# Phylogenomic data reveal how a climatic inversion and glacial refugia shape patterns of diversity in an African rain forest tree species

Andrew J. Helmstetter[1]*, Biowa E. N. Amoussou[1,2], Kevin Bethune[1], Narcisse G. Kandem[3], Romain Glèlè Kakaï[4], Bonaventure Sonké[3], Thomas L. P. Couvreur[1,3,5]

[1] IRD, UMR DIADE, Univ. Montpellier, Montpellier, France

[2] Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 04 BP 1525, Cotonou, Benin

[3] Université de Yaoundé I, Ecole Normale Supérieure, Département des Sciences Biologiques, Laboratoire de Botanique systématique et d'Ecologie, B.P. 047, Yaoundé, Cameroon

[4] Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, 03 BP 2819, Cotonou, Bénin

[5] Facultad de Ciencias Exactas y Naturales, Pontificia Universidad Católica del Ecuador, Av. 12 de Octubre 1076 y Roca, Quito, Ecuador

* Corresponding Author

Email: andrew.j.helmstetter@gmail.com

Address:

UMR DIADE, équipe DYNADIV

Institut de recherche pour le développement (IRD)

911 avenue d'Agropolis, BP 64501

34394 Montpellier Cedex 5

France

**ABSTRACT**

The world's second largest expanse of tropical rain forest is in Central Africa and contains incredible species diversity. Population genetic studies have consistently revealed significant structure across central African rain forest plants, in particular a North-South genetic discontinuity close to the equator at the level of a climatic inversion. Here, we take a phylogeographic approach using 351 nuclear markers in 112 individuals across the distribution of the African rain forest tree species *Annickia affinis* (Annonaceae). We show for the first time that the North-South divide is the result of a single major colonisation event across the climatic inversion from an ancestral population located in Gabon. We suggest that environmental differences across the inversion and associated trait divergence may have contributed to this phylogenetic discontinuity. We find evidence for inland dispersal, predominantly in northern areas, and variable demographic histories among genetic clusters, indicating that populations responded differently to past climate change. We show how newly-developed genomic tools can provide invaluable insights into our understanding of tropical rain forest evolutionary dynamics.

Keywords: Phylogenomics, phylogeography, rain forest, sequence capture, Africa, dispersal

**Abbreviations**

TRF = Tropical rain forest

CAR = central African rain forest

LGM = last glacial maximum

SFS = site frequency spectrum

## 1 INTRODUCTION

Tropical rain forests (TRFs) possess an incredibly diverse flora and fauna making up half of the world's biodiversity. Understanding how this diversity is generated is critical if we are to protect it (Mace et al. 2003). Central Africa hosts the world's second largest continuous extent of TRF (Linder 2001). Climatic fluctuations during the Pleistocene and associated glacial forest refugia are suggested to have influenced how genetic diversity is distributed in Central African rain forests (CAR) (Hardy et al. 2013). However, the nature (Anhuf et al. 2006; Diamond and Hamilton 1980; Maley 1996; Bonnefille 2007) and importance of forest refugia continue to be intensely debated (Cowling et al. 2008; Lezine et al. 2019). Population genetic studies within CAR plant species document differing levels of response to past climatic fluctuations (reviewed in Hardy et al. 2013). Conversely, one major phylogeographic pattern common to many CAR plant species studied is the existence of a phylogeographic barrier along a North-South axis around 0-3 degrees north (see Fig. 1A; Hardy et al. 2013; Faye et al. 2016; Heuertz et al. 2014). There appears to be no visible geographic barrier to explain this break as continuous rain forest exists across the entire area. This North-South phylogeographic barrier corresponds, however, to the central African climatic hinge, an inversion zone between the northern and southern rainy seasons (Hardy et al; 2013). Interestingly, this barrier is rarely recovered in phylogeographic

studies of animals and thus seems to affect a greater effect on plants groups (e.g. Fuchs and Bowie 2015; Bohoussou et al. 2015; Bell et al. 2017; Blatrix et al. 2017).

Three main hypotheses have been suggested to explain how this North-South genetic discontinuity originated (Hardy et al. 2013). First, TRF might have disappeared (repeatedly) along the climatic hinge during past climatic fluctuations, isolating allopatric north/south populations, which subsequently recolonised the area. Second, because seasons are inverted across the hinge, flowering times might be displaced between northern and southern populations preventing interbreeding and gene flow. Third, successful colonisation across the climatic hinge may be limited by factors such as environmental differences (e.g. changing levels of water stress). At present, little is known about the relative importance of these three scenarios in generating the observed genetic discontinuity.

The phylogeographic approach (Avise et al. 1987) can unravel the history of populations, and ultimately uncover which processes have shaped current patterns of diversity. It is therefore an ideal framework to study how the climatic inversion, and other factors, have shaped patterns of intraspecific diversity in CAR plants. High-throughput sequencing allows the generation of phylogeographic datasets consisting of large numbers of independently segregating nuclear loci that can account for coalescent stochasticity (Edwards et al. 2016) where studies with small numbers of markers fall short. Phylogenomic data can also be used to reconstruct the spatial evolutionary history of species by inferring phylogenetic trees among populations and dispersal dynamics. These approaches can help us to understand relationships among

populations on either side of the climatic inversion and how often lineages traversed this barrier.

If glacial refugia have played an important role in CAR plant dynamics we would expect to find evidence of dispersal inland because most putative CAR refugia are located in the Atlantic Guineo-Congolian region (Fig. 1A; Maley 1996; Anhuf et al. 2006). For example, within the palm species *Podococcus barteri*, modelling past ranges and genetic data supported the hypothesis of one large coastal refugia in western Gabon and southwestern Cameroon (Faye et al. 2016). Furthermore, we would expect population size to increase towards the present as populations spread out from climatically stable areas.

To develop upon our current understanding of the phylogeographic patterns introduced above, we present, for the first time using nuclear phylogenomic approaches, the evolutionary dynamics of a central African tree species, *Annickia affinis.* This species belongs to the pantropical plant family Annonaceae (Chatrou et al. 2012) growing up to 30 metres tall and typically inhabits primary, secondary and degraded rain forests (Versteegh and Sosef 2007). The species is widespread and common across Lower Guinea, from southern Nigeria to the western tip of Democratic Republic of Congo and is therefore ideal for studying CAR phylogeography and the nature of the climatic inversion as a phylogeographic barrier.

Here, we used a newly developed baiting kit (Couvreur et al. 2019) to sequence hundreds of nuclear markers in 112 individuals covering most of the distribution of *A. affinis*. First, we identified the major genetic clusters within *A. affinis* and their

distribution to test if *A. affinis* shows a North-South genetic structuring along the climatic inversion. Second, we built a phylogenetic hypothesis of relationships among genetic clusters and conducted spatiotemporal diffusion analyses to test if dispersal has been frequent across the climatic inversion, or if it has been restricted over time. We also test for an inland (west to east) dispersal pattern, congruent with expansion out of climatically stable areas. Finally, we reconstruct effective population size through time to infer the past demography of each identified genetic cluster to test for recent population expansion, and if demographic histories are congruent among clusters.

## 2 MATERIAL AND METHODS

### 2.1 Sample collection

A total of 112 individuals of *Annickia affinis* were sampled across most of the species distribution range in Central Africa (Table S1). In addition, two individuals were sampled from the sister species *Annickia polycarpa* as outgroups (Couvreur et al. 2019).

### 2.2 DNA extraction, gene capture and sequencing

DNA was extracted from silica gel dried leaves using the MATAB (Sigma-Aldrich, Saint Louis, Missouri, USA) and chloroform separation methods following Couveur et al. (2019). Nuclear markers were captured using the Annonaceae bait kit (Couvreur et al. 2019) made of 11,583 baits 120 bp long targeting 469 exonic regions. Barcoded Illumina libraries were constructed based on a modified protocol of Rohland and Reich (2012). See supplementary methods for details.

### 2.3 Read filtering, contig assembly and multi-sequence alignment

Reads were cleaned and filtered following the protocol in Couvreur et al. (2019) and Hybpiper (Johnson et al. 2016) was used to prepare sequence data for phylogenetic inference. Alignments were conducted using MAFFT (Katoh and Standley 2013) and cleaned with GBLOCKS (Castresana 2000). Putative paralogs for *A. affinis* that were flagged by Hybpiper were verified and removed during this processes. Further information can be found in the supplementary methods.

### 2.4 Phylogenetic inference

We filtered our dataset by choosing only those exons that had 75% of their length reconstructed in 75% of *A. affinis* individuals. We then used the corresponding supercontigs (i.e. targeted regions and surrounding off-target sequences) for phylogenetic inference. We added empty sequences when individuals were missing from locus alignments and we concatenated loci with the pxcat function from phyx (Brown et al. 2017). We assigned a different GTR+GAMMA model to each locus to account for differences in substitution rates. We then ran RAxML (v8.2.9) (Stamatakis 2014) using the '-f a' option with 100 replicates. The tree was rooted using *A. polycarpa* as outgroup. For comparison we also conducted a coalescent-based phylogenetic analysis using ASTRAL-III (Zhang et al. 2017), which uses individual gene trees to infer a species tree. Finally, we constructed a phylogenetic network using splitstree (v4.14.6; Huson and Bryant 2006) and the full SNP dataset (see below) using the neighbour-net algorithm.

