Botanical Journal of the Linnean Society, 2021, **197**, 498–512. With 5 figures.

Genomic and niche divergence in an Amazonian palm species complex

CHRISTINE D. BACON[1,2,3*,†], JULISSA RONCAL[4,†], TOBIAS ANDERMANN[1,2], CHRISTOPHER J. BARNES[5], HENRIK BALSLEV[6], NATALIA GUTIÉRREZ-PINTO[3,7], HERNÁN MORALES[8], LUIS ALBERTO NÚÑEZ-AVELLENEDA[9], NATALIA TUNAROSA[9] and ALEXANDRE ANTONELLI[1,2,10]

[1]*Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden*
[2]*Gothenburg Global Biodiversity Centre, Gothenburg, Sweden*
[3]*Laboratorio de Biología Molecular CINBIN, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia*
[4]*Department of Biology, Memorial University of Newfoundland, St. John's, Canada*
[5]*Analytical Biosciences, Department of Pharmacy, University of Copenhagen, Copenhagen, Denmark*
[6]*Department of Bioscience, Aarhus University, Aarhus, Denmark*
[7]*School of Biological Sciences, University of Nebraska, Lincoln, USA*
[8]*Centre for Marine Evolutionary Biology, Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden*
[9]*Departamento de Ciencias Basicas, Universidad de la Salle, Bogotá, Colombia*
[10]*Royal Botanic Gardens, Kew, Richmond, UK*

Received 26 June 2020; revised 5 January 2021; accepted for publication 17 January 2021

Environmental heterogeneity across the landscape can cause lineage divergence and speciation. The *Geonoma macrostachys* (Arecaceae) species complex has been proposed as a candidate case of ecological speciation in Amazonia due to evidence of habitat partitioning and pre-zygotic reproductive barriers between co-occurring morphotypes at a local scale. In this study, we provide a continent-wide perspective of the divergence patterns in *G. macrostachys* by integrating data from morphological traits, target sequence capture, climate, soil and reproductive biology. A morphometric analysis revealed four morphogroups, defined by traits related to leaf shape. A coalescence-based phylogenetic analysis did not recover the morphogroups as monophyletic, indicating independent evolution of leaf shape across geographical space. We demonstrate scale-dependent habitat differentiation for two of the morphogroups, in which segregation driven mostly by climate was complete at the regional scale but incomplete at the continental scale. Contrary to previous evidence of reproductive isolation in the form of different pollinators and flowering times between sympatric *G. macrostachys* forms in Peru and Ecuador, these were not found in Colombia, suggesting reproductive barriers have evolved multiple times across its geographical range. Taken together, our findings suggest that ecological divergence and local adaptation is driving diversification in *G. macrostachys*, and that hyperdiverse regions such as Amazonia are probable arenas for ecological divergence in sympatry.

ADDITIONAL KEYWORDS: Arecaceae – climate – coalescence – ecological divergence – habitat differentiation – morphometrics – pollinators – pre-zygotic reproductive barriers – soil – target sequence capture.

INTRODUCTION

Ecological divergence often evolves early in speciation as divergent natural selection leads to high performance of populations occupying different ecological niches, ultimately leading to reproductive isolation (Schluter, 2001; Rundle & Nosil, 2005; Nosil, 2012; Faria *et al.*, 2014). Ecological divergence leading to speciation can occur in any spatial arrangement, from a single geographical origin in allopatry or sympatry to multiple origins in sympatry

*Corresponding author. E-mail: christinedbacon@gmail.com
†These authors contributed equally.

(Coyne & Orr, 2004). A hallmark study of ecological divergence identified sister species of palms (Arecaceae) on an oceanic island that diverged in sympatry, driven by habitat and pollinator preference (Savolainen *et al.*, 2006). Islands have long served as laboratories for understanding adaptation and divergence (Warren *et al.*, 2015), but whether similar processes occur in continental systems remains unanswered.

The tropical portion of the American continent (the Neotropics) harbours extremely high biodiversity, with more than a third of the world's tropical species (Antonelli & Sanmartín, 2011). This area comprises many different biogeographical regions, including Amazonia, the largest evergreen tropical forest. As Wallace (1853) described in his seminal book, Amazonia is emblemized by Arecaceae, and they are recognized as models of tropical forest evolution (Bacon, 2013; Couvreur & Baker, 2013; Eiserhardt, Couvreur & Baker, 2017). Among these, *Geonoma macrostachys* Mart. is a common and widespread, small (mean = 1.3 m) understorey species, found across western Amazonia in Bolivia, Brazil, Colombia, Ecuador, Peru and Venezuela (Fig. 1; Henderson, Galeano & Bernal, 1995). *Geonoma macrostachys* is the most morphologically variable species in the genus, with up to nine morphotypes informally recognized primarily on the basis of leaf morphology (Henderson, 2011). These morphotypes are probably artificial, and the same morphotype may have evolved independently in different sites (Roncal, Francisco-Ortega & Lewis, 2007; Henderson, 2011). At a larger spatial scale, intermediate phenotypes break down morphotype classification. High morphological variation has made taxonomy difficult and *G. macrostachys* is largely treated as a species complex (Wessels Boer, 1968; Henderson *et al.*, 1995; Henderson, 2011). In addition, *G. macrostachys* is not monophyletic as *G. poiteuana* Kunth from the Guiana Shield was resolved as nested in this species complex (Loiseau *et al.*, 2019).

Geonoma macrostachys has been proposed as a candidate case of ecological speciation (Listabarth, 1993; Roncal, 2006; Roncal *et al.*, 2007; Ostevik *et al.*, 2012). Habitat preferences regarding soil factors and flood regimes have been identified between sympatric morphotypes (Kahn & de Granville, 1992; Svenning, 1999; Vormisto, Tuomisto & Oksanen, 2004; Roncal, 2006; Borchsenius, Lozada & Knudsen, 2016), suggesting that edaphic factors may drive ecologically driven selection. These habitat preferences, however, are not consistent throughout their distribution as at different sites the same morphotype may have different habitat preferences (Vormisto *et al.*, 2004; Henderson, 2011). Despite evidence of habitat segregation, information on whether morphotypes are reproductively isolated is scarce. Pollinator differences attributed to variation in floral scent chemistry and flowering time among sympatric morphotypes are evidence of pre-zygotic reproductive barriers at sites in Ecuador and Peru (Listabarth, 1993; Knudsen, 1999a, b; Borchsenius *et al.*, 2016). Furthermore, distinct morphotypes of *G. macrostachys* from the same location were more closely related than individuals of the same morphotype at distant sites, suggesting that the same morphotype evolved independently, at least across four Peruvian sites (Roncal *et al.*, 2007). Different levels of genetic divergence have also been observed between local sympatric morphotypes

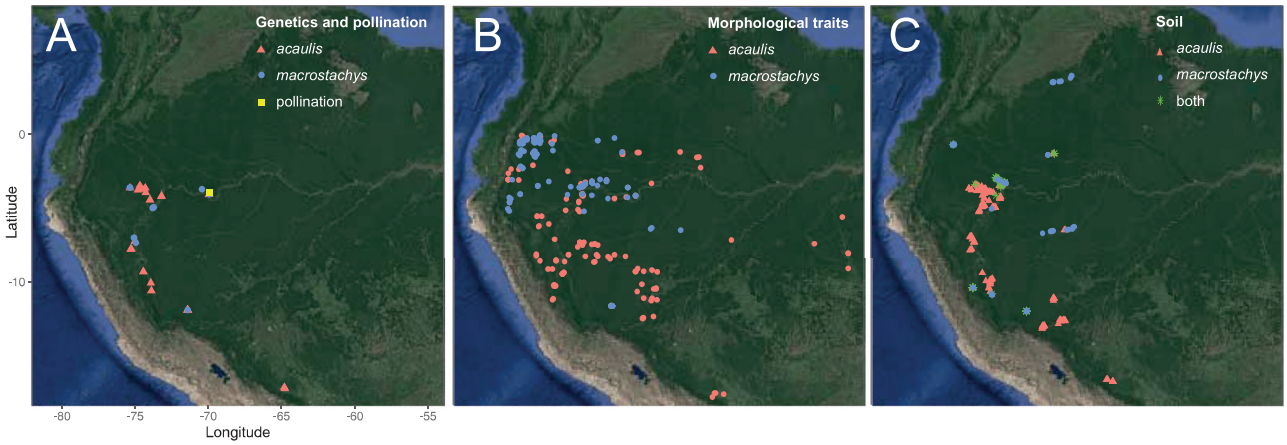

Figure 1. Map of western Amazonia showing the sampling locations. A, sampling of each morphogroup for genetic analysis. Our pollination study was conducted on the *acaulis* and *macrostachys* morphogroups (see Methods). B, sampling of each morphogroup for the morphometric analysis. C, sampling of each morphogroup for habitat differentiation, where 'both' represents transects where the two morphogroups co-occur. In total, 72% of the genetically sampled individuals were collected from the niche transects. The other datasets did not overlap.

