Cradles and museums of generic plant diversity across tropical Africa

Léo-Paul Dagallier, Steven Janssens, Gilles Dauby, Anne Blach-Overgaard, Barbara Mackinder, Vincent Droissart, Jens-Christian Svenning, Marc Sosef, Tariq Stévart, David Harris, et al.

To cite this version:

Léo-Paul Dagallier, Steven Janssens, Gilles Dauby, Anne Blach-Overgaard, Barbara Mackinder, et al., Cradles and museums of generic plant diversity across tropical Africa. New Phytologist, Wiley, 2020, 225 (5), pp.2196-2213. 10.1111/nph.16293. hal-02407084
Cradles and museums of generic plant diversity across tropical Africa

Léo-Paul M. J. Dagallier1*, Steven B. Janssens2*, Gilles Dauby3, Anne Blach-Overgaard4,5, Barbara A. Mackinder6, Vincent Droissart3, Jens-Christian Svenning4,5, Marc S. M. Sosef2, Tariq Stévart2,7,8, David J. Harris6, Bonaventure Sonké9, Jan J. Wieringa10, Olivier J. Hardy11 and Thomas L. P. Couvreur11,10

1DIADE, IRD, University of Montpellier, 911 Avenue Agropolis, 34394, Montpellier, France; 2Meise Botanic Garden, Nieuwelaan 38, BE-1860, Meise, Belgium; 3AMAP, IRD, CIRAD, CNRS, INRA, University of Montpellier, Bd de la Lironde, 34398, Montpellier, France; 4Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000, Aarhus C, Denmark; 5Section for Econinformatics and Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000, Aarhus C, Denmark; 6Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, UK; 7Herbarium et Bibliothèque de Botanique Africaine, Université Libre de Bruxelles, Boulevard du Triomphe, B-1050, Bruxelles, Belgium; 8Africa & Madagascar Department, Missouri Botanical Garden, St Louis, MO 631109, USA; 9Laboratoire de Botanique et d’Ecologie, Département des Sciences Biologiques, École Normale Supérieure, Université de Yaoundé I, BP 047, Yaoundé, Cameroon; 10Naturalis Biodiversity Center, Darwinweg 2, 2333 CR, Leiden, the Netherlands; 11Evolutionary Biology and Ecology, Faculté des Sciences, Université Libre de Bruxelles, Av. F.D. Roosevelt 50, 1050, Brussels, Belgium

Summary

• Determining where species diversify (cradles) and persist (museums) over evolutionary time is fundamental to understanding the distribution of biodiversity and for conservation prioritization. Here, we identify cradles and museums of angiosperm generic diversity across tropical Africa, one of the most diverse regions on Earth.
• Regions containing nonrandom concentrations of young (neo-) and old (paleo-) endemic taxa were identified using distribution data of 1719 genera combined with a newly generated time-calibrated mega-phylogenetic tree. We then compared the identified regions with the current network of African protected areas (PAs).
• At the generic level, phylogenetic diversity and endemicism are mainly concentrated in the biogeographically complex region of Eastern Africa. We show that mountainous areas are centres of both neo- and paleo-endemism. By contrast, the Guineo-Congolian lowland rain forest region is characterized by widespread and old lineages. We found that the overlap between centres of phylogenetic endemicism and PAs is high (> 85%).
• We show the vital role played by mountains acting simultaneously as cradles and museums of tropical African plant biodiversity. By contrast, lowland rainforests act mainly as museums for generic diversity. Our study shows that incorporating large-scale taxonomically verified distribution datasets and mega-phylogenies lead to an improved understanding of tropical plant biodiversity evolution.

Introduction

Understanding the ecological and evolutionary processes that shape biodiversity is crucial to establishing conservation priorities (McNeely et al., 1990; Myers et al., 2000). Endemic taxa are defined as being geographically restricted and their spatial concentration often highlights evolutionary processes. Traditional methods have mainly focused on using taxonomic diversity (i.e. taxon richness) to depict biodiversity patterns. However, adding the evolutionary history allows a more detailed picture of such patterns (Faith, 1992; Forest et al., 2007). Combining data from phylogenetic trees and taxon diversity enables the calculation of phylogenetic diversity (PD), that is, the total length of branches of a phylogenetic tree connecting taxa in a particular area (Faith, 1992). Regions with high PD contain important levels of evolutionary history (Sechrest et al., 2002; Forest et al., 2007). Phylogenetic endemicism (PE) is the spatial restriction of phylogenetic lineages, that is, the degree to which a part of the evolutionary history is range-restricted (Rosauer et al., 2009). In this context, two types of endemic taxa can be distinguished (Kruckeberg & Rabinowitz, 1985; Nekola, 1999; Ferreira & Boldrini, 2011; Mishler et al., 2014). Neo-endemics are taxa that diverged recently (i.e. sustained by short phylogenetic branches) and whose narrow range is generally linked to a lack of time to disperse and expand. By contrast, paleo-endemics are taxa that diverged much earlier and which are also range-restricted, either

*These authors contributed equally to this work.
because of range reduction or because of a lack of adaptation or dispersal opportunities. Spatial concentrations of neo- or paleo-endemic taxa are referred to as centres of neo-endemism or paleo-endemism, respectively (Fig. 1a,b), and concentrations of both neo- and paleo-endemic taxa are called centres of mixed-endemism (Fig. 1c) (Mishler et al., 2014). Centres of neo-endemism can be viewed as ‘cradles of diversity’, that is, regions with high speciation rates, and are centres of diversification, whereas centres of paleo-endemism can be viewed as ‘museums of diversity’, where taxa have persisted over time. By studying the distribution of endemic taxa together with their evolutionary history, we can generate hypotheses about historical processes that shaped biodiversity (Mishler et al., 2014; Thornhill et al., 2016, 2017; Scherson et al., 2017). Such knowledge is crucial for the improvement of conservation management in the context of an unprecedented biodiversity decline (Humphreys et al., 2019).

