its highly fragrant, greenish white flowers attract night-pollinating sphinx moths (Sphingidae; Dressler 1981), which in Costa Rica transport pollen up to 10–15 km (Janzen 1983). Wind-borne seeds are released during the dry season (December to April; DWT, pers. obs.) when trade winds are the strongest, blowing from the northeast at a mean speed of ~18 km/hr (1999–2012; Estación Mecánica Aeropuerto Liberia, Instituto Meteorológico Nacional) and occasionally reaching speeds > 90 km/hr (<http://www.worldheadquarters.com/cr/climate/>; <http://www.myweather2.com/City-Town/Costa-Rica/Guanacaste/climate-profile.aspx>).

We collected leaf tissue from 28–159 individuals from each of 18 populations (mean = 68.7; total = 1237) throughout northwestern Costa Rica (Table 1; Fig. 1). Study populations occurred at low elevations (range = 1–292 m asl) and are separated by a mean of 38.8 km (range = 0.2–99.7 km). Each population consisted of orchids growing on one to twelve trees (mean = 4). Leaves were snap frozen in liquid nitrogen and transported to The University of Georgia for genetic analyses.

Chloroplast genotyping and analyses

We analyzed 21–75 individuals/population (mean = 35.3) for chloroplast sequence diversity using total genomic DNA extracted from allozyme wicks or isolated from frozen leaf tissue using a CTAB protocol (Doyle and Doyle 1987). Two non-coding regions of the chloroplast (the rpl32-trnL intergenic spacer and the trnG intron) were amplified according to Trapnell et al. (2013).

We assessed nucleotide diversity (π), number of segregating sites (S), and Watterson's θ without indels using *pegas* v.0.5 (Paradis 2010) in R v.3.0.2 (R Core Development Team 2013). The effective number of haplotypes (HA_e) and haplotype diversity (HH_e) were estimated at population and species levels. Treating each haplotype as a single allele, partitioning of haplotype variation among populations was estimated with G_{STc} (Nei 1973), AMOVA (GenAlEx 6.503; Peakall and Smouse 2006), and SAMOVA 2.0 (Dupanloup et al. 2002). We calculated G_{STc} and pairwise G_{STc} values using F_{STAT} (Goudet 2001). Sequences were prepared for analysis using Fluxus' DNA Alignment software (<http://www.fluxus-engineering.com>) with all polymorphisms weighted equally and a haplotype network was generated using the Median-Joining method in NETWORK 4.6.1.2 (Bandelt et al. 1999).

Nuclear genotyping and analyses

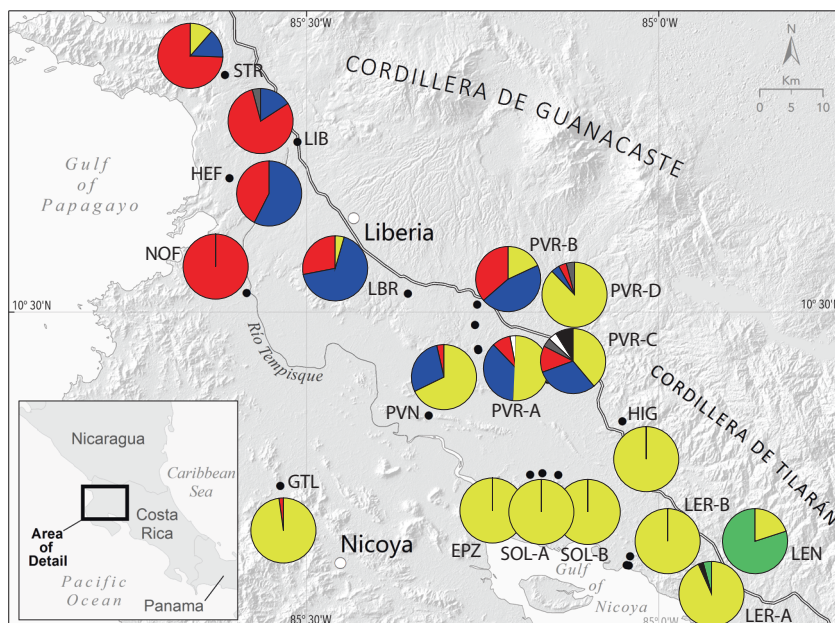
We analyzed all samples with 19 allozyme loci (Trapnell et al. 2013), 18 of which were polymorphic, yielding two to

Table 1 *Brassavola nodosa* population locations in Costa Rica arranged from north to south, within three regions, and summary statistics of chloroplast diversity

Population	Latitude	Longitude	<i>N</i>	<i>H</i>	HA_e	HH_e	π	<i>S</i>	θ
Northwest (NW)									
STR	10.8346	-85.6134	43	3	1.70	0.413	0.0015	9	2.08
LIB	10.7415	-85.5129	25	3	1.50	0.333	0.0000	0	0.00
HEF	10.6899	-85.6093	26	2	1.95	0.488	0.0000	0	0.00
NOF	10.5270	-85.5850	46	1	1.00	0.000	0.0000	0	0.00
LBR	10.5260	-85.3558	43	3	1.87	0.465	0.0006	9	2.08
NW Mean	—	—	36.6	2.4	1.60	0.340	0.0004	3.6	0.83
Transition Zone (TZ)									
PVR-B	10.5103	-85.2575	33	3	2.69	0.628	0.0022	9	2.22
PVR-D	10.4815	-85.2605	25	4	1.28	0.221	0.0010	9	2.38
PVR-A	10.4475	-85.2565	33	4	2.46	0.593	0.0036	9	2.22
PVR-C	10.4458	-85.2560	23	6	3.65	0.726	0.0037	9	2.44
PVN	10.3530	-85.3263	28	3	1.84	0.457	0.0032	9	2.31
TZ Mean	—	—	28.4	4.0	2.38	0.525	0.0027	9.0	2.31
Southeast (SE)									
HIG	10.3447	-85.0513	33	1	1.00	0.000	0.0000	0	0.00
SOL-A	10.2709	-85.1649	22	1	1.00	0.000	0.0000	0	0.00
EPZ	10.2688	-85.1823	39	1	1.00	0.000	0.0000	0	0.00
SOL-B	10.2685	-85.1424	21	1	1.00	0.000	0.0000	0	0.00
GTL	10.2530	-85.5373	51	2	1.04	0.038	0.0003	9	2.00
LER-B	10.1533	-85.0403	44	1	1.00	0.000	0.0000	0	0.00
LEN	10.1407	-85.0451	25	2	1.47	0.320	0.0000	0	0.00
LER-A	10.1400	-85.0423	75	3	1.14	0.127	0.0000	0	0.00
SE mean	—	—	38.8	1.5	1.08	0.061	0.0000	1.1	0.25
Overall pop mean	—	—	35.3	2.4	1.59	0.267	0.0009	4.0	0.99
Species level	—	—	635	10	2.48	0.597	0.0030	9.0	1.28

N sample size, *H* number of haplotypes, HA_e effective number of haplotypes, HH_e expected haplotype diversity, π nucleotide diversity, *S* number of segregating sites, Watterson's θ

Fig. 1 Distribution and frequencies of chloroplast haplotypes among eighteen *Brassavola nodosa* populations in Costa Rica. Color designations for the haplotypes are H1 yellow, H2 blue, H3 red, H4 white, and H8 green. Private haplotypes confined to a single population are represented by black (H5 [PVR-C], H6 [PVR-C], H7 [PVR-D], H9 [LIB], and H10 [LER-A])



six alleles. Nuclear (nDNA) genetic diversity statistics were estimated using LYNPROG, designed by M.D. Loveless and A.F. Schnabel. Measures of genetic diversity were percent polymorphic loci, P ; mean number of alleles/polymorphic locus, AP ; observed heterozygosity, H_o ; and genetic diversity, H_e (proportion of loci heterozygous per individual under Hardy–Weinberg expectations; Nei 1973). Species level values for these parameters were calculated by pooling data from all populations.