(Roncal *et al.*, 2007; Borchsenius *et al.*, 2016). To summarize, where two or more morphotypes grow together at a local site, they appear and behave as different species due to their distinct morphology and potential distinct habitat preference, phenology and pollination. These characteristics make *G. macrostachys* an ideal system to better understand the contribution of ecological divergence to the hyperdiversity in Amazonia.

Previous studies on this species complex have focused on genetic and niche divergence at a local spatial scale (< 100 km^2, but see Cámara-Leret *et al.*, 2016). In this study, we increased the sampling to provide a continent-wide perspective of the divergence patterns in *G. macrostachys*. Our objectives are as follows. (1) We aim to test statistically the nine informal morphotypes of *G. macrostachys* using the morphological dataset of Henderson (2011). In an earlier classification, Henderson *et al.* (1995) recognized two 'varieties': a wide-angled, pinnate (*acaulis*) and a narrow-angled, undivided leaf morphology (*macrostachys*). Here, we hypothesize that we can distinguish these two most common varieties or morphogroups through morphometric analysis. (2) We aim to identify the genetic structure and phylogenetic relationships among representatives of the *G. macrostachys* complex sampled throughout its distribution. We hypothesize that representatives will group geographically and not by morphology, as found in Roncal *et al.* (2007). (3) We evaluate the role of habitat (climate and soil) on morphogroup differentiation at continental (western Amazonia; *c.* 1 500 000 km^2) and regional (Napo, Peru; *c.* 2400 km^2) scales. Based on the suggestion that morphotype habitat preferences are not consistent throughout their distribution (Henderson, 2011), we do not expect to find complete habitat segregation at the continental scale. (4) We test the previously identified prezygotic reproductive barriers (i.e. different pollinator guild and flowering times) between forms at the local scale in Colombia (*c.* 0.2 km^2). Under the ecological divergence hypothesis, we expect to find reproductive barriers throughout the distribution of *G. macrostachys*. Taken together, our results show ecological divergence and adaptive evolution in *G. macrostachys*. Hyperdiverse regions such as Amazonia are probable arenas for ecological divergence, even in sympatry.

MATERIAL AND METHODS

Morphometrics

To test the nine informally proposed morphotypes in *G. macrostachys* statistically (aim 1), we used a dataset of 29 traits measured from 301 herbarium specimens (Henderson, 2011; Fig. 1; Supporting Information, Table S1). The data were filtered so that only traits present in > 30 samples were included, and samples with fewer than four traits were removed, which led to the complete removal of the 'large-raised' morphotype. A distance matrix was constructed using the *daisy* functions in the R statistical platform (R Core Team, 2014) package 'cluster v.2.0.6' (Maechler *et al.*, 2015). Gower's coefficient was calculated by constructing a distance matrix, allowing for quantitative, qualitative and missing data (Podani, 1999). The distance matrix was subsequently used in non-metric multidimensional scaling (nMDS) using the *metaMDS* function in 'vegan v.2.4–6' (Oksanen *et al.*, 2016). Two dimensions of the nMDS were retained with a stress value < 0.2 (Wickelmaier, 2003). The nMDS points were then used to identify the most probable number of clusters within the trait data using the *mclust* function in 'mclust v.5.4' (Fraley *et al.*, 2012). Fourteen models were constructed using default settings with one to nine clusters, and the optimal solution was identified using the Bayesian information criterion. Using the *veganCovEllipse* function in 'vegan', group means (i.e. the clusters produced from the optimal clustering model) and within-group covariances were calculated from the nMDS points. nMDS plots were visualized using 'ggplot2 v.2.2.1' (Wickham, 2016) and ellipses were overlaid for each cluster, with the centroid representing the cluster mean and the outer perimeter of the ellipses representing the within-group covariance (the 95% confidence interval). Finally, a chi-square test was performed to test whether the expected and observed frequencies of morphotypes differed between clusters (meaning the morphotypes are randomly distributed among the clusters or more likely to be assigned to certain clusters), using the *chisq.test* function within the native 'stats v.3.5.0'.

Our morphometric analysis yielded four clusters (morphogroups; see Results below). As predicted, the two largest clusters were congruent with the two most common and easily recognizable varieties of Henderson *et al.* (1995). These two clusters are referred to in our study as morphogroups *acaulis* and *macrostachys*, and these were subsequently used to label samples in the coalescent phylogenetic, habitat differentiation and reproductive biology analyses described below.

Coalescent phylogenetic analysis

To infer a coalescent tree of *G. macrostachys* (aim 2), we sampled 43 individuals from across its distribution (Fig. 1; Supporting Information, Table S2); 72% (31 samples) of the genetically sampled individuals were collected from the field transects described below, and the remaining were collected from further fieldwork in Colombia, Ecuador and Peru. Herbarium vouchers

are detailed in Table S2. DNA was isolated from silica gel-dried tissue using the DNeasy plant mini kit (Qiagen). Library preparation and target enrichment of 837 exons from 176 nuclear genes followed Heyduk *et al.* (2015). Sequencing was completed in a single Illumina Hiseq lane with 250-bp paired-end reads at the University of California Berkeley QB3 sequencing facility. Sequencing reads were processed using the SECAPR pipeline (Andermann *et al.*, 2018). We used the *find_target_contigs* function to identify all *de novo* contigs that matched any of the targeted 837 exons. To avoid paralogous loci, we extracted only those contigs that represented a unique match to a single target exon. We recovered an average of 313 targeted exons per sample, probably due to poor quality leaf material for DNA extraction. The Supporting Information text file presents more details on DNA sequence data processing. Short DNA read data are available at the NCBI Short Read Archive (SRP132119).

We generated multiple sequence alignments for each exon using the extracted contig sequences. We performed an additional reference-based assembly to control for read coverage and paralogous read contamination for each exon locus. The *reference_assembly* function was used to create a new reference library from the contig alignments, which in effect was specific to *G. macrostachys*. Cleaned reads were then mapped to this new reference using the BWA mapper (Li & Durbin, 2010), only allowing reads that had at least 95% similarity to the reference across 100% of the read length. We chose these strict mapping thresholds to avoid mapping of paralogous reads. Although this approach may also lead to the potential loss of useful orthologous read information, the loss is kept at a minimum due to use of a species-specific reference library. This approach resulted in an average of 207 exons per sample that had an average coverage of more than three reads per base throughout the complete locus. We used the resulting BAM assembly files to produce consensus sequences for each sample from the reads covering each exon, only making base calls that were supported by at least three reads. The reference-based assembly recovered an average of 5.6 reads per site calculated across all loci and samples (SD = 4.7).

Because coalescence is not sensitive to missing data (Wiens & Morrill, 2011), we built alignments for all exon loci that were recovered in at least three of the total 43 samples, using MAFFT (Katoh *et al.*, 2002). This resulted in 680 multiple sequence alignments with varying degrees of missing data. We concatenated all exon alignments belonging to the same gene, resulting in 145 separate gene sequence alignments (Supporting Information, Table S3). From these 145 alignments we inferred a coalescent phylogeny in BEAST v.2.4.4 (Bouckaert *et al.*, 2014) using STACEY (Jones, 2017). We used no a priori species assignments and ran all loci with the simplest models to minimize the number of free parameters and overall model space (Jukes–Cantor model and strict clock), fixing '1' for the first locus in the dataset and estimated the clock rate for all other loci relative to that locus. We used the default parameters in STACEY, set as defined in Jones, Aydin & Oxelman (2015), and the analysis was run for 1 000 000 000 generations, sampling every 100 000 generations.

Genetic structure

We extracted single nucleotide polymorphisms (SNPs) from the exon BAM assembly files for the genetic structure analysis (aim 2). SNP calling was performed with GATK UnifiedGenotyper v.3.7-0 (DePristo *et al.*, 2011) using default parameters with a minimum base quality filter of 20 (Illumina quality score, base call accuracy 99%) and retaining only biallelic SNPs. This resulted in a dataset of 7129 SNPs for the 43 individuals. We filtered SNPs with vcftools (Danecek *et al.*, 2011) by removing positions covered by fewer than four reads and > 30 reads. This was done to avoid extremely low confidence on SNP calls and positions with overly high coverage that might be on repetitive regions of the genome. Given the low efficiency of our target capture experiment, support of four reads was a reasonable compromise. GATK is capable of calling SNPs with high confidence, despite low coverage, because calls are assessed across all samples simultaneously, i.e. prior information from all reads across all individuals is considered for any given SNP call. Finally, we removed individuals with > 65% missing data, then SNPs with > 50% missing data and SNPs with a minor allele frequency < 5%, resulting in a dataset of 950 SNPs for 29 individuals (mean missing: per individual = 21% and per site = 27%; Supporting Information, Table S4).