### 2.5 Phylogeographic Diffusion in Continuous Space

We reconstructed the spatiotemporal dynamics of *A. affinis* using BEAST v1.8.4 (Drummond and Rambaut 2007) and spreaD3 v0.9.6 (Bielejec et al. 2011). As this analysis is computationally intensive we used a subset of our dataset by selecting only the five most informative loci based on number of phylogenetic informative sites. We added a partition consisting of longitude and latitude coordinates as continuous trait data. We used a HKY+G substitution model and a strict clock model for each genetic locus and an exponential growth coalescent tree prior. We ran the analysis for 100 million generations and assessed effective sample sizes (ESS) using Tracer v1.7 (Rambaut et al. 2018). We then used spreaD3 to visualize the output at several time points during the history of *A. affinis*. We repeated this analysis with the next five most informative loci to ensure similar patterns were recovered across datasets.

**2.6 SNP calling**

We used SeCaPr (v1.1.4; Andermann et al. 2018) to call SNPs as it generates a psuedoreference made up of the consensus sequences for each target locus (paralogs removed) that is tailored to the given dataset, which is more efficient than the bait kit reference made from distantly related Annonaceae species. We then mapped our cleaned, paired reads to this psuedoreference using BWA (v0.7.12; Li and Durbin 2009). Duplicates were removed and SNPs were called using HaplotypeCaller in GATK (v4.0; McKenna et al. 2010). We used bcftools (v1.8; Li 2011) to apply thresholds of mapping quality (>40%) depth (>25), quality by depth (>2) to filter SNPs. We also removed those SNPs with a minor allele frequency < 0.01, kept only biallelic SNPs and excluded monomorphic sites .

**2.7 Population genetic clustering and statistics**

We examined the genetic structure of *A. affinis* using three approaches. First, we undertook a Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010). We used the function *find.clusters* in the R package 'adegenet' (Jombart 2008) to identify clusters using successive K-means with 100,000 iterations per value of k and a maximum k value of 20. We identified the most appropriate number of clusters by examining the change in Bayesian Information Criterion (BIC) with increasing values of k (number of clusters). We then used the function *dapc* in order to define the diversity between the groups identified using *find.clusters*. We performed cross-validation of our DAPC analysis to ensure our chosen number of PCs was reliable. We used cluster membership inferred using DAPC to define populations for calculating summary statistics (see supplementary methods) and downstream analyses.

Second, we used TESS3 - an approach that takes into account geographic location information when inferring population clusters (Caye et al. 2016). TESS3 was implemented using the R package 'tess3r' (Caye et al. 2016). We used the projected least squares algorithm and a maximum k of 20. We examined the cross-validation score for each value of K to identify the appropriate number of clusters.

Third, we used fastSTRUCTURE (Raj et al., 2014) on a reduced set of unlinked SNPs (one per locus). We ran fastSTRUCTURE using the default settings and the simple prior. The script 'chooseK.py' was used to identify the number of clusters that maximized the marginal likelihood and explained the structure in the data.

**2.8 Demographic history**

We used stairway plot (v2; Liu and Fu 2015), a model-flexible approach that uses site frequency spectra (SFS) to infer changes in effective population size ($N_e$) through time. We generated filtered VCF files representing each cluster as detailed above but did not apply a minor allele frequency filter. We then calculated folded SFS for each cluster. Stairway plot uses SNP counts to estimate the timing of events and changes in $N_e$ so the removal of SNPs with missing data may skew counts. To overcome this we modified each SFS by first calculating the minor allele frequency at each SNP and then multiplying this by the mean number of sequences (haploid samples) at each site. This results in a new SFS that makes use of all observed site frequencies and minimizes the number of SNPs removed. The total of samples is slightly reduced based on the amount of missing data. The number of random breakpoints were calculated as recommended in the manual. We used 67% of sites for training and performed 200 bootstrap replicates. The number of observed sites was calculated as the total length of the pseudoreference. We used an angiosperm wide mutation rate of $5.35 \times 10^{-9}$ sites/year (De la Torre et al. 2017) and a generation time of 15 years based on the generation time of the Annonaceae species *Annona crassiflora* (Collevatti et al. 2014). In addition, sequencing error can skew the SFS by inflating the number of singletons. We ran analyses using the entire SFS, and then reran with singletons removed to ensure similar histories were reached.

**2.9 Modelling of current and past ranges**

Current and Last Glacial Maximum (21k years ago; LGM) potential distributions were modelled using MaxEnt (v3.3.3; Phillips et al. 2006) as implemented in 'biomod2' (v3.3-7.1;Thuiller et al.2009). Current and LGM (MIROC global circulation model) climatic data were downloaded from WordClim ver. 1.4 (Hijmans

et al. 2005) at a resolution of 10*10 arc-minute. The LGM period represents the latest unfavourable climate for tropical species and is therefore a good period to model the impact of past climate change on potential range. A total of 346 presence data points (Table S1) covering the known distribution of *A. affinis* were spatially filtered to one point per cell to avoid overfitting due to sampling bias. Model overfitting was constrained by using the β regularization parameter in Maxent, which limits model complexity (Radosavljevic and Anderson 2014), and was set to 2.00 and 4.00, rather than the default MaxEnt value of 1.00. Modelling with all 19 bioclim variables produced unrealistic results and failed to properly model the current species range independent of the regularization parameter (results not shown). Using just eight bioclim variables (four precipitation and four temperature, see supplementary methods) greatly improved the accuracy of the models to the known distribution. Model performance was evaluated using a cross-validation procedure (Ponder et al. 2001, Muscarella et al. 2014, see supplementary methods). Model fit was assessed using area under curve (AUC; Elith et al. 2006) and the true skill statistics (TSS, Allouche et al. 2006). The best fitting model was then projected into the LGM.

## 3 RESULTS

### 3.1 Sequencing

A total of 124.7 million reads were generated for 112 *A. affinis* individuals at an average coverage depth of 77.5x across all targeted loci. Using HybPiper we identified 366 loci where 75% of the exon length was recovered in at least 75% of individuals. A total of 15 loci showed signs of paralogy and were removed, leading to a final dataset of 351 supercontigs totalling 756 kb of sequence data. After cleaning

and filtering our SNP calling approach yielded 6,787 high-quality SNPs from 280 different loci.

### 3.2 How are populations structured across the range of *A. affinis*?

After cross-validation, we chose to keep 40 PCA axes as this was shown to be appropriate for accurately inferring clusters (Fig. S1A). Changes in BIC greatly decreased after k = 4 (Fig. S1B) suggesting that four clusters best fit our data (Fig. 1A, Fig. S2). Two major clusters contained 35 and 63 individuals that were located primarily in Western Gabon (cluster WG) and Cameroon (cluster CA) respectively. Two smaller clusters of seven individuals each were located in eastern Gabon (cluster EG) and Gabon / Republic of Congo (cluster GC). There is a clear discontinuity in genetic structure across the equator, separating cluster CA from the rest, except for a pair of individuals belonging to cluster EG.

The TESS3 analysis also found that four clusters best defined our data (Fig. 1B, S3) with geographic discontinuities generally congruent between analyses (Fig. 1; Fig. S4). Individual admixture proportions revealed limited mixed ancestry within samples (Fig. 1B), except at a single location. The two northern most individuals belonging to cluster EG, found in at Meyo Centre in Cameroon (labelled in Fig. 1A), had a considerable proportion of their ancestry from cluster CA (Fig. 1B). The inverse was true of the two individuals from cluster CA that were from the same location. The fastSTRUCTURE analysis supported aforementioned analyses, even with a reduced SNP dataset. We found that k was between 3 and 5 and results closely mirrored DAPC clusters (Fig. S5). However, there was little evidence for admixture when

using this approach. Our inferred phylogenetic network (Fig. 1C) also revealed four major clusters, and that clusters WG & CA and EG & GC where grouped together.

### 3.3 How did populations of *A. affinis* disperse across central Africa?

The RAxML phylogenetic tree (Fig. 1D) was highly supported at deeper nodes, giving a reliable evolutionary history between major clades. The tree topology reflected our clustering inferences, lending further support to our four identified clusters and robust evidence for phylogeographic structuring. The topology of our ASTRAL tree was very similar to the RAxML tree (Fig. S6).