Population values were calculated and then averaged across all populations. Heterogeneity of allele frequencies among populations was tested using χ^2 (Workman and Niswander 1970). For each polymorphic locus in each population, observed (H_o) and Hardy–Weinberg expected heterozygosity (H_e) were compared using Wright's inbreeding coefficient (F_{IS} ; Wright 1922, 1951). Deviations from Hardy–Weinberg expectations were tested for significance using χ^2 (Li and Horvitz 1953). The Bonferroni correction for multiple comparisons was applied using FSTAT (Goudet 2001). Allelic diversity was determined both as the total number of alleles observed (A_T) and as the number of alleles adjusted for population sample size (AR ; i.e., rarefaction) using FSTAT (Goudet 2001). For the rarefaction analysis (i) only polymorphic loci were used, (ii) all populations were treated as consisting of 28 individuals (i.e., 56 alleles) to reflect the size of the smallest population (SOL-B), and (iii) four loci (*Pgm-1*, *Pgm-2*, *Aat-2*, and *6Pgd-1*) were omitted due to missing data from some populations. A UPGMA phenogram of genetic identity values (Nei 1972) was generated using NTSYS-pc 2.1 (Rohlf 2002).

Genetic diversity among populations was partitioned using Nei's (1973) G_{STn} and AMOVA (GENALEX 6.503; Peakall and Smouse 2006). Pairwise G_{ST} values were calculated using FSTAT (Goudet 2001) using only polymorphic loci and with *Pgm-1*, *Pgm-2*, *Aat-2*, and *6Pgd-1* omitted because of missing data from some populations. A Bayesian clustering approach (STRUCTURE 2.3.4; Pritchard et al. 2000) was used to estimate genetic admixture among populations and the number of genetically distinct clusters (K). Ten independent Markov chain Monte Carlo (MCMC) simulations at each K value ranging from $K = 1$ to 18 were performed, with a burn-in of 50,000 and a run length of 50,000. A model based on correlated allele frequencies and admixture was selected. The most likely K value was identified using criterion for maximizing ΔK according to Evanno et al. (2005). Genetic admixture among populations was also analyzed using MAVERICK 1.0 (Verity and Nichols 2016), which utilizes the same mixture modeling framework as STRUCTURE but employs thermodynamic integration to estimate K . Values of K from 1 to 18 were tested in 10 independent runs using the admixture model, a burn in of 50,000 and 50,000 MCMC steps. The posterior probability and log-evidence for each possible K were used to identify the most likely value.

Isolation by distance (IBD) was estimated by plotting $G_{STn}/(1-G_{STn})$ (Rousset 1997) for each pair of populations against their geographic distance using IBDWS version 3.23 (Jensen et al. 2005). Isolation by environment (IBE) was also examined to test the relative importance of environmental versus geographic isolation on genetic differentiation among populations. We analyzed nuclear genetic data within the context of 19 bioclimatic variables (S1 Table; WorldClim v. 2.0; www.worldclim.org; 2.5 arc minute resolution), as well as geographic location data in the R package *Sunder* (Botta et al. 2015; R Core Development Team 2013). This model-based approach assesses the ecological influence, independent of geographic distance, on genetic distance among populations. We used initial values of $\alpha = 1.0$, $\beta_D = 2.0$, $\beta_E = 1.0$, $\gamma = 0.0$, and $\delta = 0.01$. During the model selection process, parameter values were: $\alpha \in [0, 10]$, $\beta_D \in (0, 996.5]$, $\beta_E \in (0, 2010.0]$, $\gamma \in [0, 1.0]$, and $\delta \in [0, 1)$. The MCMC walk consisted of 50,000 iterations, with a burn in of 1000.

A weakness of *Sunder* is that even in the absence of IBE and IBD it selects one of three proposed models where allelic covariance is a function of (i) environmental distance, (ii) geographic distance, or (iii) environmental and geographic distances. To test whether model selection was random, we performed 60 independent runs for each bioclimatic variable and tested whether the distribution of models selected was non-random, using the *xmulti* function from R package *XNomial* (Engels 2015).

Pollen vs. seed-mediated gene movement

The relative contribution of pollen (m_p) and seed dispersal (m_s) to overall gene flow was estimated by,

$$m_p/m_s = [(1/G_{STn} - 1)(1 + F_{IS}) - 2(1/G_{STc} - 1)] / (1/G_{STc} - 1)$$

where G_{STn} and G_{STc} are among population nDNA and cpDNA differentiation, respectively (Petit 1992; Petit et al. 1992; Ennos 1994). When pollen and seed migration rates are equivalent in an outcrossing hermaphrodite with strict maternal chloroplast inheritance, as G_{STn} approaches zero haplotypes should have a smaller effective population size and three times as much fixation as nuclear alleles (Hamilton and Miller 2002). The expected difference between chloroplast and nuclear G_{ST} values was calculated by $G_{STc} = 6 G_{STn} / [2 + 4 G_{STn}]$ under the null hypothesis of equal pollen and seed flow (Hamilton and Miller 2002). Significance of the difference between expected G_{STn} , and observed G_{STn} values was tested with a 95% confidence interval generated by bootstrapping procedures.

Genetic structure relative to wind patterns

Average monthly wind speed data for January–April from 2012 to 2017 was obtained from TerraClimate (Abatzoglou

et al. 2018) at a resolution where 1 pixel represents 4 km². Mean wind speed across years was calculated for each pixel across the range of study populations. Daily wind direction data was also obtained for the period between 1 January 2012 and 30 April 2017 from the NOAA/NCEP Global Forecast System Atmospheric Model using the R package *rWind* (Fernández-López and Schlieo 2018). Wind direction data were formatted to the same resolution as the wind speed data (4 km²) using the *resample* function from the raster R package (Hijmans 2018). Wind direction for each pixel was averaged for each month across the six years using the *mean.circular* function in the circular package in R (Agostinelli and Lund 2017).

Lines connecting pairs of populations were drawn using the *SpatialLines* (SL) function in R and the bearing of each SL was determined using *bearingRhumb* from the geosphere R package, which avoids the problem of a bearing changing along a line (i.e., rhumb line) on a curved surface (Hijmans 2017). The *extract* function from the raster R package was used to determine the wind direction for a given month in each pixel crossed by a SL.

To find the smallest angle between the SL bearing and mean wind direction (angle range) per pixel per month for population pairs, the *range.circular* function from the circular R package was used. The *dist.circular* function from the circular R package was then used to convert angular ranges between pairs of populations to a distance metric. Significance in the relationship between angular range distance and both nuclear (G_{STn}) and chloroplast (G_{STc}) genetic distances, were measured using a Mantel test with 323 permutations.