To determine the genetic structure of samples we used the SNP dataset and admixture model with correlated allele frequencies as implemented in STRUCTURE v.2.3.4 (Pritchard, Stephens & Donnelly, 2000). We removed outlier loci that may be subject to directional selection with PCAdapt (195 outlier loci at a 1% false discovery rate; Duforet-Frebourg *et al.*, 2016). This program uses a hierarchical Bayesian model to determine population structure with latent factors [*K*, analogous to principal components analysis (PCA) axes] and identify outlier loci that contribute disproportionately to explaining each of the *K* factors. Given that an initial inspection of 20 *K* factors revealed that the first six *K*s explain the majority of the genetic variation (Supporting Information, Fig. S1), we

performed the final analysis with $K = 6$ and extracted outliers with $Q < 0.01$. Finally, to meet STRUCTURE assumptions that loci are in linkage equilibrium we subsampled one SNP per locus (i.e. per capture probe), rendering a final dataset of 257 neutral SNPs for 29 individuals. We ran STRUCTURE for one to six genetic clusters (K) with ten independent Markov chains of 400 000 iterations of burn-in and 1 000 000 recorded iterations for each K. Convergence of the alpha and log likelihood parameters across chains was confirmed (Fig. S2). Results were summarized and the optimal number of genetic clusters estimated with the Evanno test (Evanno, Regnaut & Goudet, 2005) in STRUCTURE HARVESTER v.0.6.94 (Earl & vonHoldt, 2012).

Habitat differentiation

To evaluate the role of habitat (climate and soil) on morphogroup differentiation (aim 3), we used data from 272 transects of 5 × 500 m in size established in Bolivia, Brazil, Colombia, Ecuador and Peru (Balslev *et al.*, 2011; Kristiansen *et al.*, 2012; Fig. 1; Supporting Information, Table S5). Chemical analyses of soil samples collected within transects were pH, organic matter content, exchangeable acidity, exchangeable aluminium, exchangeable bases, cation exchange capacity, base saturation and phosphorus. Texture analyses including particle size fractions were estimated from visual/near-infrared spectroscopy [VIS/NIR spectrophotometer (350–2200 nm), Veris Technologies, USA]. We averaged values from all soil samples within each transect and log-transformed them prior to analysis. Thirty-three *G. macrostachys* individuals within these transects were sampled for the genetic analyses described above.

Since the process of speciation may vary across geographical scale, we conducted habitat differentiation analyses at two spatial scales: (1) a continental scale (*c.* 1.5×10^6 km^2), with all 272 transects (containing 25 161 *G. macrostachys* individuals in total) representing the complete distribution of *G. macrostachys*, and (2) a regional scale (*c.* 2400 km^2), where 25 transects (containing 1774 individuals) located in the Napo River area of Peru were analysed. No other regional areas could be analysed because of the low number of transects (< 14) within a similar area size where both morphogroups were recorded. At each spatial scale, a linear discriminant analysis (LDA) using the *lda* function in the R package 'MASS v.7.3-48' (Venables & Ripley, 2002) was conducted to compute the factor that maximizes the inter-morphogroup variance while minimizing the within-morphogroup variance, thus representing the ecological factor axis along which the morphogroups differentiate the most. We used 20 uncorrelated (Pearson correlation, $R \leq 0.80$) soil and climatic variables as predictors in the model. Climatic variables were obtained for each georeferenced transect from the 1970–2000 WorldClim database v.2 at 1-km^2 resolution (Fick & Hijmans, 2017) and log-transformed. Missing soil data (13.8%) were imputed across the entire dataset using the *mice* function in the R package 'mice v.2.46.0' using 50 iterations and the predictive mean matching method (van Buuren & Groothuis-Oudshoorn, 2011). LDAs were conducted with the soil and climatic variables separately before combining them. LDA plots were visualized using 'ggplot2' (Wickham, 2016).

Furthermore, we quantified niche overlap at the continental spatial scale using Schoener's D statistic (Schoener, 1970), which ranges from 0 (no niche overlap) to 1 (complete overlap), as implemented in 'ecospat' (Broennimann *et al.*, 2012). We defined sympatric transects as those where two morphogroups co-occur, and allopatric transects as those where only one morphogroup was recorded. We estimated niche overlap for the following pairwise comparison of transects: (1) all *macrostachys* vs. all *acaulis* (allopatric and sympatric transects); (2) *acaulis* allopatric vs. sympatric; (3) *macrostachys* allopatric vs. sympatric; and (4) *acaulis* allopatric vs. *macrostachys* allopatric. We also quantified niche equivalency and similarity for the above pairwise comparisons using the niche overlap metric D following Kirchheimer *et al.* (2016), each with 1000 simulations. Specifically, niche equivalency tests whether the overlap between compared niches is higher than between two random niches, whereas the niche similarity test evaluates if the overlap between niche 1 and niche 2 is higher than when comparing niche 2 with the background. The niche similarity tests were run in both directions (niche 1 vs. 2 and niche 2 vs. 1), with the complete dataset used as background data. The untransformed uncorrelated variables used here were the same as for the LDA.

Pre-zygotic reproductive barriers at a new local site

We compared the reproductive strategies of two morphogroups occurring in sympatry at El Zafire biological station in Colombia (aim 4; Fig. 1; Supporting Information text). During three field trips in November 2013, April 2014 and February 2015, we marked 95 individuals (28 and 67 from *acaulis* and *macrostachys*, respectively) in reproductive phase. These marked individuals were not used in the morphometric, genetic or habitat differentiation analyses. We compared floral morphology and reproductive efficiency (fitness) between morphogroups (Supporting Information

text). We also estimated flowering synchrony, time of anthesis, flowering rhythm, floral longevity, stigmatic receptivity, the presence of pollen and pollen viability for each morphogroup (Supporting Information text).

We made daily observations to identify floral visitors from 20 individuals of *macrostachys* and seven of *acaulis*. We recorded behavioural observations and collected all insects visiting inflorescences in 70% alcohol for species identification. We estimated the frequency and abundance of each insect species (Supporting Information text). Visitors and pollinators were differentiated by estimating pollen flow, measured from pollen on insect bodies, using the pollinator importance value (PIV). PIV is estimated by multiplying pollen transport capacity out of the male flower, pollen transport capacity towards the female flower, abundance, fidelity and floral constancy of the female phase (Núñez-Avellaneda & Rojas-Robles, 2008). PIV was transformed to relative importance of pollinators (RIP) expressed as the percentage of PIV for an insect species over the sum of all PIV values for all visitors per morphogroup. Pollinator efficiencies were categorized as principal pollinators with the highest RIP (> 30), secondary pollinators (RIP = 1–30), occasional pollinators (RIP = 0–1) and no role on pollination (RIP = 0).

RESULTS

Morphometrics

The nMDS analysis was performed on the matrix of Gower's coefficients of the trait data from Henderson (2011), but none of his morphotypes formed a discrete group (Fig. 2A). Clustering analysis performed on the nMDS distance matrix classified the samples into four clusters or morphogroups based on their trait data with no a priori morphotype information (Fig. 2B). Chi-square tests revealed a non-random distribution of Henderson's morphotypes among these four clusters (χ^2 = 480.3, df = 24, P < 0.001). Cluster 1 was associated with morphotypes named *acaulis*, *grandiflora* and *tapajotensis*. Cluster 2 had *atrovirens*, *macrostachys* and *tamandua* morphotypes