Tropical Africa contains some of the most biologically diverse regions of the world (Myers et al., 2000; Linder, 2001; Küper et al., 2004). Although it has lower species diversity when compared with the Neotropics and Southeast Asia, tropical Africa still harbours an estimated 30 000–35 000 species of plants (Couvreur, 2015). High species richness and centres of taxonomic endemism are found in Cameroon, Gabon, eastern Democratic Republic of the Congo (DRC), and eastern Tanzania (Linder, 2001; Küper et al., 2004; Sosef et al., 2017). Furthermore, Africa contains the second largest continuous tropical rainforest block after the Amazon basin (Malhi et al., 2013). Heterogeneous topography and climate across the continent gave rise to intricate biogeographical patterns forming numerous bioregions and transition zones (White, 1983; Linder et al., 2012; Fayolle et al., 2014; Droissart et al., 2018). Tropical African plant diversity patterns have been relatively well explored by taxonomic approaches (Brenan, 1978; White, 1983; Lovett et al., 2000; Linder, 2001; Küper et al., 2004; Klopper et al., 2007; Droissart et al., 2018). However, in an evolutionary context our understanding about where diversity originated and where it is maintained is limited.

The lowland rainforests of the Guineo-Congolian region have been suggested to be Pleistocene cradles of diversity (Table 1) with glacial forest refugia triggering speciation (Diamond & Hamilton, 1980; Maley, 1996; Robbrecht, 1996; Sosef, 1996; Huntley & Voelker, 2016). However, molecular dating estimates in several plant and animal clades have shown that many species originated before the start of the Pleistocene, suggesting that forest refugia patterns reflect mainly Pliocene climatic fluctuation (Plana et al., 2004; Voelker et al., 2010). This led to the idea that lowland rainforests are museums of diversity (Table 1, Murienne et al., 2013). In other studies, these lowland regions have been found to be concentrations of old and widespread lineages, especially in birds (Table 1; Fjeldså, 1994; Fjeldså & Lovett, 1997; Fjeldså & Bowie, 2008; Fjeldså et al., 2012).

Cradles of diversity have also been suggested in montane regions of Africa (Table 1), such as the Albertine Rift and the Eastern Arc Mountains (Fjeldså & Lovett, 1997; Roy, 1997; Dimitrov et al., 2012), thereby supporting a ‘montane speciation model’ (Roy, 1997). Indeed, mountains are important centres of global diversification that represent heterogeneous and dynamic landscapes, as well as isolation and reconnection processes that are linked with past climate changes (Peterson et al., 1997; Fjeldså et al., 2012; Hoorn et al., 2013; Antonelli et al., 2018; Muellner-Riehl, 2019). This diversification is the consequence of different mechanisms, for example, fragmented habitats promoting allopatric divergence by isolation of small populations (Hughes & Eastwood, 2006; Schwery et al., 2014; Merckx et al., 2015; see Wen et al., 2014 for a review of all the mechanisms). Mountains have also been acknowledged as museums of diversity as their complex topography supports relatively stable habitats by buffering climatic fluctuations and by allowing the movement of habitats over relatively short altitudinal distances to respond to temperature shifts, and thereby allowing their persistence over long periods of time (Peterson et al., 1997; Fjeldså & Lovett, 1997; Hewitt, 2000; Tzedakis et al., 2002; Loarie et al., 2009; López-Pujol et al., 2011; Tolley et al., 2011). However, studies focusing on bird and few plant clade distributions have identified co-occurrence of neo- and paleo-endemic taxa in African mountain areas (Wasser & Lovett, 1993; Fjeldså & Lovett, 1997;
Table 1 Hypotheses and categorical analysis of neo- and paleo-endemism (CANAPE) predictions tested across the tropical African angiosperm flora.