Divergence time estimates

Divergence time (T_{div}) between northwestern and southeastern regions identified with chloroplast data (see Results) was estimated using the Bayesian coalescent approach implemented in IMA2 (Hey 2010). Our goal was to estimate population splitting time (t). IMA2 also models the rate of mutation-scaled migration to each population (m_1 , m_2) and population mutation rate (θ), taking effective population size into account for current (θ_1 , θ_2) and ancestral populations (θ_A) (Hey and Nielsen 2004). Several short runs were conducted to establish suitable priors and evaluate parameters for MCMC sampling. Prior upper limits were $\theta = 20$, $m = 10$, and $t = 6$ (~95% high posterior density [HPD] estimate of t_{MRC} based on preliminary runs). Three runs were conducted with different random starting seeds. Each run had MCMCs with 50 chains and heating parameters of 0.95 and 0.80. Burn in periods of 10^6 generations were followed by 10^8 steps in which every 10 iterations were sampled. We randomly selected 100 individuals (50 northwestern and 50 southeastern) for our IMA2 analyses.

Appropriate chain mixing, low autocorrelation over the course of each run, high effective sample sizes, and convergence on similar parameter estimates between independent runs were observed, so runs were combined in an L mode run. Divergence time was estimated from the peak posterior density ($\pm 95\%$ HPD) according to equation $T_{div} = t/\mu k$, where μ is the mutation rate (substitutions/site/year) and k is sequence length. Because no independent fossil or historical evidence is available for calibration, we estimated T_{div} based on both slow (1×10^{-9} substitutions/site/year (s/s/y); Wolfe et al. 1987) and fast (8.24×10^{-9} s/s/y; Richardson et al. 2001) chloroplast mutation rates.

Ecological niche modeling

We combined *B. nodosa* location information from our samples as well as occurrence data from herbarium specimens and the Southeast Regional Network of Expertise and Collections (<http://sernecportal.org/portal/collections/index.php>) to generate a presence-only database of 62 populations. Species distribution models were created using maximum entropy in MAXENT 3.1.0 (Phillips et al. 2006; Phillips and Dudik 2008) for the present day, the Holocene (~6000 YBP) and the last glacial maximum (LGM; ~22,000 YBP). We used the default convergence threshold (10^{-5}) and maximum number of iterations (500). Jackknifing was used to estimate variable importance. Fifteen of 19 bioclimatic variables from the Worldclim global climate database, with 30 arc second spatial resolution, were used as predictors of suitable habitat (Hijmans et al. 2005). Varela et al. (2015) suggest removing four variables (BIO2, BIO3, BIO14, and BIO15; S1 Table) for construction of models projected onto previous time periods (i.e., Holocene and LGM) due to high variance of inferred values among earth system models. Model goodness-of-fit was evaluated using the area under the receiver operating characteristics curve (AUC).

MAXENT can only use one occurrence record per grid cell (~1 km²) of climatic layers in BIOCLIM, consequently usable occurrence records were reduced to 28. Because of the reduced number of localities, we did not withhold a training set for model testing. However, preliminary runs with 20% of the localities withheld for model testing resulted in AUC graphs with virtually identical training and testing trend lines, suggesting that the model was effective for inferring suitable habitat.

The resulting model was projected onto rasters of bioclimatic variables representing the Holocene and LGM as inferred by the Community Climate Systems Model ver. 4.0 (CCSM4; Gent et al. 2011), the Model for Interdisciplinary Research on Climate (MIROC; Watanabe et al. 2011), and the Max Planck Institute for Meteorology Earth Systems Model (MPI-ESM; Giorgetta et al. 2013). Because the three

models showed considerable variation, we calculated the probability of existence of suitable habitat per cell by averaging the inferred probability of suitable habitat at that position across the three models (CCMS4, MPI-ESM, MIROC) for the Holocene and LGM (Varela et al. 2015). MAXENT produces a continuous probability value, based on cumulative values, as an estimate of relative habitat suitability for the species.

Results

Chloroplast genetic data

Ten haplotypes were identified, with each population having one to six haplotypes (mean = 2.4; Table 1; Fig. 1). Three haplotypes (H1, H2, H3) characterized 95% of the individuals. Five private (i.e., found in a single population) haplotypes (H5, H6, H7, H9, H10) occurred in four populations (LIB = 1, PVR-C = 2, PVR-D = 1, LER-A = 1); each haplotype occurred in only one or two individuals (Figs. 1, 2). An additional haplotype, H4, was documented in one individual from each of two populations (PVR-A and PVR-C) separated by ~192 m. The final haplotype, H8, was found in two populations separated by ~316 m (20 individuals in LEN and 3 in LER-A). Mean population and species level diversity (HH_e) were 0.267 and 0.597, respectively. For the eight populations with segregating chloroplast sequence sites, Watterson's θ ranged from 2.00 to 2.44 (Table 1).

Chloroplast haplotypes showed distinctive spatial partitioning. Five northwestern (NW) populations (STR, LIB, HEF, NOF and LBR) possessed four haplotypes, of which two were dominant (H2, H3) and one was unique (H9) to the region. Eight southeastern (SE) populations (HIG, SOL-A, SOL-B, EPZ, GTL, LEN, LER-A, and LER-B) also had four haplotypes, with one (H1) being dominant and two being unique (H8, H10) to the region. Haplotype H8 was predominantly located in a mangrove population (LEN), but also in two individuals in LER-A, a nearby upland site. Five

populations geographically located between the NW and SE regions (PVR-A, PVR-B, PVR-C, PVR-D, and PVN) contained the highest chloroplast diversity: seven haplotypes, four of which were unique to this zone (H4, H5, H6, H7; Table 1; Fig. 1). This geographically and genetically intermediate, transition zone (TZ) had the highest cpDNA diversity with 70% of the haplotypes compared to 40% in the NW and 50% in the SE. It is noteworthy that two NW populations (STR and LBR) contained low frequencies of H1 (the dominant SE haplotype).

The three chloroplast haplotypes that characterize the NW region (H2, H3, H9) differed from one another by 1–2 mutational steps (Fig. 2). Likewise, three SE haplotypes (H1, H8, H10) differed by 1–2 mutational steps. However, the three NW haplotypes differed from the dominant (H1) and rare (H8, H10) SE haplotypes as well as from two private TZ haplotypes (H6, H7) by a minimum of 13 mutational steps (Fig. 2). The other two private TZ haplotypes (H4, H5) were more similar to haplotypes in the NW (separated by 4–6 mutational steps) than the SE (8–12 mutational steps).

Because TZ populations contain a mixture of haplotypes primarily found in the NW and the SE, and assignment of TZ populations to either the NW or SE regions would by necessity be arbitrary, TZ was treated as a third region in subsequent analyses. Designation of three regions rather than two yielded the same insights in subsequent analyses, but designation of three zones permitted greater clarity.

AMOVA revealed that 60% of cpDNA sequence diversity was partitioned among the 18 populations. Of this variation, 80% (0.48) was structured among the three regions and 20% (0.12) was partitioned among populations within regions. Nei's (1973) G_{STc} yielded similar results; partitioning of haplotype diversity among all populations was 0.570, of which 79% (0.449) was among the three regions and 21% (0.121) was among populations within regions (S2 Table). Additional support for genetic differences among regions comes from the pairwise G_{STc} data (S3 Table). Mean pairwise G_{STc} values among populations within the NW, TZ, and SE regions were 0.290, 0.142, and

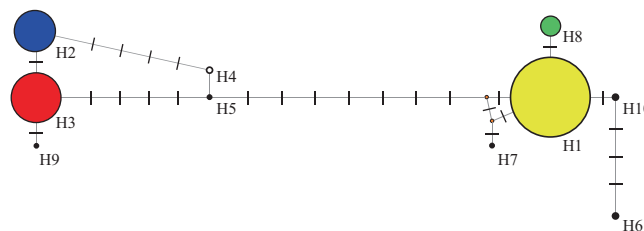


Fig. 2 Network of chloroplast haplotypes in Costa Rican populations of *Brassavola nodosa* where each circle corresponds to a unique haplotype and circle sizes represent relative frequencies. Colors agree with those used in Fig. 1. The five private haplotypes (H5, H6, H7, H9,

H10) are shown in black. Vertical notches on the lines connecting haplotypes indicate mutational differences between haplotypes. The two small orange circles between H1 and H7 indicate median vectors (i.e., missing intermediates)

0.202, respectively (S3 Table). Mean pairwise G_{STc} values between regions were considerably larger: NW and SE (0.790), NW and TZ (0.433), and TZ and SE (0.357). The G_{STc} within SE was primarily influenced by the dominance of rare haplotype H8 in LEN, the mangrove population, while the remaining seven populations were homogenous. Jackknifing procedures (Weir 1996) revealed a significant difference in haplotype diversity values between pairwise comparisons of the regions ($df = 6$, $P < 0.01$). SAMOVA produced congruent results, lending support for two groups; one includes all NW populations, as well as PVR-B while the second group contains all the SE populations as well as four TZ populations.