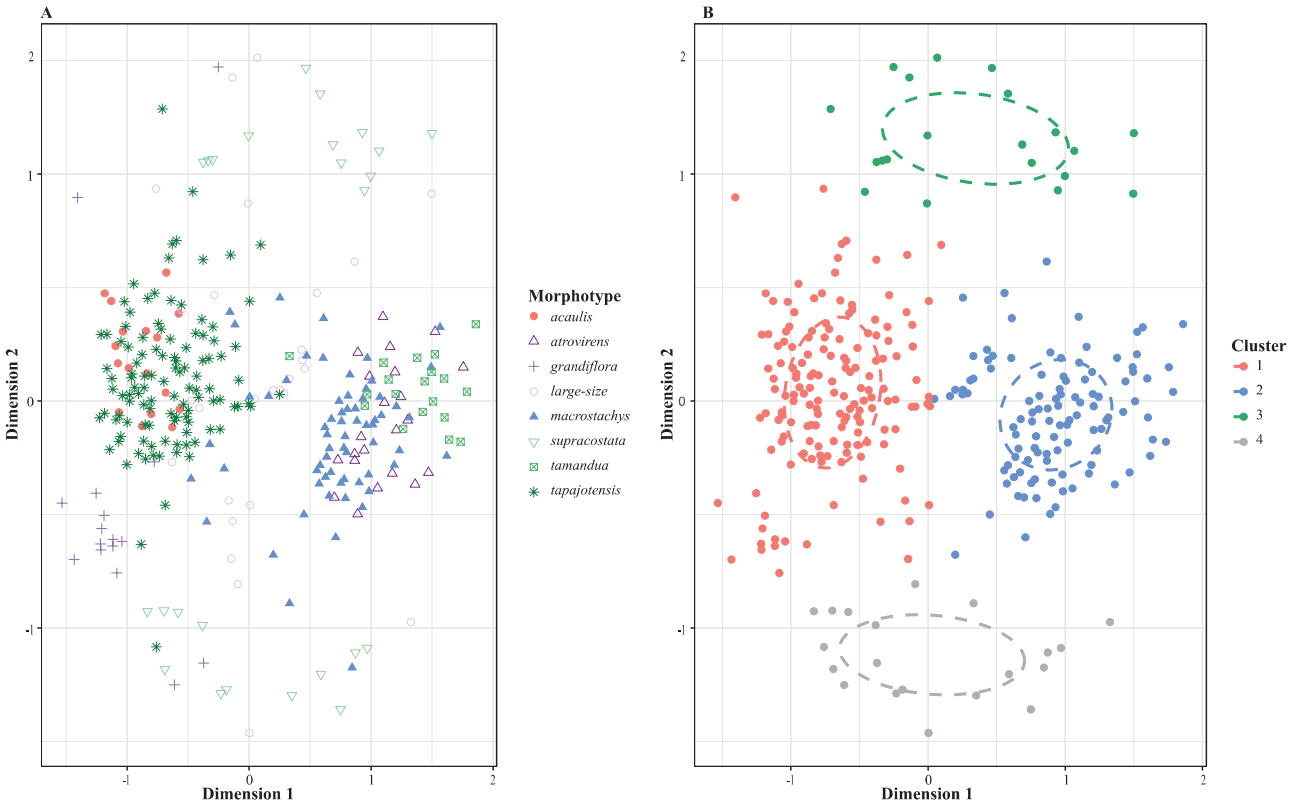

Figure 2. Visualization of the non-metric multidimensional scaling analysis that was performed on a matrix of Gower's coefficients generated from the trait data of Henderson (2011). A, colours represent the morphotype designation of the sample from Henderson (2011). B, trait-based clustering analysis was performed on the nMDS of the trait data with no a priori morphotype information. Four trait-based clusters or morphogroups were identified and illustrated on the plot by colour. Cluster 1 = *acaulis*; cluster 2 = *macrostachys*. Clusters 3 and 4 were not further studied.

assigned (Fig. 2). Clusters 1 and 2 were driven by leaf characters, specifically of basal pinna length and apical pinna length and width (Supporting Information, Table S1). Outside of the two main, tightly clustered morphogroups, loosely clustered individuals were found with intermediate trait values (clusters 3 and 4, Fig. 2B). The *supracostata* morphotype was mostly assigned to clusters 3 and 4, showing considerably more spread within the nMDS plot (Fig. 2). Increasing the number of traits per sample used in the morphometric analysis did not affect the observed trend (Fig. S3). The two largest clusters identified here were congruent with the two most common and easily recognizable varieties of Henderson *et al.* (1995). Therefore, we subsequently recognized clusters 1 and 2 as morphogroups *acaulis* and *macrostachys*. Clusters 3 and 4 were not further studied because they are less common intermediates. We report the traits that statistically distinguish each of the four morphogroups in the Supporting Information text and Table S6.

Coalescent phylogenetic tree and genetic structure

A coalescent tree was inferred for 43 *G. macrostachys* individuals from 145 individual alignments. We removed an initial 75% as burn-in, yielding effective sample sizes that were at least 200, suggesting convergence. Incongruence between the individual gene trees was evident in the coalescent tree by low branch support in many parts of the tree (Fig. 3). The topology showed that the two leaf shape morphologies have evolved numerous times across the phylogeny, as morphogroups did not form reciprocally monophyletic groups. When mapping geographical areas onto the tree, our results showed that there are groupings of local populations, many of which were well supported and genetically distinct (e.g. Loreto and Manu in Peru), but the monophyly of geographical areas in western Amazonia was only found for the Bolivian area (Fig. 3). Furthermore, the tree shows that representatives from the southern portion of the distribution (Manu in Peru, Cochabamba in Bolivia) are found in the most derived lineages of the tree, whereas early divergent lineages are found in the northern part of the distribution (Ecuador and Colombia; Fig. 3).

The most probable number of genetic clusters according to STRUCTURE was $K = 4$ (Fig. 4; Supporting Information, Fig. S4; Table S7). This genetic structure supported two main groups, a widely distributed genetic cluster occurring in the north and south of the study area (brown in Fig. 4 and Fig. S4) and a group in south-western Amazonia (green in Fig. 4 and Fig. S4). We identified two additional genetic clusters representing sub-structure within the widely distributed cluster (blue and yellow in Fig. 4 and Fig. S4). Most notably, these four genetic clusters do not correspond to the four morphological clusters as observed by their geographical locations (Figs 1, 4).

Habitat differentiation

Habitat differentiation between morphotypes using the climatic and soil variables combined was incomplete at the continental scale, but was strong at the regional scale (Fig. 5). Uncorrelated variables used in the LDA and overlap analyses are reported in Supporting Information Table S8. The LDA model for the continental scale achieved a prediction accuracy of 83%, whereas the regional-scale LDA model achieved an accuracy of 96%. The first discriminant factor for the continental-scale analysis was related to maximum temperature of warmest month (bioclimatic variable 5) and mean temperature of wettest quarter (bioclimatic variable 8). Bioclimatic variables 5, 8 and minimum temperature of coldest month (bioclimatic variable 6) were the most discriminating variables in the regional-scale analysis. Overall, the *acaulis* morphogroup grows at slightly higher temperatures than *macrostachys*. Thus, climatic variables seem to be more important than soil type in the habitat differentiation of morphogroups at both spatial scales (Fig. S5; Table S8). LDA using soil and climatic variables separately rendered similar results to the combined analysis with a stronger discrimination at the regional than at the continental scale (Fig. S5; Table S8). At the regional scale, habitat segregation was stronger when soil and climatic variables were combined than when analysed separately.

Schoener's *D* value for the niche overlap between all *macrostachys* and *acaulis* transects was 0.29, which was concordant with the continental LDA result (Supporting Information, Table S8). Allopatric vs. sympatric transects showed a higher niche overlap for *macrostachys* than for *acaulis* (Table S9), suggesting that *acaulis* has a broader niche at the continental scale. Equivalency and similarity tests were not significant, except for the allopatric *macrostachys* vs. sympatric comparison, in which niche overlap was higher than that in any of the comparisons were to the background (Table S9).

Pre-zygotic reproductive barriers at a new local site

The *acaulis* morphogroup was overall smaller than *macrostachys* (Fig. S6, Table S10, Supporting Information tex). Both morphotypes showed inflorescences and infructescences at different developmental stages in all field trips, suggesting that flowering is continuous and asynchronous (Fig. S6). Male and female floral rhythms span 18 days for both