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>CANAPE predictions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowland rainforests are museums of diversity</td>
<td>Significant concentration of old and range-restricted (paleo-endemics) or old and</td>
<td>Fjeldså (1994); Fjeldså &amp; Lovett (1997); Fjeldså &amp; Bowie (2008); Fjeldså et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>widespread taxa in lowland rainforest SUs</td>
<td></td>
</tr>
<tr>
<td>Lowland rainforests are cradles of diversity (past refugia as ‘species pump’)</td>
<td>Significant concentration of young and range-restricted taxa (neo-endemic) in lowland</td>
<td>Maley (1996); Sosef (1996); Huntley &amp; Voelker (2016)</td>
</tr>
<tr>
<td>Montane regions are centres of diversification ('montane speciation model'), i.e. cradles of diversity</td>
<td>Significant concentrations of neo-endemic taxa (centres of neo-endemism) in montane SUs</td>
<td>Roy (1997); Dimitrov et al. (2012); Schwyer et al. (2014)</td>
</tr>
<tr>
<td>Montane regions are centres of persistence of lineages, i.e. museums of diversity</td>
<td>Significant concentration of paleo-endemic taxa (centres of paleo-endemism) in montane</td>
<td>Fjeldså &amp; Lovett (1997); López-Pujol et al. (2011); Tolley et al. (2011)</td>
</tr>
<tr>
<td>Montane regions are centres of diversification and persistence of lineages</td>
<td>Significant concentration of both paleo- and neo-endemic taxa (centres of mixed-endemism) in montane SUs</td>
<td>Wasser &amp; Lovett (1993); Fjeldså &amp; Lovett (1997); López-Pujol et al. (2011)</td>
</tr>
</tbody>
</table>

SU, sampling unit (100 km²).

López-Pujol et al., 2011), suggesting that mountains are both cradles and museums of diversity (Fjeldså & Lovett, 1997) (Table 1).

For plants, these hypotheses have not yet been tested using both large distribution datasets and dated molecular phylogenies. Botanical datasets recording African plant distribution information are becoming increasingly comprehensive (Stropp et al., 2016; Dauby et al., 2016) and are significantly improving our understanding of biodiversity and phytogeographical patterns across tropical Africa (Sosef et al., 2017; Droissart et al., 2018). In parallel, the incorporation of evolutionary history information is now possible thanks to the generation of large phylogenies using existing data available on GenBank (Hinchliff & Smith, 2014; Smith & Brown, 2018). Together, this allows us to test the described hypotheses at new levels of precision.

Here, we evaluate five main hypotheses (Table 1) related to the evolutionary and historical processes shaping tropical African biodiversity. We identify significant centres of neo- and paleo-endemism across the entire tropical African angiosperm flora at generic level based on a taxonomically verified distribution dataset and a newly generated mega-phylogeny of angiosperms. At the generic level, we identified areas of significant PD, PE and centres of neo-, paleo- or mixed-endemism across tropical Africa to answer the following questions: do previously identified areas of high taxon diversity display more PD and PE than expected by chance? Are African montane regions centres of diversification, centres of lineage persistence, or both? Are African lowland rainforests museums or cradles of diversity? Finally, as centres of PE are of high conservation interest, we explored the overlap between the protected area network across tropical Africa and the different centres of endemism identified.

Materials and Methods

Study area

Here, tropical Africa is defined by the ecoregions characterized by Olson et al. (2001) with the ‘south Saharan steppe and woodlands’ ecoregion and Mauritania as northern limit and the ‘Drakensberg montane grassland’, ‘Highveld grassland’, ‘Kalahari Acacia Baikiaea woodland’, ‘Kalahari xeric savannah’ and ‘Namibian savannah woodlands’ ecoregions as southern limits (Sosef et al., 2017; Droissart et al., 2018). This study includes continental Africa and the Guinean Gulf islands (Bioko, Sao Tomé and Principe), but excludes Madagascar (Fig. 2).

Taxonomic level

We conducted this study on tropical African plants at the genus level. This taxonomic level presents advantages and drawbacks. Conceptually, genera should represent clear morphological differences of the flora and older evolutionary events than species, which is the main focus of this study. From a methodological point of view, generic names are comparatively stable taxonomic entities, which circumvent biases as a result of taxonomic uncertainties and potential misidentifications of species (but see Discussion). Moreover, mega-phylogenies are better resolved at the genus than at the species level (Hinchliff & Smith, 2014; Smith & Brown, 2018). Finally, genus-level studies are useful for addressing broad-scale biodiversity patterns (Forest et al., 2007; Thornhill et al., 2016; Scherson et al., 2017). A drawback is that really recent species cradles will be missed by this approach (see Discussion).

Datasets

Plant occurrence data We used the RAINBIO dataset (Dauby et al., 2016), a database of georeferenced occurrences of African vascular plants with comparatively high taxonomic quality. It represents a compilation of 614 022 botanical records collected from 1782 to 2015 providing distribution information for 25 356 species and 3158 genera within 273 families. This represents c. 89% of all known tropical African plant species.

Species occurrence data were projected using the Africa Albers Equal Area Conic coordinate reference system (ESRI:102022, http://spatialreference.org/). For that, we transformed coordinates from decimal degrees into metres, the base unit of this coordinate reference system. The study area was divided into 100 × 100 km equal-area square sampling units (SUs).
Using the RAINBIO dataset, we applied three filters, leading to the RAINBIO_{filtered} dataset we used for all our analyses:

- **Angiosperm filter.** As our study and phylogeny focused on Angiospermae, Pteridophyta and Gymnospermae were excluded.
- **Edge effect filter.** When considering patterns of relative endemism in a delimited zone, bias towards over-representation of artificial endemic taxa at the boundaries, is frequent. This artefact is a result of the rough cut within the geographic distribution of the taxa occurring both in and outside the study area, thereby artificially leading to the identification of these taxa as endemics although some of them may occur on both sides of the boundary. In order to avoid this edge effect, we kept only genera with > 90% of their occurrences falling within our study area.
- **Occurrence filter.** Specimen density in RAINBIO is heterogeneous across tropical Africa. Poorly sampled SUs occur in poorly known regions and thus are more likely to include taxa that were collected only once (Sosef et al., 2017). This can artificially increase the endemism of these SUs. To avoid potential bias caused by poorly sampled SUs, we included in our analyses only

Fig. 2 Map of Africa. The red dotted line delineates the study zone. Green shading delineates the Guineo-Congolian bioregion as defined by Droissart et al. (2018). Brown shading delineates ‘montane regions’ (GMBA mountain inventory_V1.2; Körner et al., 2017). Locations cited in the main text: AR, Albertine Rift; CVL, Cameroon Volcanic Line; DG, Dahomey Gap; EAM, Eastern Arc Mountains; EH, Ethiopian Highlands; GG, Guinean Gulf; Kat., Katanga; Map., Maputaland; S. Malawi, South Malawi. The map is projected in the Africa Albers Equal Area Conic coordinate reference system (ESRI:102022).
SU: high RPD indicates an over-representation of long-branched genera whereas low RPD indicates over-representation of short-branched genera (Mishler et al., 2014):

$$\text{RPD} = \frac{\text{PD}_{\text{orig}}}{\text{PD}_{\text{theor}}}$$

Phylogenetic endemism (PE) is the sum of branch lengths, weighted by the inverse of the branch’s range ($R_g$), for each branch ($c$) connecting the roots of the phylogenetic tree to the genera (tips of the phylogeny) within a SU (Rosauer et al., 2009). $R_g$ is measured as the number of SUs in which the branch $c$ occurs:

$$\text{PE} = \sum_{c \in C} \frac{L_c}{R_g}$$

Relative PE (RPE) is the ratio of the PE measured on the original phylogenetic tree ($\text{PE}_{\text{orig}}$) divided by the PE measured on a theoretical tree ($\text{PE}_{\text{theor}}$). The theoretical tree has the same topology as the original but all branches are of equal length (Mishler et al., 2014). This index is the base for the categorical analyses of neo- and paleo-endemism (see later):

$$\text{RPE} = \frac{\text{PE}_{\text{orig}}}{\text{PE}_{\text{theor}}}$$

Null model – randomizations

To test which SUs had significantly higher or lower observed values than expected given the genus richness of the SU and the geographical range of all genera, we ran 999 randomizations using BIODIVERSE v.2.0 (Laffan et al., 2010). For every run, the algorithm randomly reassigned all the genera to each SU without replacement. In order to respect both genus richness patterns and the geographical range of taxa, we constrained the procedure to keep the original number of genera in each SU and to keep the original number of SUs in which each genus occurs (‘rand_struct’ option in BIODIVERSE).

Then, for each metric, the observed value is compared with the 999 randomization values. Significantly greater or lower than expected was defined as being > 97.5% or < 2.5% of the random values, respectively (two-tailed test, $x = 0.05$). High degrees of significance were established for observed values > 99% or < 1% of the random values.
This randomization test was carried out for WE, PD, PE and RPD, resulting in the assignment of a significance class for each of these metrics for each SU: significantly very low (<0.01), significantly low (<0.025), ‘not significant’, significantly high (>0.975) or significantly very high (>0.99).

Categorical analysis of neo- and paleo-Endemism (CANAPE)

The categorical analysis of neo- and paleo-endemism (CANAPE) discriminates SUs with significantly high PE in neo- or paleo-endemism based on taxon occurrences and the dated phylogenetic tree (Mishler et al., 2014). First, we calculate the PE based on the original phylogenetic tree (PEorig) and the PE based on the theoretical tree (PEtheor, branches of equal length) for each randomization of genera composition and subsequently compare those with the observed values. This allows us to select SUs that are centres of high PE, that is, showing either significantly high PEorig or PEtheor, or both (observed values of PEorig and PEtheor >95% of the random values; one-tailed test, z = 0.05). Second, significant SUs are categorized into four nonoverlapping categories. If the RPE ratio (PEorig/PEtheor) is higher than expected (two-tailed test, z = 0.05), the SU contains significantly more endemic genera on long branches and is identified as a centre of ‘paleo-endemism’. If the RPE ratio is lower than expected (two-tailed test, z = 0.05), the SU contains significantly more endemic genera on short branches and is identified as a centre of ‘neo-endemism’. If the RPE is not significantly high or low, but both PEorig and PEtheor are significantly high at the level of z = 0.05, SU is tagged as centre of ‘mixed-endemism’ (i.e. a mix of endemic genera with both long and short branches). Finally, if a ‘mixed-endemism’ SU presents both high PEorig and PEtheor at the level of z = 0.01, the SU is tagged as a centre of ‘super-endemism’ (i.e. highly significant concentration of endemic long and short branches).

As CANAPE results might be sensitive to the SU size, CANAPE was also carried out with four different SU sizes across the study zone: 50, 75, 200 and 300 km square. The SUs contain at least 50, 75, 200 and 300 occurrences, respectively.