Nuclear genetic data

Nineteen allozyme loci were resolved, yielding 72 alleles (Table 2). Ten unique alleles were observed in eight populations: three in PVR-A and one in each of seven populations (HEF, PVN, HIG, SOL-A, LEN, LER-A, LER-B). The NW, TZ, and SE regions had 57, 59, and 64 alleles, respectively, and one, four, and five private alleles, respectively. Mean within-population genetic diversity was moderately high as was the pooled species-wide sample (i.e., across all populations; Table 2). Tests revealed significant heterogeneity in allele frequencies among populations for 17 of 18 polymorphic loci ($df = 34-102$ per locus, $P < 0.005$). Mean F_{IS} across polymorphic loci was 0.004, indicating a slight heterozygote deficiency relative to Hardy–Weinberg expectations (HWE). After Bonferroni correction, only 2.2% (6 out of 271) of F_{IS} estimates differed significantly from HWE (adjusted $P < 0.00015$).

Nei's (1972) genetic identity values (I) between pairs of populations ranged from 0.886 (LIB and GTL) to 0.993 (LIB and HEF), with mean $\bar{I} = 0.958$. GTL had the lowest mean genetic identity (0.928) with the other 17 populations while LBR and LEN had the highest mean values (0.972 and 0.970, respectively). The UPGMA phenogram (S1 Figure) failed to show a pattern of genetic similarity that corresponded to the spatial distribution of the populations.

At the broadest geographic scale, nDNA genetic structure among the 18 populations (G_{STn}) was 0.065 (S2 Table). Of that, 0.012 (18.5%) was among NW, TZ, and SE while $G_{STn} = 0.053$ (81.5%) resulted from variation among populations within regions. AMOVA yielded a similar result, with 0.050 of genetic variation partitioned among the 18 populations. Mean pairwise G_{STn} values among all populations ranged from 0.042 (PVR-B) to 0.090 (LIB) (S4 Table); there was little difference between the overall mean pairwise values (0.064) and the mean pairwise estimates within NW, TZ, and SE (0.074, 0.042, and 0.072, respectively) (S3 Table).

Although the results of the nDNA analyses were constrained by the low overall among population differentiation

Table 2 Summary of nuclear genetic diversity values in Costa Rican populations of *Brassavola nodosa*

Pop	N	A_T	AR	P	AP	H_o	H_e	F_{IS}
Northwest (NW)								
STR	52	36	27	63.2	2.42	0.210	0.227	0.075
LIB	48	44	31	79.0	2.67	0.196	0.209	0.062
HEF	88	54	36	89.5	3.06	0.245	0.246	0.004
NOF	51	41	30	88.9	2.31	0.227	0.203	-0.118
LBR	55	46	33	84.2	2.69	0.256	0.292	0.123
NW Mean	58.8	44.2	31.4	80.9	2.63	0.227	0.235	0.029
Transition zone (TZ)								
PVR-B	159	45	34	73.7	2.86	0.231	0.222	-0.041
PVR-D	67	47	37	73.7	3.00	0.284	0.277	-0.025
PVR-A	152	55	37	89.5	3.12	0.248	0.257	0.035
PVR-C	120	54	38	89.5	3.06	0.236	0.246	0.041
PVN	38	41	30	73.7	2.57	0.259	0.273	0.051
TZ Mean	107.2	48.4	35.1	80.0	2.92	0.252	0.255	0.012
Southeast (SE)								
HIG	50	51	37	84.2	3.00	0.319	0.332	0.039
SOL-A	50	45	33	84.2	2.63	0.264	0.277	0.047
EPZ	43	41	31	73.7	2.57	0.254	0.247	-0.028
SOL-B	28	40	29	68.4	2.62	0.283	0.251	-0.127
GTL	55	43	35	79.0	2.60	0.259	0.258	-0.004
LER-B	47	43	30	79.0	2.60	0.283	0.273	-0.037
LEN	55	44	33	79.0	2.67	0.283	0.277	-0.022
LER-A	79	47	32	79.0	2.87	0.255	0.252	-0.012
SE mean	50.9	44.2	32.5	78.3	2.69	0.275	0.271	-0.018
Overall pop mean	68.7	45.4	32.9	79.5	2.74	0.255	0.257	0.004
Species level	1237	72	—	94.7	3.94	—	0.284	0.008

All F_{IS} values are non-significant. AR was based on rarefaction analysis where $N = 28$ individuals. All estimates are based on 19 loci except AR , which is based on 15 loci

N sample size, A_T total number of alleles, AR allelic richness, P percent polymorphic loci, AP alleles per polymorphic locus, H_o observed heterozygosity, H_e expected heterozygosity, F_{IS} inbreeding coefficient

($G_{STn} = 0.065$), examination of the pairwise G_{STn} matrix indicated that the nDNA results were similar to the cpDNA results (S2, S3 and S4 Tables). Both genomes had higher genetic differentiation between the NW and SE regions than either region had with TZ. Also mean pairwise G_{STn} and G_{STc} values among populations within TZ were lower than the mean pairwise G_{STn} and G_{STc} values among populations within NW and SE (S3 Table). Finally, two populations (STR and LBR) from NW with haplotype H1 had lower mean pairwise G_{STn} values with populations of TZ ($G_{STn} = 0.040$ and 0.033, respectively) than they had with SE populations (0.077 and 0.044) and the other three NW populations (0.075 and 0.073). Furthermore, STR and LBR were genetically similar to each other ($G_{STn} = 0.030$). These results are consistent with the IBD analyses, which indicated that there was a non-significant ($r^2 = 0.013$, $df = 152$, $p = 0.097$) trend towards higher genetic differentiation (i.e., G_{STn}) between populations separated by greater geographic distances.

IBE analyses did not yield consistent results for each of the 19 bioclimatic variables. The models selected in 60 independent runs for each bioclimatic variable did not deviate significantly from random, with Pr (observed model counts | random model selection) ranging from 0.092 to 0.986. This suggests that there is no significant IBE or IBD. This is consistent with the absence of significant IBD found using IBDWS. Furthermore, the absence of IBE is unsurprising given the small magnitude of differences in the bioclimatic variables across populations (S1 Table).