We assessed the geographic locations in these clusters at a finer scale by mapping each tip of the RAxML phylogenetic tree to its collection site (Fig. S7). We subdivided clusters WG and CA into major clades (I-IV) as points of reference (see Fig. 1). In cluster WG (Fig. S7C) the earliest diverging individuals are found at Mt. Cristal and in coastal rain forests in northwest Gabon. The remaining individuals in cluster WG formed a monophyletic group (clade I) and are found to the South and East, as far as the southern tip of the Republic of Congo. We then examined the geographic locations in cluster CA and identified three clades with distinct geographic distributions (Fig. S7D) going up the Cameroon's Atlantic coast. The middle and largest cluster extended inland. Given this structure we repeated DAPC clustering analyses using only cluster CA individuals and revealed fine-scale genetic structure that supported these three clades (Fig. S8).

Our diffusion analysis was based on 47.3kb of sequence data across five partitions and converged with the ESS > 200 for all parameters after 100 million generations.

The root was inferred to be around central Gabon (Fig. 2A). We estimated a single lineage crossed the climatic inversion, from South to North, establishing cluster CA (Fig. 2). Late in the evolutionary history of *A. affinis* another dispersal event crossed the barrier at Meyo Centre (see Fig. 1A). Our repetition of the diffusion analysis with different loci matched these patterns (Fig. S9) indicating our results are reliable and unlikely to have been biased by particular gene histories.

**3.4 Do different populations share similar demographic histories?**

We estimated the demographic history of the four DAPC clusters (Fig. 3; Fig S10-13 for full plots). Over the last 100 thousand years (Ka) three clusters (GC, WG and CA) experienced similar demographic histories with population decline around 60-80 Ka followed by an increase in $N_e$ towards the present. We found this increase began at different times across these three clusters though all show rapid increases in population size close to the end of the LGM, around 20 Ka. Cluster CA showed evidence of a rapid growth very close to the present, in the last 4 Ka. Cluster EG had a much different history, exhibiting a relatively constant population size throughout its history with a gradual decline in the last 10 Ka. Results were very similar when singletons were removed indicating that sequencing errors were not affecting our analyses (results not shown).

**3.5 Which areas have remained climatically stable over time?**

A total of 113 data points were retained after filtration. The best predictors were Precipitation of Wettest Month (Bio13) and Precipitation Seasonality (Bio15) (Table S2). A regularization multiplier of 2 generated a better model fit than with 4, showing a good visual match with the known distribution of the species at present (Fig. 4A).

The mean value of the AUC for the training and test data were respectively 0.77 and 0.76. The mean value of TSS was 0.454, indicating that the model is better than a random model. During the LGM, the highest presence probabilities were all located along the Atlantic coast in Cameroon, Equatorial Guinea and Gabon (Fig. 4B).

# 4 DISCUSSION

## 4.1 Limited dispersal across the climatic hinge

Intraspecific diversity, based on phylogenomic nuclear sequence data, within the widespread tree species *Annickia affinis* is highly structured with a clear North-South divide between identified genetic clusters (Fig. 1). This is the first time this has been observed in plants using genomic data and adds to the growing evidence of an important phylogeographic barrier around a climatic hinge between 0 and 3°N in numerous CAR distributed plants (Hardy et al. 2013; Heuertz et al. 2014; Faye et al. 2016; Ley et al. 2017; Pineiro et al. 2017). This North-South discontinuity is, however, generally not recovered in CAR distributed animals except in rare cases (e.g. Portik et al. 2017). This suggests that the processes taking place in relation to this barrier affect the flora of CARs more than the fauna. Indeed, Blatrix et al. (2017) showed that this barrier was more abrupt in the studied tree species (*Barteria fistulosa*) than within the associated symbiotic ants. However, the reasons for this genetic break in a seemingly continuous rain forest region remain little understood (Hardy et al. 2013).

Here, we show that, throughout the evolutionary history of *A. affinis*, a single major northward cross-hinge colonisation event occurred leading to the successful establishment of the Cameroon population (Fig. 2C). This result lends support to the

third hypothesis of Hardy et al. (2013), that possible environmental differences have prevented multiple establishments of populations crossing the hinge. Indeed, the small red to black fleshy fruits of *Annickia affinis* (Versteegh & Sosef, 2007) are frugivore-dispersed (Poulsen et al. 2001; Holbrook and Smith 2000) and can potentially travel long distances (> 500 m) for example by hornbills. Thus, the genetic structure of *A. affinis* in general, and the North-South divide in particular, is not linked to seed dispersal limitation *per se*.

We detected that one genetic cluster (cluster EG) extends across the climatic hinge into south Cameroon (Meyo Centre site, Fig. 1A) leading to a more recent, northwards migration event into the climatic hinge area (Fig. 2D). This indicates that the barrier is not entirely impassable, agreeing with other studies (Hardy et al. 2013; Duminil et al. 2015; Pineiro et al. 2017). The Meyo Centre site lies within the inversion zone and several individuals with mixed ancestry are found here (Fig. 1A, B). A similar result was found in *B. fistulosa*, with 20% of individuals sampled near 1°N being hybrids (Blatrix et al. 2017). In addition, the general lack of evidence for admixture found between genetic clusters on either side of the inversion (Fig. 1B; Fig. S5) suggest that gene flow is nevertheless rare. The existence of hybrids in the absence of gene flow between clusters could be the result of intrinsic (developmental; environment independent) or extrinsic (environment dependent) post-zygostic isolation due to lower fitness of hybrids (Turelli et al. 2001; Blatrix et al. 2017). The example of Meyo Centre provides some evidence that even if dispersal across the hinge is possible, it doesn’t result in the successful establishment of new populations, reinforcing the phylogenetic break over time. Hardy et al.’s (2013) hypotheses were not mutually exclusive and given that there is no clear barrier to dispersal of pollen or

seeds, there may be a role for divergence in traits such as flowering time causing reproductive isolation. This may in turn be linked to why successful establishment across the inversion is rare. However, more fine scale sampling and information on ecological differences among populations across the hinge will be needed to test these ideas. Interestingly, similar north-south phylogeographic breaks are also known from the Atlantic rain forests of Brazil, due to differing climatic regimes and floral compositions (Carnaval et al. 2014; Leite et al. 2016). This suggests that similar processes, though not necessarily driven by exactly the same factors, might be driving patterns of intraspecific diversity in different TRF regions.

**4.2 Out-of-refugia migration in northern forests**

The inferred evolutionary dynamics of *A. affinis* support a role for Pleistocene forest oscillations in shaping intraspecific genetic diversity patterns. The four retrieved clusters are found in allopatry or parapatry (Fig. 1A). This supports the hypothesis of incomplete mixing after post-glacial expansion and is similar to patterns found in other CAR species (Hardy et al. 2013) and within species from other TRF regions (Carnaval et al. 2009; Leite and Rogers 2013). Evidence was found for recent demographic expansion in three clusters (CA, GC and WG, Fig. 3), as would be expected if *A. affinis* expanded out of refugia. These expansions were estimated to have taken place 15-25 Ka but we note that further work is needed to determine a more accurate mutation rate and generation time for *A. affinis* to verify the timing of these events. Therefore we avoid interpreting the exact timing of demographic events and instead focussed on the population size trends. Sampling sizes were also small (n =7) for clusters GC and EG meaning we are less confident in the patterns reconstructed for these clusters. Similar patterns of recent expansion were detected in

populations of central African plants (Pineiro et al. 2017) and animals (Bell et al. 2017) as well as in studies on neotropical flora (Vitorino et al. 2016) and fauna (Batalha-Filho et al. 2012). The refuge hypothesis has received support from population genetic studies of CAR plants, showing concordance between putative refugia and regions of high or unique allele/haplotype diversity (Lowe et al. 2010; Dauby et al. 2014; Heuertz et al. 2014; Faye et al. 2016). However, cluster EG showed constant population size with a slight decline towards the present, perhaps indicating that refugia have not played in important role in its demographic history. Similar demographic patterns were found populations of two central African *Erythrophleum* species (Duminil et al. 2015) though these exhibited a more pronounced decline in the last 50 Ka. Overall, our results indicate that demographic responses to past climate change have been different among populations of *A. affinis* across central Africa. Similar patterns of recent expansion were detected in populations of central African plants (Pineiro et al. 2017) and animals (Bell et al. 2017) as well as in studies on neotropical flora (Vitorino et al. 2016) and fauna (Batalha-Filho et al. 2012). The refuge hypothesis has received support from population genetic studies of CAR plants, showing concordance between putative refugia and regions of high or unique allele/haplotype diversity (Lowe et al. 2010; Dauby et al. 2014; Heuertz et al. 2014; Faye et al. 2016).

While our results are mixed, we do find evidence to support the scenario presented by Anhuf et al. (2006) who proposed that coastal rain forests in central Africa acted as refugia during the LGM. The modelled LGM distribution of *A. affinis* indicates that suitable habitat was concentrated continuously along the coast, from Cameroon to Gabon (Fig. 4B), like in the understory palm species *Podococcus barteri* (Faye et al.