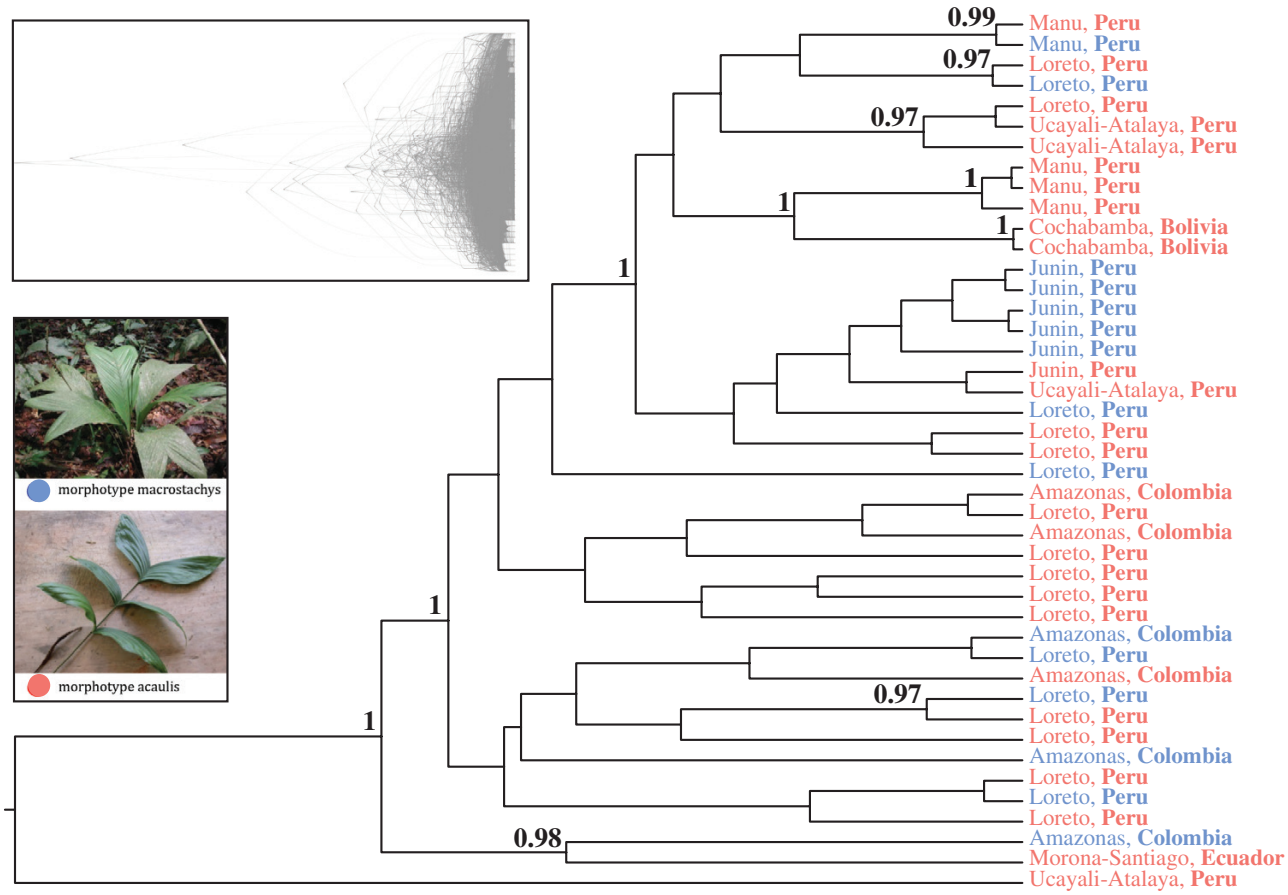

Figure 3. Coalescent tree from STACEY for 43 *Geonoma macrostachys* individuals using 145 genes derived from target sequence capture. Posterior probabilities of branch support > 0.85 are shown above nodes. The geography of samples is visualized at the tips, and morphogroups (*acaulis* in red and *macrostachys* in blue; inset) do not form monophyletic groups. The posterior distribution of trees, which did not converge for the majority of nodes, is displayed in the inset, indicating potential gene flow and/or incomplete lineage sorting in these data.

morphogroups (Fig. S7, Supporting Information text). We found a timelag between morphogroups regarding daily anthesis. Male flowers from *macrostachys* reached anthesis earlier than those of *acaulis*, but with an overlap period of 2.5 h in the morning (Table S10, Supporting Information text). Reproductive efficiency was significantly lower in *acaulis* (mean = 4.9%) than in *macrostachys* (mean = 7.8%) ($t = 3.25$, $P < 0.001$).

The diversity of insects visiting the inflorescences was the same for both morphogroups. Most visitors used flowers for food, mating and oviposition. The most abundant visitor for both morphogroups was the bee *Oxytrigona mellicolor*, and another bee (*Euglossa* sp.) was more abundant and frequent for *macrostachys* than for *acaulis* (Supporting Information, Table S11). These two species were principal pollinators for both morphogroups because they carried copious amounts of pollen (Table S11). Occasional pollinators and insects that played no role in pollination were also the same for both morphogroups (Table S11).

DISCUSSION

We used a comprehensive dataset to provide insight on the divergence patterns of an abundant and taxonomically complex palm species from western Amazonia. In doing so, we demonstrate the potential of comparing diverse datasets from genomics, habitat use and reproductive biology in understanding the process of ecological divergence at the continental scale. Identifying patterns of diversification is important for understanding biodiversity assembly, especially in hyperdiverse regions such as Amazonia.

Morphometric analysis

In his revision of *Geonoma*, Henderson (2011) used unique discrete trait combinations and/or geographical disjunctions to formally recognize subspecies within species complexes. His subspecies recognition was usually supported by quantitative data. In the case

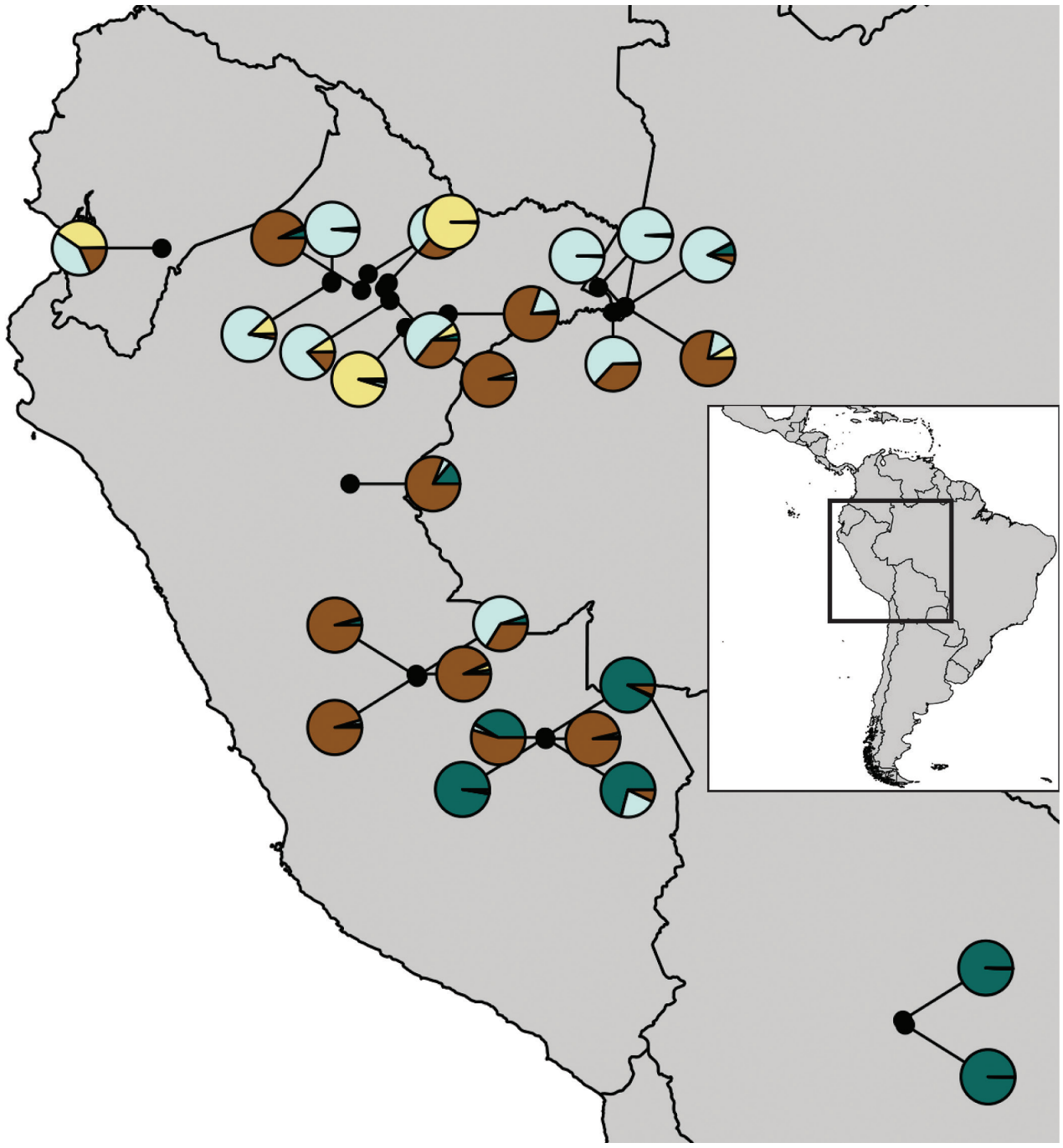

Figure 4. Distribution of 29 *Geonoma macrostachys* genetic samples. Pie charts show individual assignment probability scores to four genetic clusters marked in colours according to the STRUCTURE admixture model (Pritchard *et al.*, 2000) with putative neutral genetic loci. Inset: map of south America showing location of study site. See Supporting Information, Figure S4 for bar plot format.

of *G. macrostachys*, subspecies delimitation using his method was problematic because of the large morphological variation, which led Henderson (2011) to group similar specimens informally into nine morphotypes. Thus, his morphotype definition did not follow a formal statistical analysis. Our ordination analysis (nMDS) yielded four morphological clusters of low taxonomic value in light of our genomic results. However, morphological clusters 3 and 4, not sampled in our genomic study, must be included in future