Structuring of individual nDNA multi-locus genotypes using STRUCTURE was equivocal. The distribution of $LnP(D) = L(K)$ values failed to show a discontinuity before plateauing, while maximization of ΔK using the Evanno et al. (2005) approach yielded 13.33 for $K = 2$ (S2 Figure). For all other K values, ΔK ranged from 0.44 to 7.05. MAVERICK also revealed that $K = 2$ had the highest posterior likelihood. The two genetic clusters identified by STRUCTURE and MAVERICK, as well as the relative Q value cluster assignments, were virtually identical. Populations LIB, PVR-A, PVR-B, PVR-C, PVR-D, HEF, and LEN were in one cluster, with assignment values > 0.5 , while the second cluster included STR, SOL-A, SOL-B, LBR, NOF, EPZ, HIG, PVN, LER-A, LER-B, GTL. Clusters do not correspond with the delineation between regions revealed by the chloroplast data and have little correspondence with the geographic proximity of populations.

Pollen vs. seed-mediated gene movement

At all spatial scales pollen-mediated gene flow contributed significantly ($P < 0.05$) more than seed dispersal to overall gene movement. Among the 18 populations, spanning distances of up to 99.7 km, the ratio of pollen to seed movement (m_p/m_s) was 17.0 (Table 3). The ratio was substantially higher (116.1) between the NW and SE regions. Lower m_p/m_s ratios were found between the NW and TZ (87.3) and between the SE and TZ (9.8).

Genetic structure relative to wind patterns

Mean wind speeds ranged from 2.52 m/s to 5.67 m/s during the dry season when *B. nodosa* seeds are released (S5 Table; S3 Figure). Furthermore, the wind consistently blows from the northeast to the southwest during this period. Wind intensity declines in April before the start of the rainy season. The Mantel test showed a highly significant relationship between haplotype structure (G_{STc}) and angular ranges for all four months (January [$p = 0.003$], February [$p = 0.003$], March [$p = 0.003$], and April [$p = 0.006$]). Genetic divergence between neighboring pairs of populations separated by 10–30 km and lying perpendicular to the trade winds was substantially higher than for population

Table 3 Summary of the mean pairwise G_{STn} values among *Brassavola nodosa* populations and regions (northwest [NW], transition zone [TZ], and southeast [SE]) within Costa Rica, and the ratio of pollen movement to seed movement (m_p/m_s)

	Mean pairwise G_{STn}	m_p/m_s
Among 18 populations throughout Costa Rica	0.064	17.0 ^a
Between NW and SE regions	0.076	116.1 ^a
Between NW and TZ regions	0.052	87.3 ^a
Between TZ and SE regions	0.057	9.8 ^a
Among populations within NW	0.074	4.1 ^a
Among populations within TZ	0.042	2.1
Among populations within SE	0.072	28.6 ^a

All G_{STn} and G_{STc} values are significant ($p = 0.001$).

^a95% confidence intervals generated by bootstrapping indicate that m_p/m_s differs significantly from 1.0 (Hamilton and Miller 2002)

pairs positioned parallel to the winds. Within NW, the mean G_{STc} of perpendicular pairs is 367% higher than parallel pairs of populations (0.284 versus 0.078). In SE, the mean G_{STc} of perpendicular pairs is 118% higher than parallel pairs of populations. In TZ, the parallel pairs G_{STc} was 0.034 but there were no perpendicular pairs available for comparison. For all comparisons, the mean G_{STc} of neighboring pairs of populations that are oriented perpendicular to the trade winds is 283% greater than the mean for pairs parallel to the prevailing wind direction. However, the relationship between nuclear genetic structure (G_{STn}) and angular ranges was non-significant for all months (January [$p = 0.201$], February [$p = 0.194$], March [$p = 0.213$], and April [$p = 0.213$]).

Divergence time estimates

Divergence time estimates suggest a relatively recent split between NW and SE, ~10,000–100,000 YBP based on fast and slow substitution rates (Fig. 3; S6 Table). Despite good MCMC mixing, low autocorrelation, and long runtimes, a long asymptote of low but non-zero probability extended beyond the peak of the t estimate. Additionally, posterior estimates of two parameters, θ_A and m_1 , plateaued without forming a peak within the prior range. This is a common signature of divergence time estimates when the data contains relatively limited information, and thus estimates of t should be considered a minimally recent split with some probability of earlier divergence (Won et al. 2005; Poelchau and Hamrick 2013; Bagley and Johnson 2014a). Although posterior distributions for t may be broad, the IMA2 method is sufficient to discriminate between alternative hypotheses distinguishing long time scales and suggests Pleistocene rather than Pliocene divergence.

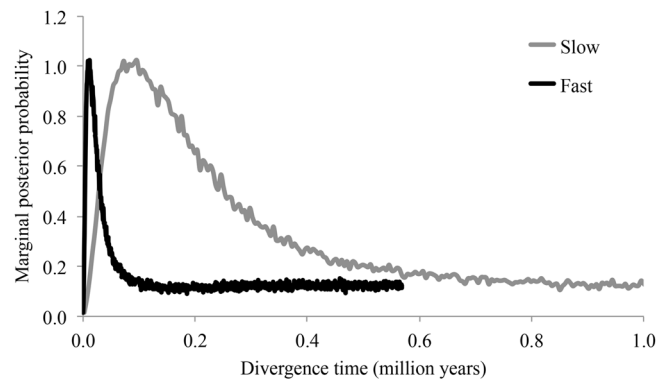


Fig. 3 Marginal posterior probabilities for divergence times between northwestern and southeastern populations of *Brassavola nodosa* in Costa Rica based on IMA2 and alternative cpDNA substitution rates.

Although the x -axis is truncated at 1 million YBP, the posterior distribution of T_{div} assuming a slow mutation rate extends to 4.68 Mya

Ecological niche modeling

The AUC score for the present day model was high (0.977). The two most influential bioclimatic variables were precipitation during the driest quarter (BIO17) and precipitation during the coldest quarter (BIO19), contributing 62.7% and 16.0% to the model, respectively. The inference of habitat suitability in Costa Rica during the Holocene is much weaker, and during the LGM weaker still.

The niche projection on climatic layers for the LGM, when conditions were cooler and drier, indicates a considerably reduced range and reduced habitat suitability, with a maximum probability of occurrence (PrO) of 0.23. The most likely habitat for *B. nodosa* was in western Nicaragua (maximum $PrO = 0.23$), followed by Panama in the region near present day Santiago ($PrO = 0.17$; Fig. 4a) and the Pacific coastline of Costa Rica in the area of today's Nicoya Peninsula ($PrO = 0.11$; Fig. 4b; S4 Figure).

By the Holocene, both the inferred range and suitability of habitat increased, with the most suitable habitat occurring along the Nicaraguan Pacific coastal region (maximum $PrO = 0.75$; Fig. 4c) and northwestern Costa Rica, particularly in the area west of present day Liberia (maximum $PrO = 0.52$ (Fig. 4d; S4 Figure). At this time, no suitable habitat for *B. nodosa* remained in western Panama.

The present day model shows optimal habitat in Costa Rica that overlaps with our study populations (Fig. 4e, f), as well as the area near Santiago, Panama just inland of Bahía de Parita (maximum $PrO = 1.00$; Fig. 4e; S4 Figure).