2016). In addition, we uncovered fine-scale genetic structure and evidence for dispersal eastwards in Cameroon (Fig. 2), demonstrating a possible out-of-refugia pattern in this area.

In contrast, an inland pattern of migration was not found in Gabon where dispersal was both towards the east and west from a central area. This may be because there was a large amount of highly-suitable area (>0.8) during the LGM that extended further from the coast in Gabon than in Cameroon (Fig. 4), meaning that populations could persist and expand out of this area. In addition, we inferred more pronounced East-West clustering (Fig. 1) in Gabon than in Cameroon, which has been observed in at least four other CAR tree species (Hardy et al. 2013). Bringing our results together, it appears that refugia may have played a different role for populations in different areas, and that each has responded to past climate in change in its own way.

## 5 CONCLUSIONS

This study uncovered the evolutionary dynamics and demographic history of the CAR tree species *Annickia affinis*. Our approach is the first to use genome-wide data from hundreds of nuclear loci to infer population-level phylogeographic patterns in CAR plants. We found high levels of genetic structure consistent with a pattern of North-South genetic discontinuity often recovered in this region. We highlighted how a climatic inversion limits colonisation and shapes patterns of population structure across a continuous rain forest region. We also show that the current distribution of extant populations is the result of different demographic histories and, in northern regions, migration inland from putative refugia in coastal rain forests. This study provides a proof-of-concept for future work taking advantage of recent genomic

resources, such as the sequence capture kit used here, to improve our understanding of TRF evolution, at the population level and above.

**FUNDING**

This study was supported by the Agence Nationale de la Recherche (grant number ANR-15- CE02-0002-01 to TLPC).

**ACKNOWLEDGEMENTS**

We thank Prof. Moutsambote, Raoul Niangadouma, Théophile Ayole for help in the field. The authors acknowledge the IRD itrop HPC (South Green Platform) at IRD Montpellier for providing HPC resources. We are grateful to the Centre National de la Recherche Scientifique et Technique (CENAREST), the Agence National des Parques Nationaux (ANPN) and Prof. Bourobou Bourobou for research permits (AR0020/16; AR0036/15 (CENAREST) and AE16014 (ANPN). Fieldwork in Cameroon was undertaken under the “accord cadre de cooperation” between the IRD and Ministère de la Recherche Scientifique et Technique (MINRESI). Prof. Bouka Biona of the Institut National de Recherche en Sciences Excates et Naturelles (IRSEN) of the Republic of Congo is thanked for research permits.

The authors have no competing interests.

**FIGURES**

Colour should be used for all figures

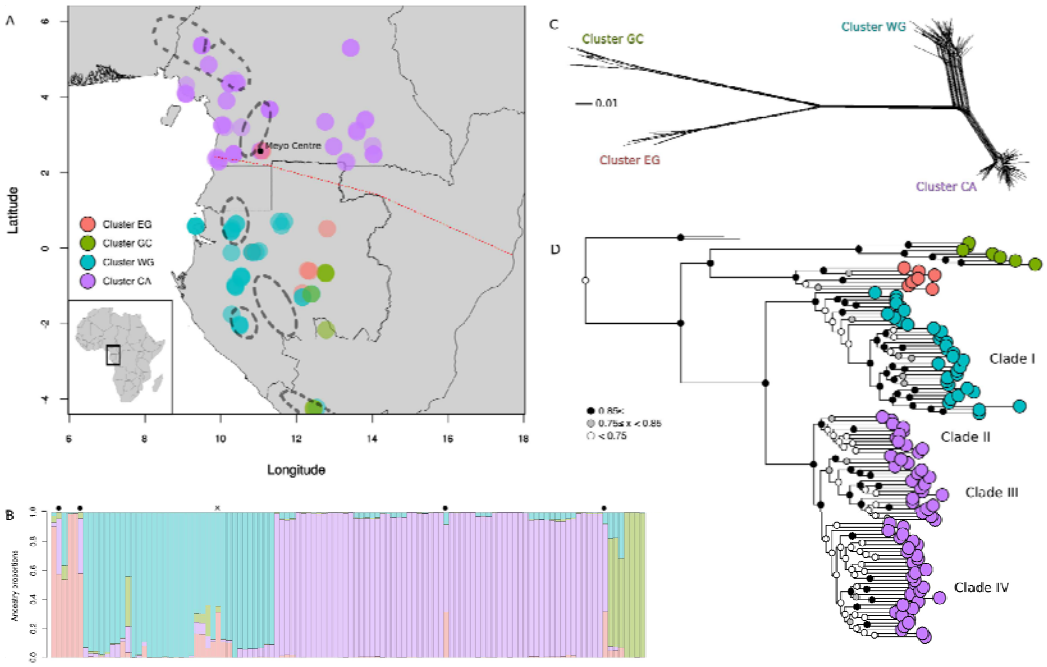



**Fig. 1.** (A) Map of the study region showing genetic clusters inferred using Discriminant analysis of principal components (DAPC, k=4). Individuals are colour coded by cluster membership. Superimposed upon the map are the locations of putative glacial refugia (adapted from Faye et al. 2016). The climatic hinge is shown by a dashed red line. Inset is a map of the African continent showing highlighting the study area. (B) Barplot of ancestry proportions inferred using TESS (k=4). Colours were made to correspond to those in Fig. 1A as clustering was almost identical between approaches. A single individual, “A_affinis_Ndjole_5” (marked with an x), was inferred as cluster EG (red) in DAPC but TESS suggests the majority of its ancestry is instead from cluster WG (blue). Individuals from the “Meyo Centre” collection site (marked with a black circle in panel A) show evidence of admixture between clusters across the North-South climatic inversion. (C) Phylogenetic network

among *A. affinis* individuals constructed in splitstree using NeighbourNet algorithm based on 6787 SNPs. (D) RAxML tree representing relationships among *Annickia affinis* samples, rooted on two *A. polycarpa* samples. Support values are shown and tips are coloured based on genetic clustering results.

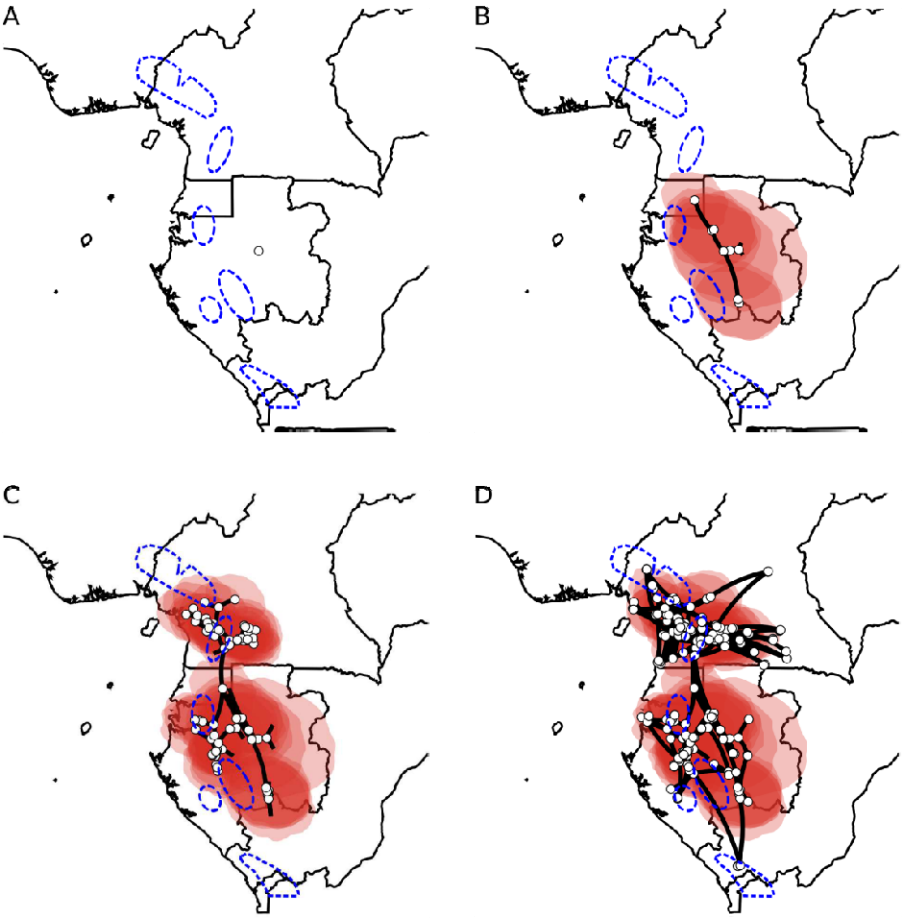



**Fig. 2.** Phylogeographic diffusion analysis split into four time slices. Images were rendered using spreaD3 and move forward through time starting from the (uncalibrated) time of the most recent common ancestor (A) to the present day (D). White circles represent ancestrally estimated geographic locations for nodes in the inferred phylogenetic tree, as well as current, real locations at tips. Polygons around points represent uncertainty of estimated ancestral locations at 80% highest posterior density (HPD). Putative refugia following Maley (1996) are shown in dashed blue lines.