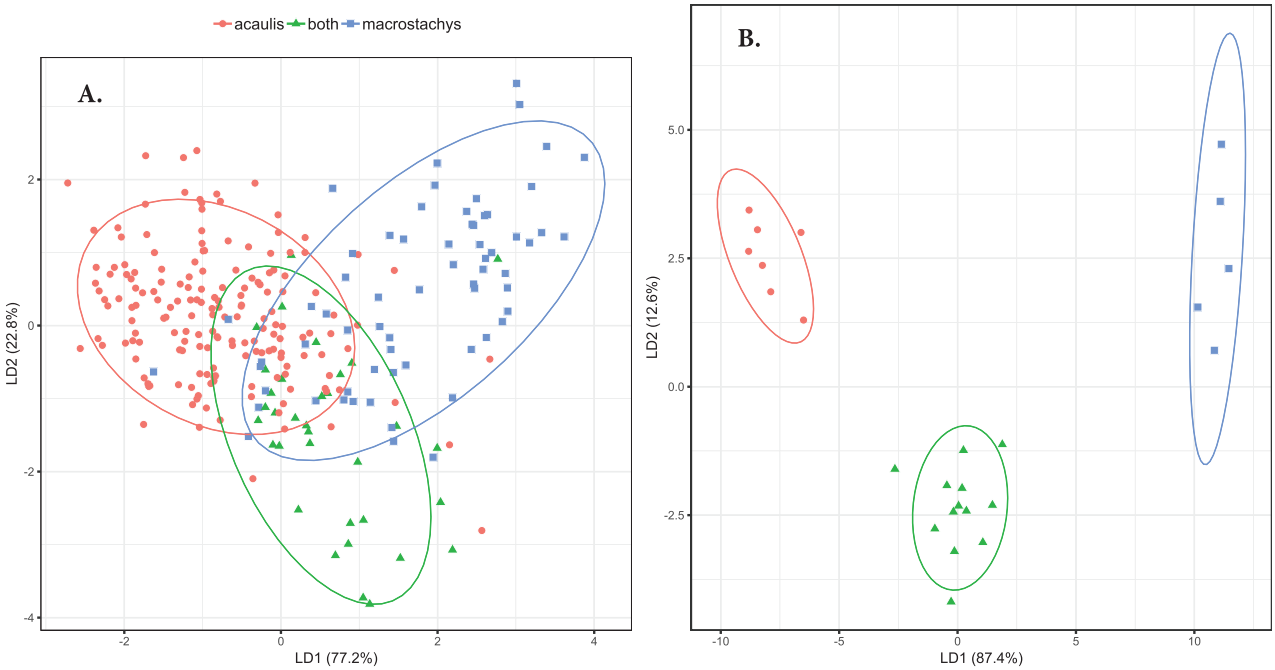

Figure 5. Linear discriminant analysis (LDA) plot showing habitat differentiation between *Geonoma macrostachys* morphogroups at (A) the continental scale and (B) regional scale. Ellipses are the 90% confidence intervals. In parentheses along the axes are the percentages of the between-group variance explained by each discriminant factor. Acaulis (red circles) = transects where only the *acaulis* morphogroup occurs. Macrostachys (blue squares) = transects where only the *macrostachys* morphogroup occurs. Both (green triangles) = transects where two morphogroups co-occur.

phylogenomic analyses to test their monophyly and history of population divergence.

Morphogroups are not monophyletic and do not match the genetic clusters

As expected, morphologically similar individuals did not form groups in our coalescent tree. This pattern is also consistent with Loiseau *et al.* (2019) who sampled 12 individuals from four of Henderson's (2011) morphotypes, but none was monophyletic. Unfortunately, the low branch support in the coalescent tree prevented the evaluation of whether individuals group by geographical position, as in Roncal *et al.* (2007).

Although the coalescent tree should be interpreted with caution due to the large parameter space, the posterior distribution of the tree suggests high gene flow and/or incomplete lineage sorting. Long-distance pollen transfer among *G. macrostachys* populations is feasible through strong flying euglossine bees which are effective pollinators of this species complex (see pre-zygotic barriers below). Plants pollinated by euglossine bees tend to form large metapopulations. Seed or fruit dispersal by birds has been observed in the field (J. Roncal, pers. obs.), and this could be another way to obtain a metapopulation with high gene flow. The decoupling that we found between the morphological and genetic structure in *G. macrostachys* may be the result of ongoing speciation with gene flow or from hybridization after secondary contact of divergent morphological types. Phenotypic plasticity in *G. macrostachys* has not been tested using common garden experiments. However, different genotypes underlie the two local morphotypes specialized to different habitats at the Yasuní National Park, Ecuador (Borchsenius *et al.*, 2016), and in Loreto, Peru (Iquitos; Roncal *et al.*, 2007). In addition, we cannot exclude the possibility of random phenotypic drift in *G. macrostachys*, as was demonstrated for the evolution of morphological traits in *Geonoma* (Roncal *et al.*, 2012).

The spatial scale of habitat differentiation

Niche differentiation between morphogroups was more evident at the regional than continental scale, corroborating Henderson's (2011) hypothesis of

an inconsistent habitat preference of any given morphological form throughout its distribution range. At the continental scale, each morphogroup occurred across a wide range of environmental conditions (temperature, precipitation, soil), potentially masking niche differences between morphogroups. Our continental results agree with those of Cámara-Leret *et al.* (2016), in which, despite the different response shapes of *G. macrostachys* varieties along a gradient of Western Amazonian soil exchangeable bases, varieties had similar mean optimum exchangeable bases and their range values overlapped. Also consistent with Cámara-Leret *et al.* (2016), our niche overlap analysis suggested a broader ecological niche for *acaulis*. Kisel & Barraclough (2010) suggested that the spatial scale of population divergence is a neglected determinant of diversity patterns, showing that the probability of divergence increases (1) with the size of a given region and (2) as the spatial extent of gene flow decreases. Our results reject the first aspect, but the second remains to be tested.

Rivers and their changes through time create a dynamic mosaic of forest types. In Amazonia, river dynamics create floodplains, uplands and terraces, each with different soil characteristics and microclimatic regimes (Salo *et al.*, 1986). Amazonian plants in these heterogeneous landscapes can be subject to divergent selection, particularly related to edaphic factors, leading to ecological divergence (Fine *et al.*, 2005). Consistent with this hypothesis, Vormisto *et al.* (2004), Roncal (2006) and Borchsenius *et al.* (2016) found that soil and topographic conditions explained the niche differentiation between sympatric morphotypes of *G. macrostachys*. We found that morphogroups of *G. macrostachys* can be discriminated by soil characteristics at the regional scale, corroborating these previous studies. However, we also found that climate, notably temperature, is a stronger discriminating factor of the niche than soil properties at both spatial scales. Climate is an important determinant of palm species distribution at the landscape scale and above (> 1000 m; see Eiserhardt *et al.*, 2011, for a review), and studies from across the tree of life have shown the important role of climate as a niche discriminant factor in closely related species or populations (plants: Anacker & Strauss, 2014; frogs: Graham *et al.*, 2004; birds: McCormack, Zellmer & Knowles, 2010).

No evidence of pre-zygotic barriers at the new Colombian site

Our results did not show evidence of pre-mating barriers based on pollinator guild and flowering times. The same pollinators, flowering asynchrony and the fact that stigmas remain receptive during male anthesis of both local morphologies suggest that gene flow between the two is possible. This result contrasts those from previous *G. macrostachys* studies at other sites in western Amazonia (Ecuador and Peru) where different pollinators, floral scents and daily flowering times between local morphotypes were observed (Listabarth, 1993; Knudsen, 1999a, b; Borchsenius *et al.*, 2016). The presence and absence of these pre-zygotic barriers among sites suggest that they may have evolved multiple times. A crossing experiment in Yasuní, Ecuador, showed that pistillate flowers of either local form produced fruit with pollen from the other morphotype, suggesting that the presence of pre-mating barriers (i.e. floral scents and daily flowering times) at this site does not warrant a complete reproductive isolation of forms (Borchsenius *et al.*, 2016). Further studies that extend testing of pre-mating barriers and the level of reproductive isolation between morphological entities are needed.

Our study found less insect visitor diversity (15 visitors of which eight were pollinators) than previous studies. In Yasuní, Ecuador, all 40 insect visitor species carried pollen, making them effective pollinators (Borchsenius *et al.*, 2016). As in our study, bees had the highest pollen loads, and Drosophilidae flies carried little pollen. The higher pollinator diversity in Ecuador might explain the higher reproductive efficiency of the two local morphotypes (13 and 25%; Borchsenius *et al.*, 2016), compared to that in our study site (4.9 and 7.8%). Listabarth (1993) recorded 22 insect species in a central Peruvian forest where meliponine bees were suggested as the most effective pollinators.

A bee species in the genus *Euglossa* was one of the two principal pollinators in our study and deserves attention. Euglossine bees were previously proposed as effective pollinators in *G. macrostachys* because they collect the strong scent emitted by both floral sexes (Listabarth, 1993; Knudsen, 1999a, b, 2002; Knudsen, Andersson & Bergman, 1999), and because of the copious amounts of pollen these bees carry (Borchsenius *et al.*, 2016). Notably, euglossine bees are capable of flying long distances, enhancing gene flow among distant populations (Dressler, 1982). A second type of pollination in *G. macrostachys* occurs through Bakerian mimicry in which pistillate flowers offer no pollen reward, but resemble staminate ones in size, colour and aromatic compounds (Olesen & Balslev, 1990; Listabarth, 1993). Meliponine bees such as *Oxytrigona mellicolor* (a principal pollinator in our study) are pollinators of this category. Pollen reward could explain the higher abundance of insect visitors in staminate than pistillate flowers in our study and in previous ones (Olesen & Balslev, 1990; Listabarth, 1993; Borchsenius *et al.*, 2016).