Discussion

We found high genetic diversity in the chloroplast and nuclear genomes of *B. nodosa*. Spatial structuring of haplotypes and the sharp discontinuity between NW and SE regions of northwestern Costa Rica suggest that the

geographically intermediate TZ represents either a center of diversity from which *B. nodosa* migrated to the NW and SE, or a secondary contact zone between two formerly isolated lineages. The large number of mutational steps separating cpDNA haplotypes and the similar patterns seen for variation at the cpDNA and nDNA loci indicate that secondary contact has produced a transitional zone (TZ) with higher haplotype diversity. This pattern of genetic discontinuity within *B. nodosa* is approximately concordant with patterns seen in six other plant species with various modes of seed dispersal (wind, birds, and vertebrates). Precise locations of these discontinuities are somewhat idiosyncratic for each species; however, the discontinuities of the two lowland orchid species, *B. nodosa* and *L. rubescens*, are nearly identical. Various hypotheses regarding the cause/causes of these discontinuous genetic patterns have been discussed (Cavers et al. 2003; Fuchs and Hamrick 2010; Poelchau and Hamrick 2013; Kartzinel et al. 2013). A compelling explanation is episodic long-distance dispersal across Pliocene islands scattered along the volcanic arc of the Chorotega and Choco Blocks that underlie much of Central America, followed by prolonged population isolation (Bagley and Johnson 2014a).

A prolonged period of isolation between the NW and SE lineages would be consistent with rare, long-distance dispersal over water and colonization of spatially isolated substrates. Biogeographical evidence indicates that many Central American plant species migrated northward from South America (Bagley and Johnson 2014b) via long-distance dispersal prior to isthmus formation (Raven and Axelrod 1974), with the archipelago of volcanic islands that emerged during the Late Cretaceous (Coates and Obando 1996) serving as stepping-stones. The preponderance of *Brassavola* species in South America compared to Mesoamerica and the Caribbean (Jones 1975) strongly suggests a South American origin for this genus. That the total number of nuclear alleles, private alleles, and heterozygosity are

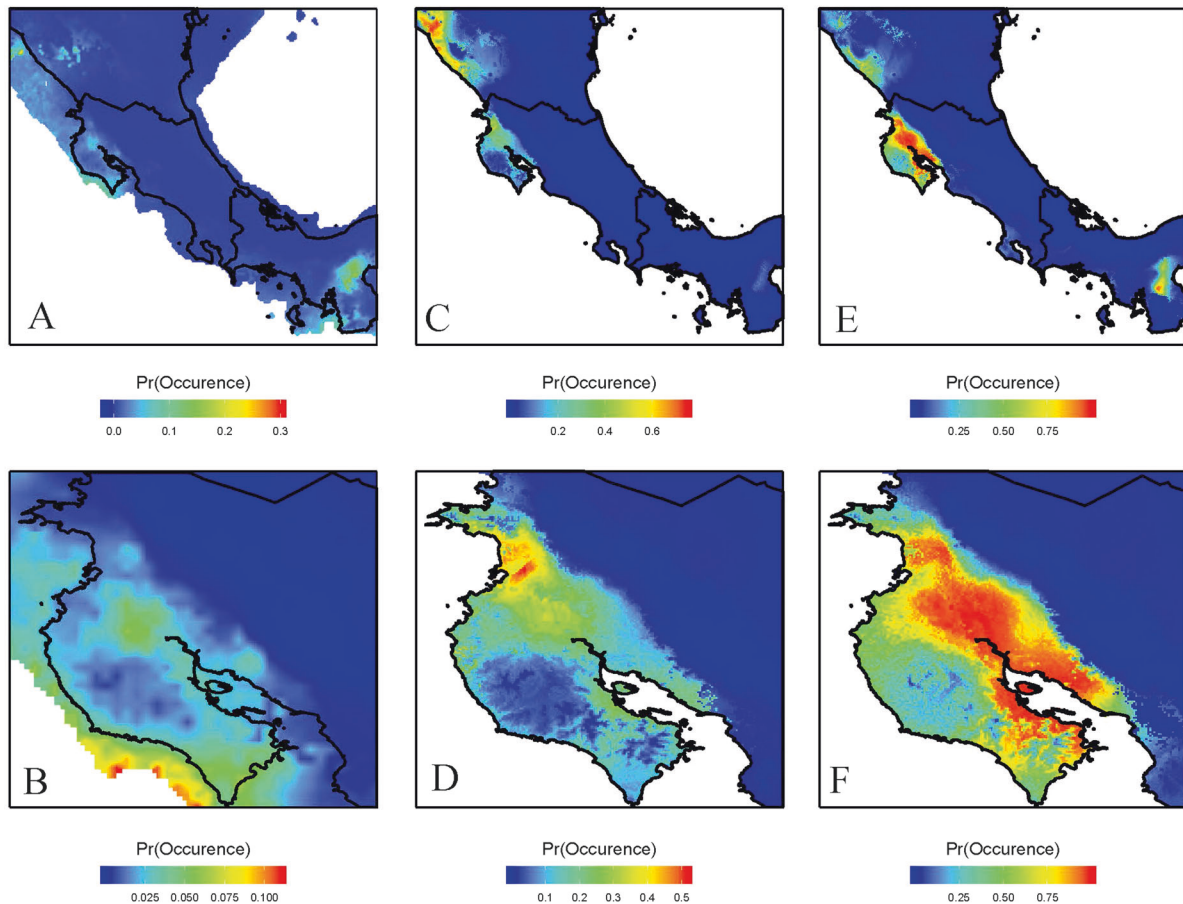


Fig. 4 Predictions of *Brassavola nodosa*'s distribution during the last glacial maximum (~22,000 YBP) and Holocene (~6000 YBP), as inferred from the average of three climatic data sets (CCSM4, MIROC, and MPI-ESM), as well as the present range as inferred from 15 WorldClim climatic variables. Panels B, D and F show the same

distribution models, at a higher resolution, as panels A (LGM), C (Holocene), and E (present) respectively. Blue indicates areas of unsuitable habitat, with warmer colors indicating increasing suitability of habitat for *B. nodosa*

higher in the SE region than the NW region of Costa Rica is consistent with this hypothesis.

The range of estimated divergence times is typical when data contains relatively limited information, and thus estimates of t should be considered a minimally recent split with some probability of earlier divergence (Won et al. 2005; Poelchau and Hamrick 2013; Bagley and Johnson 2014a). Although posterior t distributions may be broad, the IMA2 method can discriminate between alternative hypotheses with different time scales and suggests Pleistocene rather than Pliocene divergence, consistent with the estimated divergence of ~10,000–100,000 YBP. The complex interaction of dynamic factors that shaped the topography and climate of Central America over more recent time scales may have been responsible; extensive volcanism (~50 contemporary active or potentially active volcanoes; Van Wyk de Vries et al. 2007), completion of isthmus closure, fluctuating sea level (~121 ± 5 m at LGM; Geophysics Study Committee 1990), and alternating cool/dry and warm/wet periods associated with glacial maxima and

minima respectively. Active volcanoes in the Guanacaste and Central Cordilleras during the two most recent peak periods (600–400 ka and 100–0 ka; Carr et al. 2007) not only modified the topography in TZ but also altered precipitation patterns. In particular, the resulting continental divide and associated elevational gain would have (i) precluded extensive recolonization by *B. nodosa* above 600 m and (ii) created a rain shadow on the divide's leeward Pacific flanks (Bagley and Johnson 2014b) that favored *B. nodosa*.