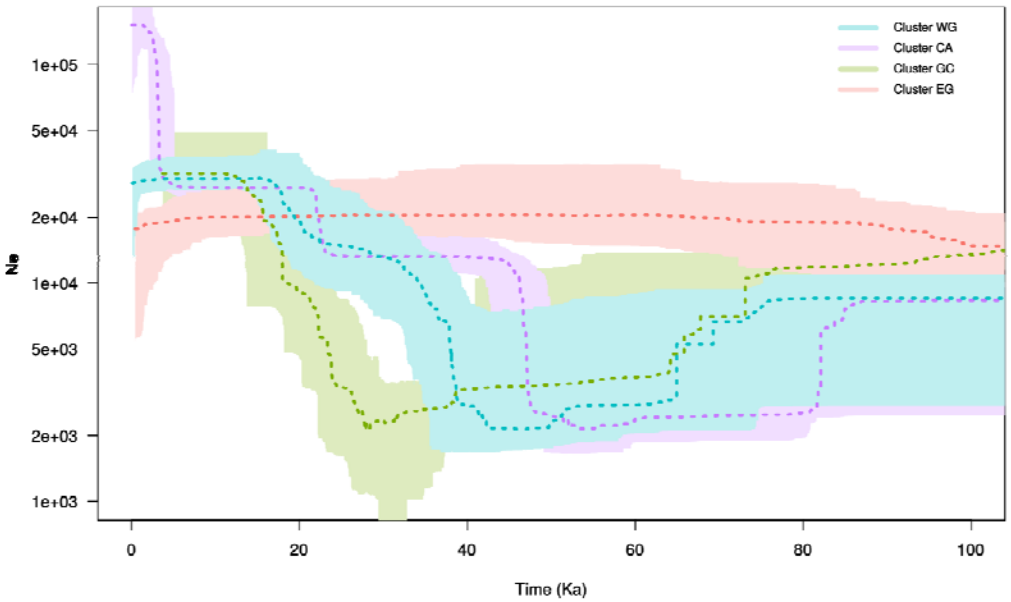



**Fig. 3.** Plots of effective population size through time for each of the four clusters inferred using stairway plot. The present is located on the left side of each graph. The dotted line represents the median population size and the shaded polygon represents the 80% central posterior density intervals. Colours correspond to the colours used in figure 1. Full plots of each species can be found in figures S10-13.

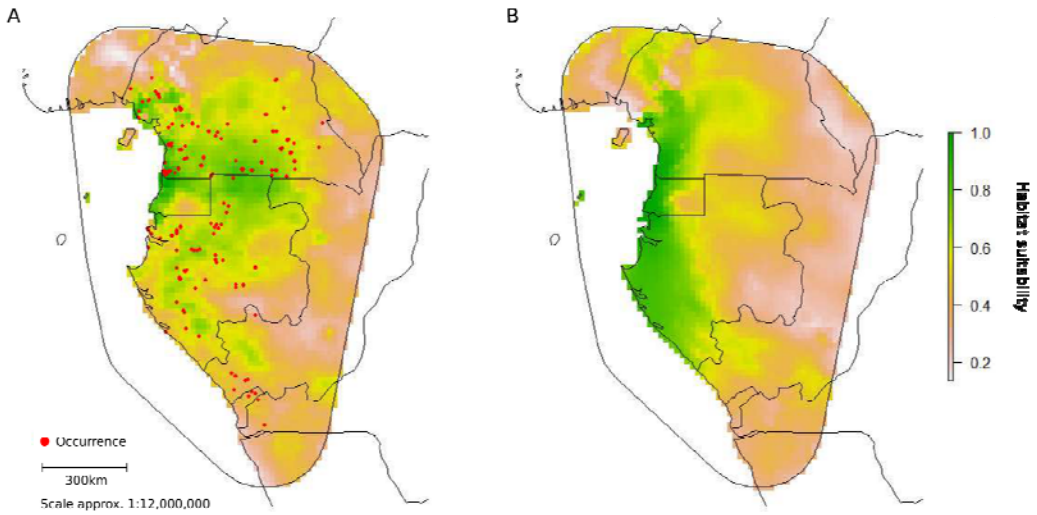



**Fig. 4.** Species distribution models (SDMs) for the present (A) and projected into the past, during the last glacial maximum (B). SDMs were constructed using MaxEnt and bioclimatic variables. The colour scale represents habitat suitability for each cell where green indicates more suitable cells. Red circles in (A) indicate sites where *A. affinis* individuals were collected and used in building the model.

## REFERENCES


Allouche O, Tsoar A, Kadmon R. 2006 Assessing the Accuracy of Species Distribution Models: Prevalence, Kappa and the True Skill Statistic (TSS). *Journal of Applied Ecology* 43(6):1223–32. https://doi.org/10.1111/j.1365-2664.2006.01214.x.

Andermann, T, Cano Á, Zizka A, Bacon C, and 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. https://doi.org/10.7717/peerj.5175.

Anhuf D, Ledru M-P, Behling H, Da Cruz FW, Cordeiro RC, Van der Hammen T, Karmann I et al. 2006 Paleo-Environmental Change in Amazonian and African Rainforest during the LGM. *Palaeogeography, Palaeoclimatology, Palaeoecology* 239(3–4):510–27. https://doi.org/10.1016/J.PALAEO.2006.01.017.

Avise, JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, and Saunders NC. 1987. Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics* 18(1):489–522. https://doi.org/10.1146/annurev.es.18.110187.002421.

Batalha-Filho H, Cabanne GS, Miyaki CY. 2012 Phylogeography of an Atlantic Forest Passerine Reveals Demographic Stability through the Last Glacial Maximum. *Molecular Phylogenetics and Evolution* 65(3):892–902. https://doi.org/10.1016/j.ympev.2012.08.010.

Bell, RC, MacKenzie JB, Hickerson MJ, Chavarría KL, Cunningham M, Williams S, and Moritz C. 2011 Comparative Multi-Locus Phylogeography

Confirms Multiple Vicariance Events in Co-Distributed Rainforest Frogs. *Proceedings of the Royal Society B: Biological Sciences* 279(1730):991–99. https://doi.org/10.1098/rspb.2011.1229.

Bielejec F, Rambaut A, Suchard MA and Lemey P. 2011 SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 27(20):2910-2912.

Blatrix R, Peccoud J, Born C, Piatscheck F, Benoit L, Sauve M, Djiéto-Lordon C et al. 2017 Comparative Analysis of Spatial Genetic Structure in an Ant–Plant Symbiosis Reveals a Tension Zone and Highlights Speciation Processes in Tropical Africa. *Journal of Biogeography* 44(8):1856–68. https://doi.org/10.1111/jbi.12972.

Bohoussou KH, Cornette R, Akpatou B, Colyn M, Peterhans JK, Kennis J, Šumbera R et al. 2015 The Phylogeography of the Rodent Genus Malacomys Suggests Multiple Afrotropical Pleistocene Lowland Forest Refugia. *Journal of Biogeography* 42(11):2049–61. https://doi.org/10.1111/jbi.12570.

Bonnefille R. 2007. Rainforest responses to past climate changes in tropical Africa. In: Tropical rainforest responses to climate change (eds Bush MB, Flenley JR), pp.117–170, Praxis Publishing, Chichester

Brown, JW, Walker JF, Smith SA. 2017 Phyx: Phylogenetic Tools for Unix. *Bioinformatics* 33(12):1886–88. https://doi.org/10.1093/bioinformatics/btx063.

Carnaval, AC, Hickerson MJ, Haddad CFB, Rodrigues MT, and Moritz C. 2009 Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. *Science* 323 (5915):785–89. https://doi.org/10.1126/science.1166955.

Carnaval AC, Waltari E, Rodrigues MT, Rosauer D, VanDerWal J, Damasceno R, Prates I et al. 2014 Prediction of Phylogeographic Endemism in an Environmentally Complex Biome. *Proceedings of the Royal Society B: Biological Sciences* 281(1792):20141461. https://doi.org/10.1098/rspb.2014.1461.

Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution* 17 (4): 540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Caye K, Deist TM, Martins H, Michel O, François O. 2016 TESS3: Fast Inference of Spatial Population Structure and Genome Scans for Selection. *Molecular Ecology Resources* 16(2):540–48. https://doi.org/10.1111/1755-0998.12471.

Chatrou LW, Pirie MD, Erkens RHJ, Couvreur TLP, Neubig KM, Abbott JR, Mols JB, Maas JW, Saunders RMK, Chase MW. 2012 A New Subfamilial and Tribal Classification of the Pantropical Flowering Plant Family Annonaceae Informed by Molecular Phylogenetics. *Botanical Journal of the Linnean Society* 169(1):5–40. https://doi.org/10.1111/j.1095-8339.2012.01235.x.