Overlap with montane regions

A ‘montane’ region can be defined according to differences in elevation, relief or steepness. We adopted the montane definition of Körner et al. (2011) based on the concept of steepness, the basic and consistent feature that reflects landscape heterogeneity found in mountains. Indeed, a definition based solely on elevation would consider highland plateaus as mountains (e.g. Tibetan Plateau, Altiplano) even if they are relatively flat landscapes. For continental-scale purposes, steepness is estimated with the ruggedness, that is, the elevation range between a grid cell and the eight adjacent cells in the grid (Körner et al., 2011). A cell is then considered as a ‘mountain’ if its ruggedness exceeds 200 m. Global mountain biodiversity assessment (GMBA) provides ruggedness data for 2.5-arcminute (c. 4.6 × 4.6 km at the equator) plots across the globe (GMBA mountain definition_V1.0 database; Körner et al., 2011) and shapefiles delineating ‘montane areas’ (GMBA mountain inventory_V1.2; Körner et al., 2017). For each SU, we calculated the mean ruggedness, the percentage of montane plots (i.e. percentage of 4.6 × 4.6 km plots whose ruggedness exceeds 200 m) and the percentage of overlap with ‘montane areas’ defined by GMBA.

By contrast, we defined ‘lowlands’ on the elevation feature to avoid considering a highland plateau as a lowland area. Elevation data were retrieved from the digital elevation modelling data produced by the NASA Shuttle Radar Topographic Mission (SRTM; http://srtm.csi.cgiar.org/). We downloaded the 250 m resolution (SRTM v.4.1) data in GeoTiff raster format, and extracted a raster file of the tropical Africa region using QGIS 2.18. From this raster file, we calculated the mean elevation for each SU.

We compared the distribution of the mean ruggedness, the percentage of montane plots and the percentage of overlap with ‘montane areas’ of the SUs across:

1. the three categories of the PD results (significantly high or very high PD; not significant; significantly low or very low PD);
2. the five categories of the CANAPE results (neo-endemism; paleo-endemism; mixed-endemism; super-endemism, not significant).

We compared the distribution of the mean elevation of the SUs across the three categories of the RPD results (significantly high or very high RPD; not significant; significantly low or very low RPD).

As normality of the residuals (one of the fundamental assumptions for ANOVAs) was not fulfilled for the three comparisons (results not shown), we used a nonparametric Kruskal–Wallis test. If at least one of the distributions was significantly different from the others (P < 0.05), then Wilcoxon pairwise comparison (two-tailed tests with Holm’s correction) was performed to disentangle which categories were significantly different from each other (P < 0.05).

CANAPE and protected areas network overlap

CANAPE results were overlapped with the African protected areas network retrieved from the World Database on Protected Areas (WDPA, https://protectedplanet.net/, accessed June 2018). Protected areas exclusively related to marine or faunal protection (e.g. ‘bird reserve’, ‘hunting reserve’, ‘faunal reserve’) were excluded. The number of SUs that contained protected areas (PA) were counted for any PA and for the most restrictive PAs in terms of conservation (reported as International Union for Conservation of Nature (IUCN) categories Ia, Ib and II; PAs that have a main focus on biodiversity) (IUCN, 2008).

Results

Datasets

Plant occurrence The RAINBIOfiltered dataset contained 547 273 occurrences of 2345 genera distributed across 638 equal-area square SUs. GR is unevenly distributed across tropical Africa, ranging from 39 to 853 genera per SU (Fig. 3). The most diverse SUs are found around the Guinean Gulf, particularly in Gabon, western Cameroon, in the Dahomey Gap, the Eastern
Arc Mountains and the Albertine Rift region, and to a lesser extent isolated SUs in the DRC and the Ethiopian Highlands.

**Phylogeny** The unfiltered phylogenetic tree contained 36,234 species and 8,357 genera. Based on the RAINBIO\_filtered dataset, this phylogenetic tree was pruned and included a final total of 1,719 genera. Of these, 36% (618/1,719) were not monophyletic in the unfiltered phylogenetic tree. In this case, we randomly sampled one of the branches of the nonmonophyletic genera. The tropical African flora at generic level represents 113,687 Myr of cumulative evolutionary history (i.e. the cumulative age of the taxa, estimated by the sum of all branch lengths on the tree).

**Biodiversity analyses**

**Weighted endemism** Raw values of WE (Fig. 4a) followed a similar pattern to GR (Fig. 3) ($R^2 = 0.764$; Supporting Information Fig. S1). Significantly high WE SUs (Fig. 4b) are located in the eastern parts of tropical Africa, in Ethiopia, Kenya-Tanzania, Katanga (southeast DRC), Maputaland and scattered in southeast Africa. This indicated that these regions harbour more endemic genera than expected from random. By contrast, regions in Central and West Africa (i.e. the Guineo-Congolian region) are characterized by a significantly less WE than expected at random (Fig. 4b).