The recency of contact zone formation between *B. nodosa* migrating fronts in TZ is supported by the sharp demarcation of haplotypes between the NW and SE regions (mean $H = 2.4$ and 1.5 respectively; Table 1), the mixture of NW and SE haplotypes found in TZ (mean $H = 4.0$; Table 1) and higher mean number of nuclear alleles (A_T) in TZ (TZ = 48.4 vs. NW = 44.2 and SE = 44.2; Table 2). These patterns are consistent with the two *B. nodosa* lineages coming into secondary contact after Quaternary volcanism and mountain building were well underway. The

MAXENT LGM model is consistent with this conclusion and suggests that there was a low probability of occurrence in this region ~22,000 YBP. The discontinuity occurs at low elevation (< 50 m asl) just southwest of, and parallel to, the transition between the Guanacaste and Tilarán Cordilleras. The Guanacaste Cordillera is separated from the Tilarán Cordillera to the southeast by a narrow, low elevation band of older pre-Pleistocene substrate (Coates and Obando 1996). Leeward foothills of the Guanacaste and Tilarán Cordilleras are bounded by this pre-Pleistocene undifferentiated volcanic rock (Tournon and Alvarado 1995; Coates and Obando 1996; S5 Figure) and it is in these older dry lowlands that half of our populations are located (STR, LIB, HEF, LBR, PVR-A, PVR-B, PVR-C, PVR-D, HIG). Interestingly these populations have nearly twice as many haplotypes (mean = 3.22 versus 1.67) and more nuclear alleles (mean = 48.0 versus 42.8) than populations (NOF, PVN, SOL-A, SOL-B, EPZ, GTL, LER-A, LER-B, and LEN) that occur on younger Quaternary Tempisque Basin sediments (Tournon and Alvarado 1995) that were submerged as recently as the late Pliocene (Coates and Obando 1996). This pattern of reduced genetic diversity is consistent with recent colonization of the younger substrates.

Brassavola nodosa releases wind-borne seeds during the dry season (December to April; DWT, pers. obs.) when trade winds are strongest, blowing from the northeast according to the NOAA/NCEP Global Forecast System Atmospheric data. It is possible that the prevailing winds intensify through this gap between the two cordilleras, creating an effective but permeable barrier (i.e., populations STR and LBR) to seed dispersal between the NW and SE lineages. The pattern seen for *B. nodosa* is similar to patterns observed in *L. rubescens* and the higher elevation *E. firmum*, in approximately the same geographic region (Kartzinel et al. 2013). The finding that cpDNA genetic divergence between pairs of populations lying perpendicular to the trade winds is substantially higher than for pairs positioned parallel to the winds supports this interpretation. It is noteworthy that genetic differentiation among populations within TZ is lower for both cpDNA and nDNA indicating higher levels of gene flow within this region.

Furthermore, although seed dispersal contributed significantly less than pollen to gene flow at the broadest geographic scale (i.e., nearly 100 km), there was nearly a 7-fold decrease in relative importance of seed-mediated gene flow between the NW and SE regions, consistent with historical isolation of the two lineages (Table 3). In *L. rubescens* there is a nearly 9-fold decrease in the relative importance of seed flow between the NW and SE (Trapnell and Hamrick 2004) while in *E. firmum* there is approximately a 4-fold decrease between the Guanacaste and Tilarán Cordilleras (Kartzinel et al. 2013).

The dramatically greater importance of pollen dispersal between NW and SE provides further evidence for a NW/SE seed dispersal barrier and/or recent secondary contact of two discrete lineages. That these three orchid taxa show the same pattern suggests that a similar mechanism is responsible. Because all three species are epiphytic and release their seeds during the dry season, their wind-borne seeds should be similarly dispersed. Thus, differences in their m_p/m_s values likely reflect movement patterns of their pollinators (moths versus hummingbirds) rather than seed dispersal variance. The relative importance of pollen movement is much greater in the moth-pollinated orchids (*B. nodosa* [Trapnell and Hamrick 2004] and *E. firmum* [Kartzinel et al. 2013]) than hummingbird-pollinated *L. rubescens*, at the broader geographic level and across the NW/SE discontinuity. Janzen (1983) reported that Costa Rican sphinx moths transport pollen up to 10–15 km. Our data and that of Kartzinel et al. (2013) are consistent with this observation and highlight the effectiveness of sphinx moths for long-distance pollen transport. Unlike insects whose diurnal flight can be restricted by prevailing trade winds, sphinx moths typically fly at dusk and early evening when the wind temporarily subsides. Thus, these strong flyers can transport pollen over much greater distances, regardless of direction. In *L. rubescens*, hummingbird pollinators disperse pollen effectively at distances that rarely exceed a few kilometers, hence the lower m_p/m_s ratio of *L. rubescens* (Trapnell and Hamrick 2005).

In conclusion, our data suggest that Costa Rican *B. nodosa* populations descended from two previously isolated lineages that diverged in the late Pleistocene to Holocene and have subsequently undergone range expansion. Genetic discontinuity between NW and SE, and the genetic profile of populations in TZ provide evidence of secondary contact between lineages and/or a cryptic seed dispersal barrier. Furthermore, concordant discontinuities among multiple plant species suggest a shared biogeographical history. A possible explanation is that with the Quaternary emergence of the Guanacaste and Tilarán Cordilleras, strong northeasterly trade winds channeled through the gap separating these cordilleras formed an effective dispersal barrier for plant species whose seeds are wind-dispersed, and released during the dry season when trade winds are strongest. However, this barrier is somewhat permeable hence the observed zone of admixture and the presence of haplotype H1 in populations STR and LBR in the NW. That a smaller discontinuity is observed in nuclear markers likely reflects that *B. nodosa*'s pollinators, sphinx moths, effectively move pollen across the NW/SE discontinuity detected in the chloroplast genome. While Kartzinel et al. (2013) proposed the possible role of strong winds in limiting seed dispersal among populations linearly distributed perpendicular to the prevailing wind direction, our research is the first to support

this hypothesis with detailed analyses of genetic differentiation in both nuclear and plastid genomes. Our work further illustrates how a dispersal vector (i.e., wind) that can potentially transport seeds considerable distances (≤ 180 km; Trapnell and Hamrick 2004) can also limit gene flow between population pairs located perpendicular to the wind's prevailing direction. This study demonstrates the value of using different genomes to investigate phylogeographic structure shaped within ancient landscapes bearing little resemblance to today's landscapes. It also illustrates how continuously distributed populations contain genetic clues concerning the role that landscape features and other abiotic factors play in shaping the evolutionary trajectory of populations.

Data archiving

Chloroplast haplotype sequences and nuclear genetic data underlying these analyses are available from Dryad: <https://doi.org/10.5061/dryad.hc218qk>.