Collevatti RG, Telles MP, Lima JS, Gouveia FO, Soares TN. 2014 Contrasting spatial genetic structure in Annona crassiflora populations from fragmented and pristine savannas. *Plant systematics and evolution* 300(7):1719-1727.

Couvreur TLP, Helmstetter AJ, Koenen EJM, Bethune K, Brandão RD, Little SA, Sauquet H, Erkens RHJ. 2019 Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes.

*Frontiers in Plant Science* 9 (January). Frontiers:1941. https://doi.org/10.3389/fpls.2018.01941.

Cowling SA, Cox PM, Jones CD, Maslin MA, Peros M, Spall SA. 2008 Simulated glacial and interglacial vegetation across Africa: implications for species phylogenies and trans-African migration of plants and animals. *Global Change Biology* 14:827–840. https://doi.org/10.1111/j.1365-2486.2007.01524.x

Dauby G, Duminil J, Heuertz M, Koffi GK, Stévart T, Hardy OJ. 2014 Congruent Phylogeographical Patterns of Eight Tree Species in Atlantic Central Africa Provide Insights into the Past Dynamics of Forest Cover. *Molecular Ecology* 23(9):2299–2312. https://doi.org/10.1111/mec.12724.

De La Torre AR, Li Z, Van De Peer Y, Ingvarsson PK. 2017 Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Molecular Biology and Evolution* 34(6):1363–1377.

Diamond AW, Hamilton AC. 1980 The Distribution of Forest Passerine Birds and Quaternary Climatic Change in Tropical Africa. *Journal of Zoology* 191 (3): 379–402. https://doi.org/10.1111/j.1469-7998.1980.tb01465.x.

Drummond AJ, Rambaut A. 2007 BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evolutionary Biology* 7(1):214. https://doi.org/10.1186/1471-2148-7-214.

Duminil J, Mona S, Mardulyn P, Doumenge C, Walmacq F, Doucet JL, Hardy OJ. 2015 Late Pleistocene Molecular Dating of Past Population Fragmentation and Demographic Changes in African Rain Forest Tree Species Supports the Forest Refuge Hypothesis. *Journal of Biogeography* 42(8):1443–54. https://doi.org/10.1111/jbi.12510.

Edwards SV, Potter S, Schmitt CJ, Bragg JG, Moritz C. 2016 Reticulation, Divergence, and the Phylogeography–Phylogenetics Continuum. *Proceedings of the National Academy of Sciences* 113(29):8025–32. https://doi.org/10.1073/pnas.1601066113.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ et al. 2006 Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. *Ecography* 29(2):129–51. https://doi.org/10.1111/j.2006.0906-7590.04596.x.

Faye A, Deblauwe V, Mariac C, Richard D, Sonké B, Vigouroux Y, Couvreur TLP. 2016 Phylogeography of the Genus *Podococcus* (Palmae/Arecaceae) in Central African Rain Forests: Climate Stability Predicts Unique Genetic Diversity. *Molecular Phylogenetics and Evolution* 105:126–38. https://doi.org/10.1016/j.ympev.2016.08.005.

Fuchs J, Bowie, RC. 2015 Concordant genetic structure in two species of woodpecker distributed across the primary West African biogeographic barriers. *Molecular phylogenetics and evolution* 88:64-74.

Hardy OJ, Born C, Budde K, Daïnou K, Dauby G, Duminil J, Ewédjé EBK et al. 2013 Comparative Phylogeography of African Rain Forest Trees: A Review of Genetic Signatures of Vegetation History in the Guineo-Congolian Region. *Comptes Rendus - Geoscience* 345(7–8):284–96. https://doi.org/10.1016/j.crte.2013.05.001.

Heuertz M, Duminil J, Dauby G, Savolainen V, Hardy OJ. 2014 Comparative Phylogeography in Rainforest Trees from Lower Guinea, Africa. *PLoS ONE* 9(1):e84307. https://doi.org/10.1371/journal.pone.0084307.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005 Very High Resolution Interpolated Climate Surfaces for Global Land Areas. *International Journal of Climatology* 25(15):1965–78. https://doi.org/10.1002/joc.1276.

Holbrook KM, Smith TB. 2000 Seed dispersal and movement patterns in two species of *Ceratogymna* hornbills in a West African tropical lowland forest. *Oecologia* 125:249–257. https://doi.org/10.1007/s004420000445

Huson DH, Bryant D. 2006 Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*. 23(2):254-267. https://doi.org/10.1093/molbev/msj030.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. 2016 HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-Throughput Sequencing Reads Using Target Enrichment. *Applications in Plant Sciences* 4(7):1600016. https://doi.org/10.3732/apps.1600016.

Jombart T. 2008 Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* 24(11):1403–5. https://doi.org/10.1093/bioinformatics/btn129.

Jombart T, Devillard S, Balloux F. 2010 Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations. *BMC Genetics* 11(1):94. https://doi.org/10.1186/1471-2156-11-94.

Katoh K, Standley DM. 2013 MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30 (4): 772–80. https://doi.org/10.1093/molbev/mst010.

Leite RN, Rogers DS. 2013 Revisiting Amazonian Phylogeography: Insights into Diversification Hypotheses and Novel Perspecitves. *Organisms Diversity & Evolution* 13(4):639–64. https://doi.org/10.1007/s13127-013-0140-8.

Leite, YLR, Costa LP, Loss AC, Rocha RG, Batalha-Filho H, Bastos AC, Quaresma VS et al. 2016 Neotropical Forest Expansion during the Last Glacial Period Challenges Refuge Hypothesis. *Proceedings of the National Academy of Sciences* 113(4):1008–13. https://doi.org/10.1073/pnas.1513062113.

Ley AC, Heuertz M, Hardy OJ. 2017 The Evolutionary History of Central African Rain Forest Plants: Phylogeographical Insights from Sister Species in the Climber Genus *Haumania* (Marantaceae). *Journal of Biogeography* 44(2):308–21. https://doi.org/10.1111/jbi.12902.

Lézine AM, Izumi K, Kageyama M, Achoundong G. 2019 A 90,000-year record of Afromontane forest responses to climate change. *Science* 363(6423):177-181.

Li H. 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21), 2987-2993.

Li H, Durbin R. 2009 Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Liu X, Fu YX. 2015 Exploring population size changes using SNP frequency spectra. *Nature Genetics* 47, 555-559.

Linder HP. 2001 Plant Diversity and Endemism in Sub-Saharan Tropical Africa. In *Journal of Biogeography*, 28:169–82. https://doi.org/10.1046/j.1365-2699.2001.00527.x.

Lowe AJ, Harris D, Dormontt E, Dawson IK. 2010 Testing Putative African Tropical Forest Refugia Using Chloroplast and Nuclear DNA Phylogeography. *Tropical Plant Biology* 3(1):50–58. https://doi.org/10.1007/s12042-010-9045-2.

Mace GM, Gittleman JL, Purvis A. 2003 Preserving the Tree of Life. *Science* 300 (5626):1707–9. https://doi.org/10.1126/science.1085510.

Maley J. 1996 The African Rain Forest - Main Characteristics of Changes in Vegetation and Climate from the Upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh Section B: Biological Sciences* 104:31–73. https://doi.org/10.1017/S0269727000006114.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K et al. 2010 The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Research* 20(9):254–60. https://doi.org/10.1101/gr.107524.110.20.

Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP. 2014 ENMeval: An R Package for Conducting Spatially Independent Evaluations and Estimating Optimal Model Complexity for Maxent Ecological Niche Models. Methods in Ecology and Evolution 5(11):1198–1205. https://doi.org/10.1111/2041-210x.12261.

Phillips SJ, Anderson RP, Schapire, RE. 2006 Maximum Entropy Modeling of Species Geographic Distributions. *Ecological Modelling* 190(3–4):231–59. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Piñeiro R, Dauby G, Kaymak E, Hardy OJ. 2017 Pleistocene Population Expansions of Shade-Tolerant Trees Indicate Fragmentation of the African Rainforest during the Ice Ages. *Proceedings of the Royal Society B: Biological Sciences* 284(1866):20171800. https://doi.org/10.1098/rspb.2017.1800.

Portik DM, Leaché AD, Rivera D, Barej MF, Burger M et al. 2017 Evaluating mechanisms of diversification in a Guineo Congolian tropical forest frog using demographic model selection. *Molecular Ecology* 26(19):5245-5263.

Ponder WF, Carter GA, Flemons P, Chapman RR. 2001 Evaluation of Museum Collection Data for Use in Biodiversity Assessment. *Conservation Biology* 15 (3): 648–57. https://doi.org/10.1046/j.1523-1739.2001.015003648.x.

Poulsen JR, Clark CJ, Smith TB. 2001 Seed Dispersal by a Diurnal Primate Community in the Dja Reserve, Cameroon. *Journal of Tropical Ecology* 17:787–808.