![Genus richness map of the tropical African angiosperms](image_url)
Fig. 4 Weighted endemism (WE) maps of 2345 tropical African angiosperm genera. (a) Observed values of WE; (b) significant WE results from the randomization test. Sampling units (SUs): 100 × 100 km. Red SUs contain less WE than expected; blue SUs contain more WE than expected; beige SUs are not significant; white regions contain <100 records and were not included in the analyses.
Randomization tests for PD  Calculations of phylogenetic metrics were conducted on the 1719 genera found both in the phylogenetic tree and in the RAINBIOfiltered dataset. PD is highly correlated with genus richness ($R^2 = 0.992$; Figs S2, S3). Higher PD than expected was found in 22 SUs (Fig. 5a), located in mountainous areas (Fig. S4; Table S1) around Lake Victoria, near the Eastern Arc Mountains, in northern Kivu, in Central Ethiopia and in the Cameroon Volcanic Line (Fig. 5a). Conversely, 230 SUs presented lower PD than expected. These are distributed broadly across the study zone, with the main centres

Fig. 5 Maps of significant (a) phylogenetic diversity (PD) and (b) relative phylogenetic (RPD) resulting from the randomization tests for 1719 tropical African angiosperm genera. Sampling units (SUs) are 100 x 100 km squares. Red SUs contain less PD or RPD than expected; blue SUs contain more PD or RPD than expected; beige SUs are not significant; white regions contain < 100 records and were not included in the analyses.
occurring in the Dahomey Gap and to the west, in northern Cameroon, central Gabon and coastal Congo, and across eastern Africa (Fig. 5a).

SUs of significantly high RPD, representing concentrations of long-branched (i.e. old) genera, are clustered in the Guineo-Congolian region, western African forests and scattered in eastern Africa (Fig. 5b). SUs of significant low RPD, representing concentrations of short-branched (i.e. young) genera, are scattered in the north of the Guineo-Congolian region and in eastern Africa (Fig. 5b).

**Categorical Analyses of neo- and paleo-endemism** As for PD and GR, PE is highly correlated with WE ($R^2 = 0.985$; Fig. S5) and presents a similar pattern of significance as the WEs (Fig. S6). A total of 155 SUs were revealed by the CANAPE analyses as containing significantly more PE than expected (red, blue and purple SUs in Fig. 6). These are concentrated mainly in the eastern part of Africa. Areas of mixed- and super-endemism are the most common significant CANAPE SUs (58 and 50, respectively). Only 11 SUs are classified as areas of neo-endemism and these are generally surrounded by areas of mixed-, super- or paleo-endemism. As for neo-endemism, we found the 36 SUs of paleo-endemism surrounded by mixed- and super-endemism SUs. The Ethiopian Highlands show a concentration of mixed, super and paleo-endemic SUs. From central Kenya towards

![Map of significant phylogenetic endemism (PE) identified by the categorical analysis of neo- and paleo-endemism (CANAPE) for 1719 tropical African angiosperm genera. Sampling units (SUs) are 100 × 100 km squares. Coloured SUs present significantly high PE: red, centres of neo-endemism; blue, centres of paleo-endemism; violet, centres of mixed-endemism (i.e. mix of neo- and paleo-endemism), with darker violet indicating centres of super-endemism; beige, not significant; white regions contain < 100 records and were not included in the analyses. Figures below the legend indicate the number of SUs included in each category.](image-url)
northern Tanzania, a similar cluster is found, with some neo-endemic SUs. Maputaland also concentrates PE (mostly paleo-). Smaller clusters occurred across eastern Africa, such as in coastal and southwestern Tanzania, Katanga and other scattered areas in southeast Africa. Concentrations of significant PE in West and Central Africa are very low and mainly located in northern Cameroon, Bioko Island (Equatorial Guinea, Guinean Gulf), Guinea and Sierra Leone.

Different SU sizes give a similar overall CANAPE results, with higher coverage of the study zone with large SUs (200 and 300 km square) and lower coverage with small SUs (50 and 75 km square) (Fig. S7).

Montane and lowland regions The mean elevation (Fig. 7) differs significantly between the RPD randomization categories (Kruskal-Wallis test, $P < 0.05$; Table S2). The high RPD SUs are significantly lower in elevation than nonsignificant and low RPD SUs (Wilcoxon pairwise comparison, $P < 0.05$; Table S2). Some SUs are outliers of high RPD distribution (solid dots), meaning these few SUs are occurring at high elevation.

The distribution of the mean ruggedness (Fig. 8), the percentage of montane plots (Fig. S8) and the percentage of overlap with ‘montane areas’ (Fig. S6) differ significantly between the CANAPE categories ($P < 0.05$; Tables S3–S5). The mean ruggedness of mixed- and super-endemic SUs is significantly higher than the ruggedness of nonsignificant SUs (Fig. 8) (Wilcoxon pairwise comparison, $P < 0.05$; Table S3). The mean ruggedness of neo- and paleo-endemic SUs is not different from all the others (Wilcoxon pairwise comparison, $P < 0.05$; Table S3).

The distributions of the percentage of montane plots and of the overlap with ‘montane areas’ of the SUs show similar differences across the five CANAPE categories, except for paleo-endemic SUs that are significantly different from nonsignificant SUs (Figs S8, S9; Tables S4, S5).