Acknowledgements The authors thank the University of Georgia Office of the Vice President of Research for financial support, Cecile Deen and Trapnell lab personnel for laboratory assistance, Graham Wyatt for assistance with analyses, and three reviewers for their valuable feedback.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC (2018) TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data* 5:170191
- Agostinelli C, Lund U (2017) R package circular: circular statistics. R package version 0.4-93. <https://r-forge.r-project.org/projects/circular/>
- Arditti J, Ghani AKA (2000) Numerical and physical properties of orchid seeds and their biological implications. *Tansley Review* no. 110. *New Phytol* 145:367–421
- Bagley JC, Johnson JB (2014a) Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. *Biol Rev* 89:767–790
- Bagley JC, Johnson JB (2014b) Testing for shared biogeographic history in the lower Central American freshwater fish assemblage using comparative phylogeography: concerted, independent, or multiple evolutionary responses? *Ecol Evol* 4:1686–1705
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:47–48
- Biodiversity occurrence data published by: Huntington Botanical Gardens Herbarium, Tulane Herbarium, University of California, Riverside Plant Herbarium, and University of Florida Herbarium, SERNEC Data Portal <http://serneportal.org/portal/collections/index.php> Accessed Dec 12 2018)
- Botta F, Eriksen C, Fontaine MC, Guillot G (2015) Enhanced computational methods for quantifying the effect of geographic and environmental isolation on genetic differentiation. *Method Ecol Evol* 6:1270–1277
- Carr MJ, Saginor I, Alvarado GE, Bolge LL, Lindsay FN, Milidakis K et al. (2007) Element fluxes from the volcanic front of Nicaragua and Costa Rica. *Geochim Geophys Geosyst* 8:1–22
- Cavers S, Navarro C, Lowe J (2003) Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Mol Ecol* 12:1451–1460
- Chapman JW, Nesbit RL, Burgin LE, Reynolds DR, Smith AD, Middleton DR, Hill JK (2010) Flight orientation behaviors promote optimal migration trajectories in high-flying insects. *Science* 327:682–685
- Coates AG, Obando JA (1996) Biotic and oceanographic response to the Pliocene closing of the Central American Isthmus. In: Jackson JBC, Budd AF, Coates AG eds. *Evolution and environment in tropical America*. University of Chicago Press, Chicago, 76–104
- Docters van Leeuwen WM (1936) Krakatau, 1833–1933. *Annales du Jardin Botanique de Buitenzorg* 46:47:1–506
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bulletin* 19:11–15
- Dressler RL (1981) *The Orchids: Natural History and Classification*. Harvard University Press, Cambridge
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 11:2571–81
- Engels B (2015) XNomial: extract goodness-of-fit test for multinomial data with fixed probabilities. R package version 1.0.4. <https://cran.r-project.org/web/packages/XNomial/index.html>
- Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 72:250–259
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Fernández-López J, Schlieo K (2018) rWind: download, edit, and include wind data in ecological and evolutionary analysis. *Eco-geography* 42:1–7
- Fuchs EJ, Hamrick JL (2010) Genetic diversity in the endangered tropical tree, *Guaiaecum sanctum* (Zygophyllaceae). *J Hered* 101:284–291
- Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley PH, Yang Z, Zhang M (2011) The community climatic system model version 4. *J Climate* 24:4973–4991
- Geophysics Study Committee (1990) *Sea-level change*. National Academy Press, Washington, D.C.
- Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Böttinger M et al. (2013) Climate and carbon changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J Adv Model Earth Syst* 5:572–597
- Goudet J (2001) Fstat, a program to estimate and test gene diversities and fixation indices (version 2.9.3). <http://www2.unil.ch/izea/software/fstat.html>
- Gregory DP (1964) Hawkmoth pollination in the genus *Oenothera*. *Aliso* 5:385–419
- Haber WA, Frankie GW (1989) The tropical hawkmoth community: Costa Rica dry forest Sphingidae. *Biotropica* 21:155–172
- Hamilton MB, Miller JR (2002) Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. *Genetics* 162:1897–1909
- Hamrick JL, Trapnell DW (2011) Using population genetic analyses to understand patterns of seed dispersal. *Acta Oecologica* 37:641–649

- Hey J (2010) Isolation with migration models for more than two populations. *Mol Biol Evol* 27:905–920
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760
- Hijmans RJ (2017) Geosphere: spherical trigonometry. R package version 1.5-7. <https://CRAN.R-project.org/package=geosphere>
- Hijmans RJ (2018) Raster: geographic data analysis and modeling. R package version 2.8-4. <https://CRAN.R-project.org/package=raster>
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climatic surfaces for global land areas. *International Journal of Climatology* 25:1965–1987
- Janzen DH (1983) Costa Rican Natural History. The University of Chicago Press, Chicago.
- Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* 6:13. v.3.23, <http://ibdws.sdsu.edu/>
- Jones HG (1975) Nomenclatural revision of the genus *Brassavola* R. Br. of the Orchidaceae. *Annalen des Naturhistorischen Museums in Wien* 79:9–22
- Kartzinel TR, Shefferson RP, Trapnell DW (2013) Relative importance of pollen and seed dispersal across a Neotropical mountain landscape for an epiphytic orchid. *Mol Ecol* 22:6048–6059
- Li CC, Horvitz DG (1953) Some methods of estimating the inbreeding coefficient. *Am J Hum Genet* 5:107–117
- Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. *Am J Bot* 83:383–404
- Nei M (1972) Genetic distance between populations. *Am Nat* 106:283–292
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323
- Paradis E (2010) pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419–420
- Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Petit RJ (1992) Polymorphisme de l'ADN chloroplastique dans un complexe d'espèces: les chênes blancs européens. Approche de génétique des populations. Ph.D. Thesis University of Paris XI, France
- Petit R, Kremer A, Bacilieri R, Ducouso A, Zanetto A (1992) Structuration génétique dans le complexe des chênes blancs européens. In: Actes du colloque international en hommage à Jean Pernès professeur à l'Université de Paris 6. Paris: Lavoisier, pp. 155–163. Colloque international "Complexes d'espèces, flux de gènes et ressources génétiques des plantes", Paris, France, January 1992
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modelling of species geographic distributions. *Ecolog Modelling* 190:231–259
- Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161–175
- Poelchau MF, Hamrick JL (2013) Palaeodistribution modeling does not support disjunct Pleistocene refugia in several Central American plant taxa. *J Biogeograph* 40:662–675
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- R Core Development Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Raven PH, Axelrod DI (1974) Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden* 61:539–673
- Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM (2001) Rapid diversification of a species-rich genus of Neotropical rain forest trees. *Science* 293:2242–2245
- Rohlf FJ (2002) NTSYS-pc: Numerical taxonomy system version 2.1. Exeter Publishing, Ltd., Tamil Nadu
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219–1228
- Schemske DE (1980) Evolution of floral display in the orchid *Brassavola nodosa*. *Evolution* 34:489–493
- Tourmon J, Alvarado GE (1995) Carte géologique - Mapa geológico de Costa Rica, Scale 1:500 000. Imprimerie La Vigie, Dieppe, France
- Trapnell DW, Hamrick JL (2004) Partitioning nuclear and chloroplast variation at multiple spatial scales in the Neotropical epiphytic orchid, *Laelia rubescens*. *Mol Ecol* 13:2655–2666
- Trapnell DW, Hamrick JL (2005) Mating patterns and gene flow in the neotropical epiphytic orchid, *Laelia rubescens*. *Mol Ecol* 14:75–84
- Trapnell DW, Hamrick JL, Ishibashi CD, Kartzinel TR (2013) Genetic inference of epiphytic orchid colonization; it may only take one. *Mol Ecol* 22:3680–3692
- Van Wyk de Vries B, Grosse P, Alvarado GE (2007) Volcanism and volcanic landforms. In: Bundschuh J, Alvarado GE (eds.) Central America: Geology, Resources and Hazards. Taylor and Francis, United Kingdom, 123–154
- Varela S, Lima-Riberio MS, Terribile LC (2015) A short guide to the climatic variables of the last glacial maximum for biogeographers. *PLoS ONE* 10(6):e0129037
- Verity R, Nichols RA (2016) Estimating the number of subpopulations (K) in structured populations. *Genetics* 203:1827–1839
- Watanabe S, Hajima T, Sudo K, Nagashima T, Takemura T, Okajima H et al. (2011) MIROC-SEM 2010: model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development* 4:845–872
- Weir BS (1996) Genetic Data Analysis II. Sinauer Associates, Sunderland
- Wolfe KH, Li W, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054–9058
- Won YJ, Sivasundar A, Wang Y, Hey J (2005) On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. *Proc Natl Acad Sci USA* 102:6581–6586
- Workman PL, Niswander JD (1970) Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. *Am J Hum Genet* 22:24–49
- Wright S (1922) Coefficients of inbreeding and relationship. *Am Nat* 56:330–338
- Wright S (1951) The genetical structure of populations. *Ann Eugenics* 15:323–354