Radosavljevic A, Anderson RP. 2014 Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography* 41:629–643. https://doi.org/10.1111/jbi.12227

Raj A, Stephens M, Pritchard JK. 2014 FastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics* 197(2):573–89. https://doi.org/10.1534/genetics.114.164350.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018 Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Systematic Biology. syy032. doi:10.1093/sysbio/syy032

Rohland N, Reich D. 2012. Cost-Effective, High-Throughput DNA Sequencing Libraries for Multiplexed Target Capture. *Genome Research* 22(5):939–46. https://doi.org/10.1101/gr.128124.111.

Stamatakis A. 2014 RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* 30(9):1312–13. https://doi.org/10.1093/bioinformatics/btu033.

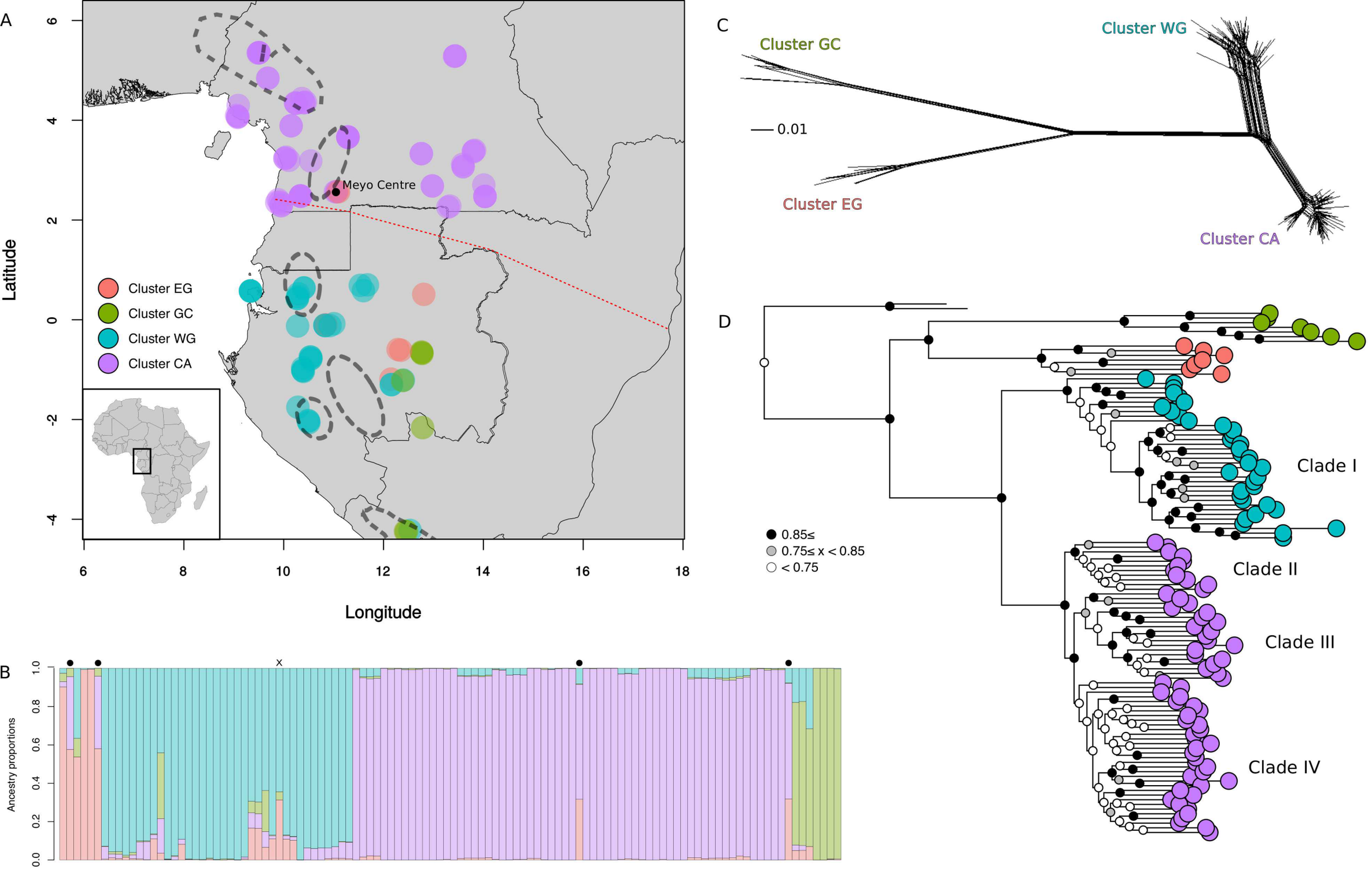
Thuiller W, Lafourcade B, Engler R, Araújo MB. 2009 BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32:369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

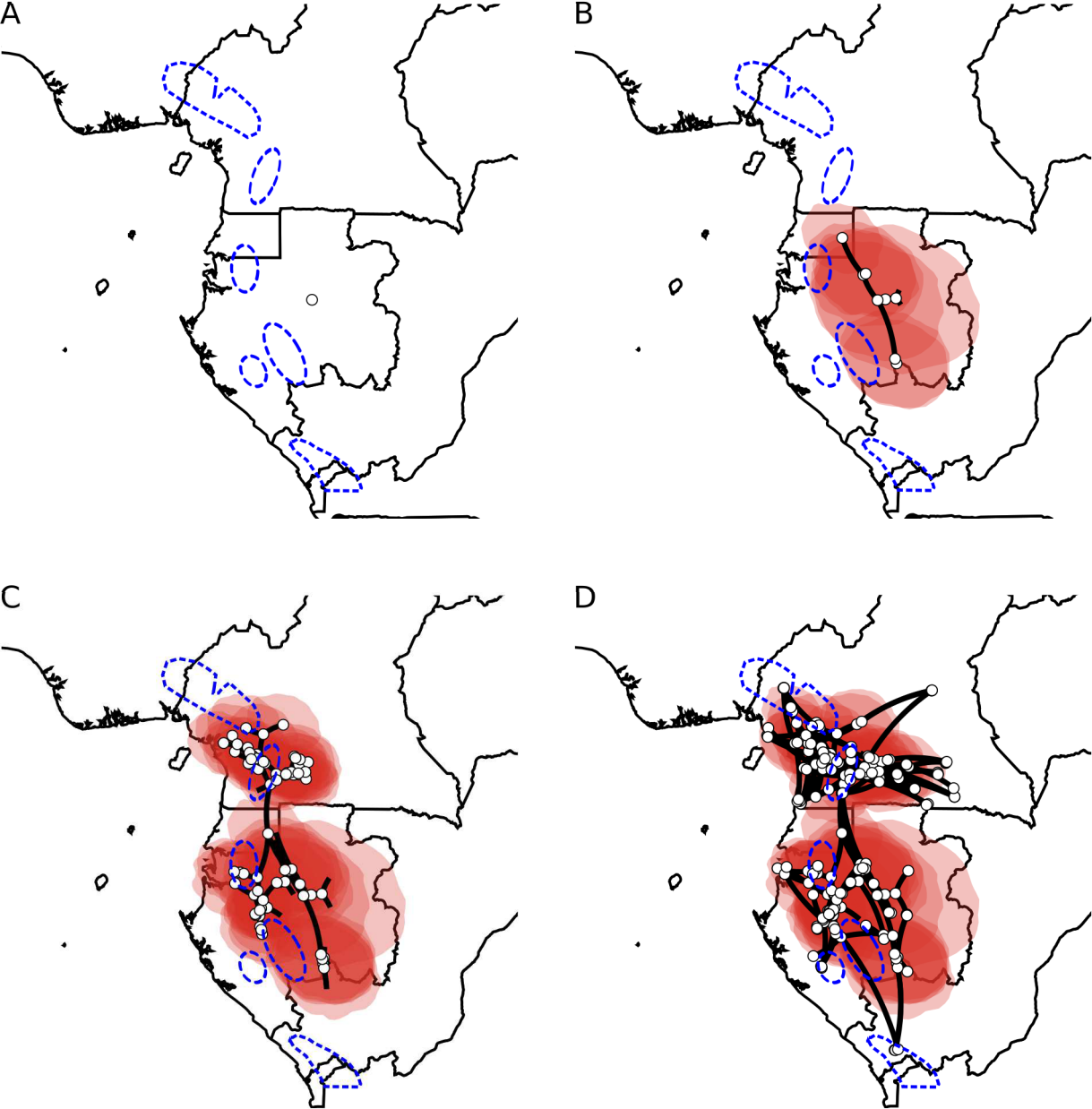
Turelli M, Barton NH, Coyne JA. 2001 Theory and speciation. *Trends in Ecology & Evolution* 16:330–343. https://doi.org/10.1016/S0169-5347(01)02177-2