Protected areas network overlap The overlap between the African PA network and centres of significant PE detected by CANAPE is partial (Fig. 9). More than 85% of the CANAPE cells contain at least a part of a PA, and > 42% of the CANAPE cells partially overlap with one of the most restrictive PA IUCN categories (Ia, Ib or II). The whole PA network covers 18.7% of the CANAPE cell surface, and 5.9% when considering only IA, Ib and II PA categories (Tables S6, S7).

Discussion

Significant concentrations of evolutionary history and taxonomic endemism

In this study, we combined for the first time at the near continental level a large floristic dataset within a phylogenetic framework to examine more than 113 000 Myr of evolutionary history of tropical African angiosperms. Our GR map (Fig. 3) corroborates previous studies highlighting Cameroon, Gabon, the Albertine Rift, the Eastern Arc Mountains and coastal Tanzania as containing high taxon richness (Linder, 2001; Küper et al., 2004; Lovett et al., 2005; Plumptre et al., 2007; Sosef et al., 2017).
Despite the correlation between GR and PD (Fig. S2), our randomization procedure reveals that in some regions, PD is higher or lower than expected by chance (Fig. 5a). A decoupled pattern between GR and PD was also observed in South Africa (Forest et al., 2007). Across tropical Africa, 230 out of 638 SUs showed significantly lower PD than expected by chance (phylogenetic clustering). These are located in a lowland rainforest in Gabon and central DRC, but also in drier regions such as the Dahomey Gap, northern Cameroon, savannah regions in western DRC and in eastern Africa (Fig. 5a). In savannah (i.e. wooded grassland) regions, phylogenetic clustering is probably caused by the dominance of Poaceae and Fabaceae (Jacobs, 2004). In lowland tropical rainforest regions, phylogenetic clustering suggests evolutionary conservatism in tropical forest adaptations.

Analogous phylogenetic clustering has been observed in some tropical forest and arid regions in Borneo, Central America, Australia and Chile (Webb, 2000; Swenson et al., 2007; Thornhill et al., 2016; Scherson et al., 2017). By contrast, only six SUs had significantly higher PD than expected by chance (phylogenetic overdispersion). These are mainly located in montane areas (Fig. S5) of the Cameroon Volcanic Line, Ethiopian Highlands, northern Kivu, central Kenya, the Eastern Arc Mountains and southern Malawi (Fig. 5a). Phylogenetic overdispersion may be related to competitive exclusion of closely related genera whose niches show a large overlap (Webb et al., 2002), extinction (Kissling et al., 2012; Couvreur, 2015), or colonization of phylogenetically distinct lineages. Another reason for this pattern could be linked to the complex topography of montane areas.

Fig. 9 Map of the protected areas of Africa overlaid onto categorical analysis of neo- and paleo-endemism (CANAPE) results. Dark grey areas are the most restrictive protected areas (PAs) (International Union for Conservation of Nature (IUCN) categories Ia, Ib or II); light grey areas are other PAs (IUCN categories III, IV, V or VI, or with no category assigned). Data were retrieved from the World Database on Protected Areas (WDPA, https://protectedplanet.net/, accessed June 2018). Sampling units (SUs) are 100 × 100 km squares. Coloured SUs present significantly high PE: red, centres of neo-endemism; blue, centres of paleo-endemism; violet, centres of mixed-endemism (i.e. mix of neo- and paleo-endemism), with darker violet indicating centres of super-endemism; beige, not significant; white SUs contain < 100 records.
containing heterogeneous habitats and various soil types harbouring distinctly adapted plant lineages (Peterson et al., 1997; Hoorn et al., 2010, 2013; Antonelli et al., 2018). Phylogenetically overdispersed SUs have also been observed in ecotone regions in the Cape region (Forest et al., 2007) and were argued to be of high conservation importance as they represent an important amount of evolutionary history (Faith, 1992; Swenson et al., 2007; Forest et al., 2015).

The regions of high taxonomic diversity mentioned earlier (Cameroon, Gabon, the Albertine Rift, the Eastern Arc Mountains and coastal Tanzania) are also known to contain a high number of endemic species (Linder, 2001; Küper et al., 2004; Plumptre et al., 2007). We indeed found a similar pattern between GR and observed WE (Figs 4a, S1). However, our randomization procedure revealed that these regions present lower amounts of generic endemism than expected (Fig. 4b) – except for southern coastal Tanzania. This does not mean, however, that Cameroon, Gabon, the Albertine Rift and the Eastern Arc Mountains are not rich in endemic genera, but that given the high taxonomic richness, the concentration of endemic genera they harbour is not exceptional. Instead, generic level distribution ranges are wider than expected, indicating an over-representation of widespread genera in the Guineo-Congolian region (Fig. 4b). By contrast, East Africa, particularly Ethiopia, Kenya, Tanzania, Katanga (DRC), Maputaland and other SUs in southeast Africa, harbour more endemic genera than expected (high WE; Fig. 4b). These represent important endemic portions of evolutionary history (high PE; Fig. 6), possibly related to the topographic and edaphic complexity of East Africa. Indeed, East Africa has a greater elevation range than West or Central Africa (Guillocheau et al., 2018) and the tectonic activity that occurred in East Africa generated places that differ edaphically from surrounding areas. Edaphic complexity plays a great role in generating endemism (Bruchmann & Hobohm, 2014; Rahbek et al., 2019), and the habitat heterogeneity and fragmentation of montane regions are often associated with range-restricted taxa (Moritz et al., 2002; Hughes & Eastwood, 2006; Kier et al., 2009; Fjeldså et al., 2012). Moreover, East Africa is acknowledged to be a ‘complex biogeographical mixture’ containing distinct taxa adapted to various localized environmental gradients generated by the complex topography and climate (Linder et al., 2012; Droissart et al., 2018). Endemic lineages are thus more likely to occur in greater proportion in the regions that are topographically and edaphically complex than in regions of low habitat heterogeneity (such as lowlands) or low edaphic complexity (such as, *inter alia*, the Eastern Arc Mountains). Finally, endemism in eastern Africa seems to have been undervalued in previous continental-scale studies (Fjeldså & Lovett, 1997; Linder, 2001; but see Küper et al., 2004). Our incorporation of evolutionary history and randomization procedures allowed the identification of new regions of high evolutionary and conservation interests (Fig. 6).