CONCLUSIONS

We corroborate the hypothesis of Henderson (2011) of certain morphological entities within *G. macrostachys*, and that they may have arisen multiple times in space and/or time. We demonstrate scale-dependent habitat differentiation of new, broadly defined morphogroups. Despite earlier findings of pre-zygotic reproductive barriers between local morphotypes, these are not widespread throughout its geographical range. Here, we demonstrate the power of comparing diverse datasets from across genomes, environmental axes and reproductive biology in understanding the process of ecological divergence. These findings are relevant for understanding the interplay between population divergence and the environment, particularly in species-rich regions such as Amazonia. More generally, identifying common patterns of population divergence allows for generalizations of biodiversity and allows for better understanding of the processes driving species richness.

ACKNOWLEDGEMENTS

We thank Rodrigo Bernal, Finn Borchsenius, Bernard Pfeil, Søren Faurby, Andrew Henderson, Karolina Heyduk, Kerstin Johannesson, Lacey Knowles, Christian Lexer, Jim Leebens-Mack, Margot Paris, Jens-Christian Svenning and María Fernanda Torres Jiménez for discussions and constructive feedback. The Instituto Nacional de Recursos Naturales (INRENA) of the Peruvian Ministry of Agriculture granted collecting and export permits to J.R. (12-C/C-2002-INRENA-DGANP, and 002561-AG-INRENA, respectively) and to H.B. (690-2007-INRENA-IFFS-DCB). The Bolivian Ministry of Environment and Water granted a research permit to the Bolivian National Herbarium for carrying out research under the PALMS 2009–2013 project led by H.B. Collecting permits in Colombia were granted by the Ministry of the Environment and Sustainable Development under the Autoridad Nacional de Licencias Ambientales (Resolution 0790) to CES University in collaboration with María Jose Sanín Perez. C.D.B. was funded by the Swedish Research Council (2017-04980), the International Palm Society, the Montgomery Botanical Center and the Biodiversity in a Changing Climate Strategic (BECC) Research Area at the University of Gothenburg. J.R. was funded by a Canadian NSERC-Discovery grant (RGPIN-2014-03976). C.J.B. was funded by the Aage V. Jensen Naturfond of Denmark (1121721001). H.B. was supported by the Danish Council for Independent Research – Natural Sciences (grant no. 4181-00158) and the European Community (FP/2007–2013, ERC Agreement no. 212631). A.A. was funded by the Swedish Research Council (B0569601), the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013, ERC Grant Agreement no. 331024), a Wallenberg Academy Fellowship, the Swedish Foundation for Strategic Research, the Faculty of Sciences at the University of Gothenburg, the Wenner-Gren Foundation and the David Rockefeller Center for Latin American Studies at Harvard University. There are no conflicts of interest.

REFERENCES

Anacker BL, **Strauss SY. 2014.** The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proceedings of the Royal Society B* **281:** e20132980.

Andermann T, **Cano Á**, **Zizka A**, **Bacon C**, **Antonelli A. 2018.** SECAPR—a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. *PeerJ* **6:** e5175.

Antonelli A, **Sanmartín I. 2011.** Why are there so many plant species in the Neotropics? *Taxon* **60:** 403–414.

Bacon CD. 2013. Biome evolution and biogeographical change through time. *Frontiers of Biogeography* **5:** 227–231.

Balslev H, **Kahn F**, **Millan B**, **Svenning J-C**, **Kristiansen K**, **Borchsenius F**, **Pederson D**, **Eiserhardt WL. 2011.** Species diversity and growth forms in tropical American palm communities. *Botanical Review* **77:** 381–425.

Borchsenius F, **Lozada T**, **Knudsen JT. 2016.** Reproductive isolation of sympatric forms of the understorey palm *Geonoma macrostachys* in western Amazonia. *Botanical Journal of the Linnean Society* **182:** 398–410.

Bouckaert R, **Heled J**, **Kuhnert D**, **Vaughn T**, **Wu CH**, **Xie D**, **Suchard MA**, **Rambaut A**, **Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:** e1003537.

Broennimann O, **Fitzpatrick MC**, **Pearman PB**, **Petitpierre B**, **Pellissier LC**, **Yoccoz NG**, **Thuiller W**, **Fortin MJ**, **Randin C**, **Zimmerman NE**, **Graham CH**, **Guisan A. 2012.** Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* **21:** 481–497.

van Buuren S, **Groothuis-Oudshoorn K. 2011.** mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* **45:** 1–67.

Cámara-Leret R, **Tuomisto H**, **Ruokolainen K**, **Balslev H**, **Kristiansen SM. 2016.** Modelling responses of western Amazonian palms to soil nutrients. *Journal of Ecology* **105:** 367–381.

Couvreur TLP, **Baker WJ. 2013.** Tropical rain forest evolution: palms as a model group. *BMC Biology* **11:** e48.

Coyne J, **Orr HA. 2004.** *Speciation*. Sunderland: Sinauer Associates.

Danecek P, **Auton A**, **Abecasis G**, **Albers CA**, **Banks E**, **DePristo MA**, **Handsaker RE**, **Lunter G**, **Marth GT**, **Sherry ST**, **McVean G**, **Durbin R**, and **1000 Genome Project Analysis Group**. **2011.** The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.

DePristo MA, **Banks E**, **Poplin R**, **Garimella K**, **Maguire JR**, **Hartl C**, **Phillippakis AA**, **del Angel G**, **Rivas MA**, **Hanna M**, **McKenna A**, **Fennel TJ**, **Kernytsky AM**, **Sivachenko A**, **Cibulskis K**, **Gabriel SB**, **Sltschluter D**, **Daly MJ. 2011.** A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43:** 491–498.

Dressler RL. 1982. Biology of the orchid bees (Euglossini). *Annual Review of Ecology and Systematics* **13:** 373–394.

Duforet-Frebourg N, **Luu K**, **Laval G**, **Bazon E**, **Blum MG. 2016.** Detecting genomic signatures of natural slection with principal component analysis: application to the 1000 genomes data. *Molecular Biology and Evolution* **33:** 1082–1093.

Earl DA, **vonHoldt BM. 2012.** STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* **4:** 359–361.

Eiserhardt WL, **Couvreur TLP**, **Baker WJ. 2017.** Plant phylogeny as a window on the evolution of hyperdiversity in the tropical rainforest biome. *New Phytologist* **214:** 1408–1422.

Eiserhardt WL, **Svenning J-C**, **Kissling WD**, **Balslev H. 2011.** Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Annals of Botany* **108:** 1391–1416.

Evanno G, **Regnaut S**, **Goudet J. 2005.** Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Faria R, **Renaut S**, **Galindo J**, **Pinho C**, **Melo-Ferreria J**, **Melo M**, **Jones F**, **Salzburger W**, **Schluter D**, **Butlin RK. 2014.** Advances in ecological speciation: an integrative approach. *Molecular Ecology* **23:** 513–521.

Fick SE, **Hijmans RJ. 2017.** WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37:** 4302–4315.

Fine PVA, **Daly DC**, **Villa Muñoz G**, **Mesones I**, **Cameron KM. 2005.** The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the western Amazon. *Evolution* **59:** 1464–1478.

Fraley C, **Raftery AE**, **Murphy TB**, **Scrucca L. 2012.** Normal mixture modeling for model-based clustering, classification, and density estimation. *The R Journal* **8:** 205–233.

Graham CH, **Ron SR**, **Santos JC**, **Schneider CJ**, **Moritz C. 2004.** Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* **58:** 1781–1793.

Henderson A. 2011. A revision of *Geonoma* (Arecaceae). *Phytotaxa* **17:** 1–271.

Henderson A, **Galeano G**, **Bernal R. 1995.** *Field guide to the palms of the Americas*. Princeton: Princeton University Press.

Heyduk K, **Trapnell DW**, **Barret CF**, **Leebens-Mack J. 2015.** Phylogeonomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* **117:** 106–120.

Jones G. 2017. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology* **74:** 447–467.

Jones G, **Aydin Z**, **Oxelman B. 2015.** DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* **31:** 991–998.

Kahn F, **de Granville J-J. 1992.** *Palms in forest ecosystems of Amazonia*. Berlin: Springer-Verlag.

Katoh K, **Misawa K**, **Kuma K**, **Miyata T. 2002.** MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30:** 3059–3066.

Kirchheimer B, **Schinkel CCF**, **Dellinger AS**, **Klatt S**, **Moser D**, **Winkler M**, **Lenoir J**, **Caccianiga M**, **Guisan A**, **Nieto-Lugilde D**, **Svenning J-C**, **Thuiller W**, **Vittoz P**, **Zimmermann NE**, **Hörandl E**, **Dullinger S. 2016.** A matter of scale: apparent niche differentiation of diploid and tetrapoloid plants may depend on extent and grain of analysis. *Journal of Biogeography* **43:** 716–726.

Kisel Y, **Barraclough TG. 2010.** Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist* **175:** 316–334.

Knudsen J. 1999a. Floral scent chemistry in geonomoid palms (Palmae: Geonomeae) and its importance in maintaining reproductive isolation. *Memoirs of the New York Botanical Garden* **83:** 141–157.