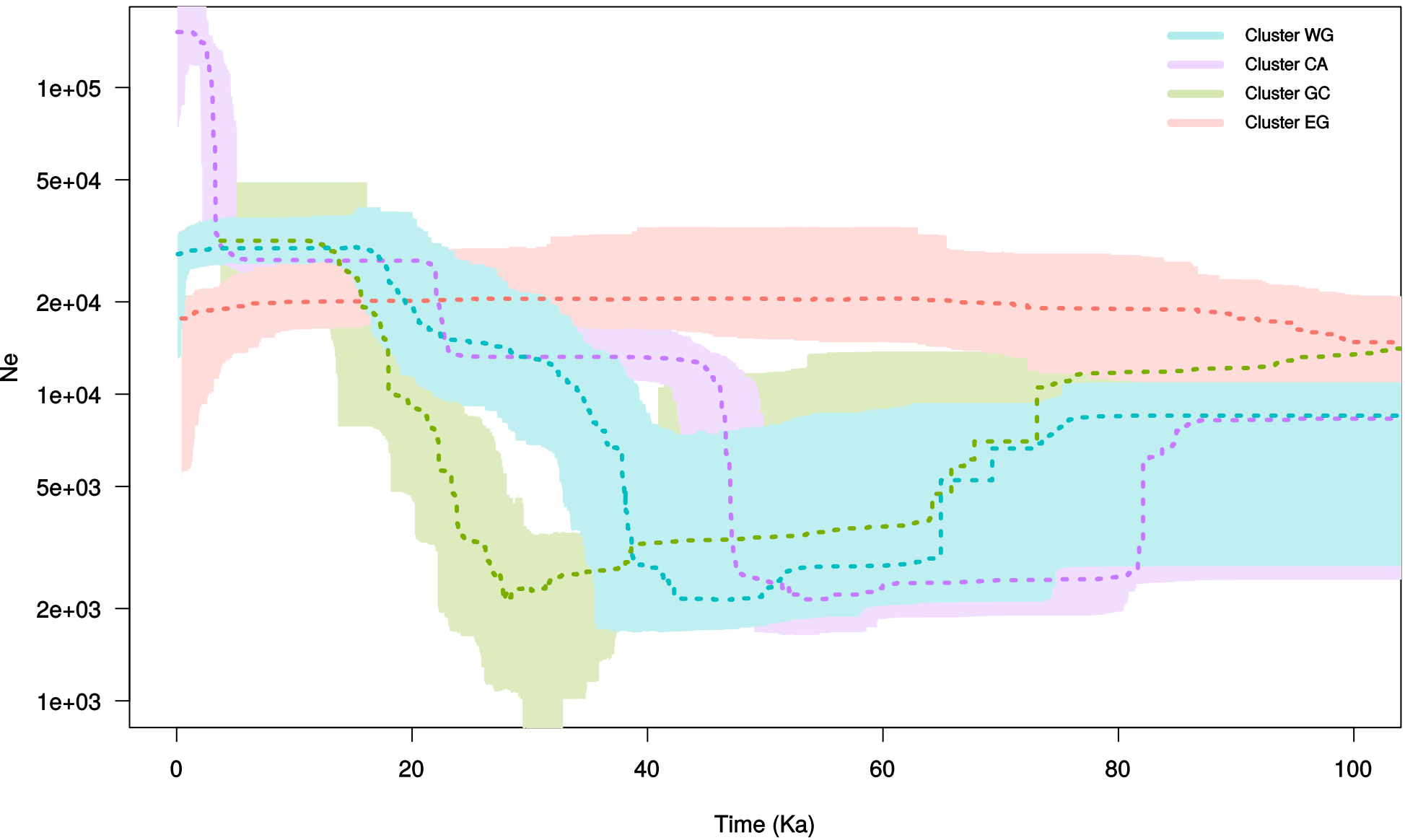
Versteegh CP, Sosef M. 2007 Revision of the African Genus *Annickia*. *Systematics and Geography of Plants* 77(1):91–118. https://doi.org/10.1600/036364411X553108.

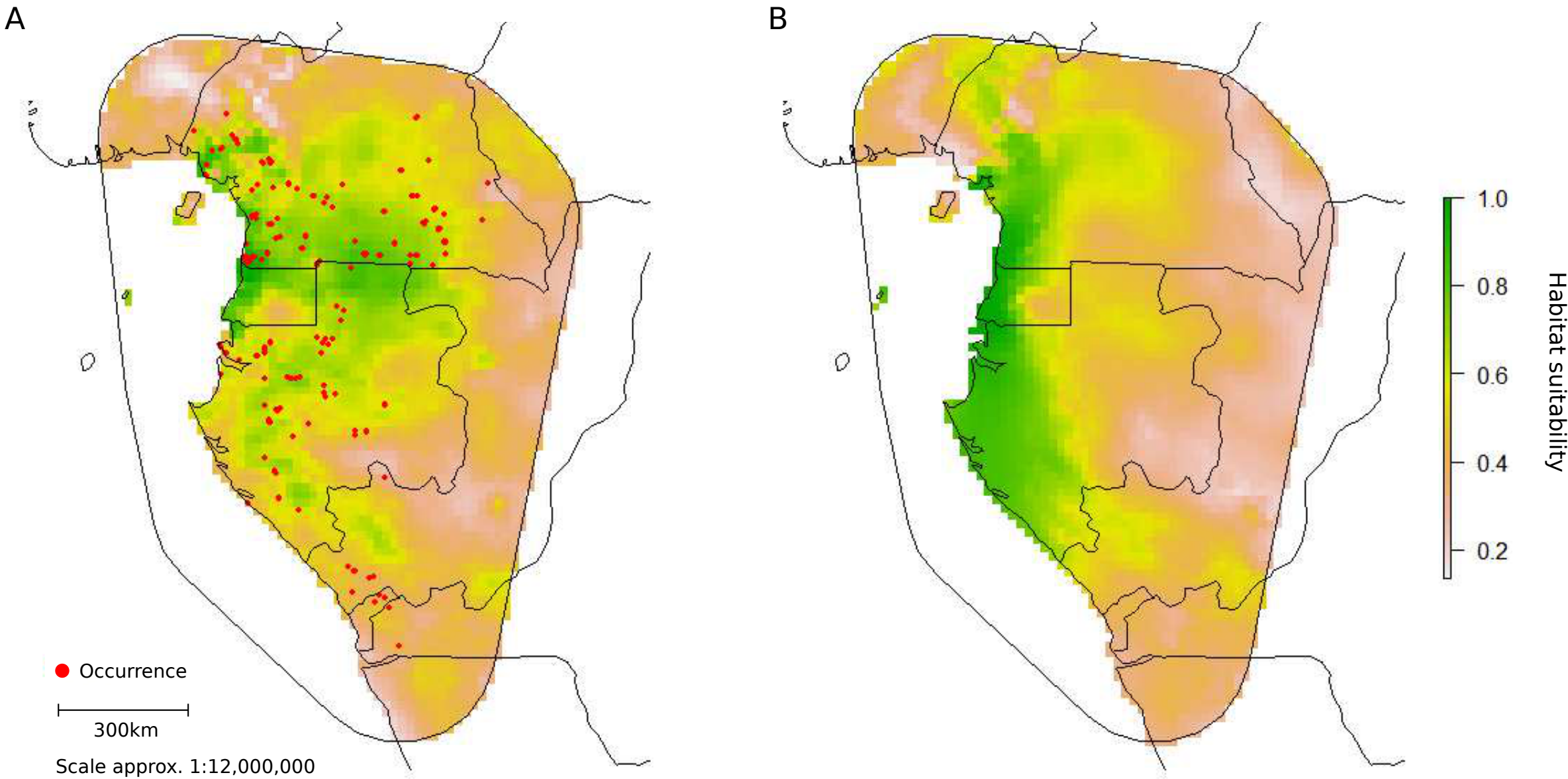
Vitorino LC, Lima-Ribeiro MS, Terribile LC, Collevatti RG. 2016 Demographical History and Palaeodistribution Modelling Show Range Shift towards Amazon Basin for a Neotropical Tree Species in the LGM. *BMC Evolutionary Biology* 16(1):1–15. https://doi.org/10.1186/s12862-016-0779-9.

Zhang C, Sayyari E, Mirarab S. 2017 ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10562 LNBI:53–75. https://doi.org/10.1007/978-3-319-67979-


A
Latitude
Longitude
Meyo Centre
Cluster EG
Cluster GC
Cluster WG
Cluster CA
B
Ancestry proportions
C
Cluster GC
Cluster WG
0.01
Cluster EG
Cluster CA
D
0.85≤
0.75≤ x < 0.85
< 0.75
Clade I
Clade II
Clade III
Clade IV


A
B
C
D


Cluster WG
Cluster CA
Cluster GC
Cluster EG
Ne
1e+05
5e+04
2e+04
1e+04
5e+03
2e+03
1e+03
0
20
40
60
80
100
Time (Ka)


A
B
Occurrence
300km
Scale approx. 1:12,000,000
1.0
0.8
0.6
0.4
0.2
Habitat suitability