Knudsen J. 1999b. Floral scent differentiation among coflowering, sympatric species of *Geonoma* (Arecaceae). *Plant Species Biology* **14:** 137–142.

Knudsen J. 2002. Variation in floral scent composition within and between populations of *Geonoma macrostachys* (Arecaceae) in the western Amazon. *American Journal of Botany* **89:** 1772–1778.

Knudsen JT, **Andersson S**, **Bergman P. 1999.** Floral scent attraction in *Geonoma macrostachys*, an understorey palm of the Amazonian rain forest. *Oikos* **85:** 409–418.

Kristiansen T, **Svenning J-C**, **Eiserhardt WL**, **Pederson D**, **Brix H**, **Kristiansen SM**, **Knadel M**, **Grandez C**, **Balslev H. 2012.** Environment versus dispersal in the assembly of western Amazonian palm communities. *Journal of Biogeography* **39:** 1318–1332.

Li H, **Durbin R. 2010.** Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589–595.

Listabarth C. 1993. Pollination in *Geonoma macrostachys* and three congeners *G. acaulis, G. gracilis,* and *G. interrupta*. *Botanica Acta* **106:** 496–506.

Loiseau O, **Olivares I**, **Paris M**, **de la Harpe M**, **Weigand A**, **Koubinova D**, **Rolland J**, **Bacon CD**, **Balslev H**, **Borchsenius F**, **Cano A**, **Couvreur TLP**, **Delnatte C**, **Fardin F**, **Gayot M**, **Mejía F**, **Mota Machado T**, **Perret M**, **Roncal J**, **Sanin MJ**, **Stauffer F**, **Lexer C**, **Kessler M**, **Salamin N. 2019.** Targeted capture of hundreds of nuclear

genes unravels phylogenetic relationships of the diverse Neotropical palm tribe Geonomateae. *Frontiers in Plant Science | Plant Systematics and Evolution* **10:** 864.

Maechler M, **Rousseeuw P**, **Struyf A**, **Hubert M**, **Hornik K. 2015.** *cluster: cluster analysis basics and extensions. R package version 2.1.0*. Available at: https://cran.r-project.org/web/packages/cluster/citation.html

McCormack JE, **Zellmer AJ**, **Knowles LL. 2010.** Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: Insights from tests with niche models. *Evolution* **64:** 1231–1244.

Nosil P. 2012. *Ecological speciation*. Oxford: Oxford University Press.

Núñez-Avellaneda LA, **Rojas-Robles R. 2008.** Biología reproductiva y ecología de la polinización de la palma milpesos (*Oenocarpus bataua*) en los Andes Colombianos. *Caldasia* **30:** 101–125.

Oksanen J, **Blanchet FG**, **Kindt R**, **Legendre P**, **Minchin PR**, **O'Hara RB**, **Simpson GL**, **Solymos P**, **Stevens MHH**, **Wagner H. 2016.** *vegan: community ecology package. R package version 2.5–6*. Available at: https://cran.r-project.org/web/packages/vegan/index.html

Olesen JM, **Balslev H. 1990.** Flower biology and pollinators of the Amazonian monoecious palm, *Geonoma macrostachys*: a case of Bakerian mimicry. *Principes* **34:** 181–190.

Ostevik KL, **Moyers BT**, **Owens GL**, **Rieseberg LH. 2012.** Parallel ecological speciation in plants? *International Journal of Ecology* **2012:** e939862.

Podani J. 1999. Extending Gower's general coefficient of similarity to ordinal characters. *Taxon* **48:** 331–340.

Pritchard JK, **Stephens M**, **Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

R Core Team. **2014.** *R: A language and environment for statistical computing*. Vienna: Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Roncal J. 2006. Habitat differentiation of sympatric *Geonoma macrostachys* (Arecaceae) varieties in Peruvian lowland forests. *Journal of Tropical Ecology* **22:** 483–486.

Roncal J, **Francisco-Ortega J**, **Lewis CE. 2007.** An evaluation of taxonomic distinctness of two *Geonoma macrostachys* (Arecaceae) varieties based on inter-simple sequence repeat (ISSR) variation. *Botanical Journal of the Linnean Society* **153:** 381–392.

Roncal J, **Henderson A**, **Borchsenius F**, **Sodre-Cardoso SR**, **Balslev H. 2012.** Can phylogenetic signal, character displacement, or random phenotypic drift explain the morphological variation in the genus *Geonoma* (Arecaceae)? *Biological Journal of the Linnean Society* **106:** 528–539.

Rundle HD, **Nosil P. 2005.** Ecological speciation. *Ecology Letters* **8:** 336–352.

Salo J, **Kalliola R**, **Häkkinen I**, **Mäkinen Y**, **Niemelä P**, **Puhakka M**, **Coley PD. 1986.** River dynamics and the diversity of Amazon lowland forest. *Nature* **322:** 254–258.

Savolainen V, **Anstett M-C**, **Lexer C**, **Hutton I**, **Clarkson JJ**, **Norup MV**, **Powell MP**, **Springate D**, **Salamin N**, **Baker WJ. 2006.** Sympatric speciation in palms on an oceanic island. *Nature* **441:** 210–213.

Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology and Evolution* **16:** 372–380.

Schoener TW. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology* **51:** 408–418.

Svenning J-C. 1999. Microhabitat specialization in a species-rich palm community in Amazonian Ecuador. *Journal of Ecology* **87:** 55–65.

Venables WN, **Ripley BD. 2002.** *Modern applied statistics with S*. New York: Springer.

Vormisto J, **Tuomisto H**, **Oksanen J. 2004.** Palm distribution patterns in Amazonian rainforests: what is the role of topographic variation? *Journal of Vegetation Science* **15:** 485–494.

Wallace AR. 1853. *Palm trees of the Amazon and their uses*. London: Taylor and Francis.

Warren BH, **Simberloff D**, **Ricklefs RE**, **Aguilée R**, **Condamine FL**, **Gravel D**, **Morlon H**, **Mouquet N**, **Rosindell J**, **Casquet J**, **Conti E**, **Cornuault J**, **Fernández-Palacios JM**, **Hengl T**, **Norder SJ**, **Rijsdijk KF**, **Sanmartín I**, **Strasberg D**, **Triantis KA**, **Valente LM**, **Whittaker RJ**, **Gillespie RG**, **Emerson BC**, **Thébaud C. 2015.** Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters* **18:** 200–217.

Wessels Boer J. 1968. The geonomoid palms. *Koninklijke Nederlandse Akademie van Wetenschappen* **58:** 1–202.

Wickelmaier F. 2003. An introduction to MDS. In: *Reports from the Sound Quality Research Unity (SQRU) No. 7*. Aalborg: Aalborg University. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.495.4629&rep=rep1&type=pdf

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. New York: Springer.

Wiens JJ, **Morrill MC. 2011.** Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology* **60:** 719–731.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. A, PCAdapt screeplot of the amount of genetic variation explained by the first 20 *K* clusters (see Methods). B–D, PCAdapt biplots showing the structure detected by the first six *K* clusters. Colours correspond to the four genetic clusters detected by STRUCTURE (see Fig. 2 in the main text) according to their highest probability score of assignment.

Figure S2. Convergence plots for STRUCTURE $K = 4$ analysis for alpha (A) and log-likelihood (B) parameters for 10 iterations. Left panels show the mixing of the chains for each iteration and right panels the distribution of sampled values for each iteration.

Figure S3. Non-metric multidimensional scaling analysis was performed on a matrix of Gower's coefficients generated from the morphological trait data of Henderson (2011) with an increased number of traits per sample to show that the observed trend of morphological clusters is not affected by missing data. A, samples labelled by morphotype as described in Henderson (2011). B, samples labelled according to the four clusters or morphogroups identified in this study. Cluster K1 = *acaulis*; cluster K2 = *macrostachys.* Clusters K3 and K4 were not further studied.

Figure S4. Genetic cluster assignment probability according to STRUCTURE for $K = 2$ to $K = 4$, with samples organized from north to south based on their georeferences.

Figure S5. Linear discriminant analyses (LDA) for *Geonoma macrostachys* morphogroups using (A) continental climatic, (B) regional climatic, (C) continental soil and (D) regional soil data. Ellipses are the 90% confidence intervals. In parentheses along the axes are the percentages of the between-group variance explained by each discriminant factor. Acaulis (red circles) = transects where only the *acaulis* morphogroup occurs. Macrostachys (blue squares) = transects where only the *macrostachys* morphogroup occurs. Both (green triangles) = transects where the two morphogroups co-occur.

Figure S6. *Geonoma macrostachys* morphogroups in sympatry at El Zafire Biological Station in Colombia. A, morphogroup *macrostachys* habit; B, morphogroup *acaulis* habit; C–D, *acaulis* inflorescence at male anthesis; E–F, *macrostachys* inflorescence at male anthesis.

Figure S7. Male and female floral rhythms span 18 days for both *Geonoma macrostachys* morphogroups: A, *acaulis*; B, *macrostachys.*