Cradles and museums of diversity

In our study, most SUs identified as containing significantly high PE are centres of mixed- or super-endemism, concentrating both neo- and paleo-endemic genera (Fig. 6). These mixed-endemism and super-endemism SUs occur significantly more frequently in montane regions than did nonsignificant SUs (Fig. 8). In addition, neo-endemism SUs also show a nonsignificant tendency to occur mainly in montane areas (Fig. 8). Thus, even though mountains have been suggested to act as cradles of diversity (Roy, 1997; Jetz et al., 2004; Fjeldså & Rahbek, 2006; Schwery et al., 2014; Merckx et al., 2015), our results support the hypothesis that montane regions, particularly in East Africa, are both museums and cradles of diversity (Wasser & Lovett, 1993; Fjeldså & Lovett, 1997; López-Pujol et al., 2011). Our results also suggest that the montane speciation model, where mountain ‘cradles’ are considered to feed surrounding (generally lowland) regions in species (Roy, 1997; Jetz et al., 2004; Fjeldså & Rahbek, 2006; Hoorn et al., 2013), is hard to support at genus level across tropical Africa. This hypothesis appears reductive as mountains can also act as ‘museums’, and centres of diversification may also occur elsewhere, such as between river networks or in lowland ecotonal zones (e.g. transitions between forests and savannahs) (Fjeldså, 1994; Plana, 2004).

Furthermore, by identifying Ethiopian highlands, Kenya, Tanzania, Katanga, southeast Africa and Maputaland as centres of mixed-endemism, we demonstrate that within these regions, processes of recent diversification occur together with favourable conditions for lineage persistence leading to a greater accumulation of evolutionary history than expected by chance. Several studies also found areas of mixed-endemism in other montane regions, such as in South America (Fjeldså, 1994; Fjeldså & Lovett, 1997; Bitencourt & Rapini, 2013), and China (López-Pujol et al., 2011). In tropical Africa, previous identification of centres of neo- and paleo-endemism in plants focused on only two plant clades *a priori* assumed to be neo- or paleo-endemics, because of a lack of dated plant phylogenetic trees (Fjeldså & Lovett, 1997). Fjeldså & Lovett (1997) reported co-occurrences of neo- and paleo-endemics in some mountains, such as the Cameroon Volcanic Line or the Albertine Rift, but they did not detect the same East African regions to act as both museums and cradles, as reported here (Fig. 6). This underlines the importance of integrating randomization and categorization procedures together with a comprehensive phylogeny and angiosperm wide datasets to detect patterns of significant neo-, paleo- and mixed-endemism (Mishler et al., 2014; Thornhill et al., 2016).

By contrast, most parts of West and Central Africa present nonsignificant CANAPE SUs, indicating that these regions harbour no more endemic genera than expected. Randomization of the RPD revealed concentrations of old genera in lowland rainforests of Guineo-Congolia (Figs 5b, 7). As mentioned earlier, these genera also appear to be widespread (low WE; Fig. 4b). Several local museums of biodiversity have been proposed to occur within this region (Fjeldså & Lovett, 1997; Murienne et al., 2013), which, according to Linder (2014), contain the oldest flora of tropical Africa, and here appears to be a museum of diversity as a whole.

Our study highlights the importance of incorporating large-scale taxonomically verified distribution datasets with mega-phylogenies, which lead to an improved understanding of tropical
plant biodiversity evolution. Similar studies in other megadiverse regions such as the Amazon basin will allow comparison of tropical biodiversity origins and maintenance processes across the tropics.

**Taxonomic and geographic resolution effects**

Biodiversity and phylogenetic analyses at genus level have already demonstrated their relevance in understanding patterns of biodiversity evolution (Forest et al., 2007; Thornhill et al., 2016; Scherson et al., 2017). However, interpretation must not be extrapolated to other taxonomic levels. This is particularly the case for analyses that are focused on species endemism, as different species within a genus may show various geographic distribution patterns. For example, the genus *Impatiens* (Balsaminaceae) is widespread in the Old World tropics and subtropics, but the majority (c. 80%) of its African species are range-restricted in regions such as the Eastern Arc Mountains (Grey-Wilson, 1980; Lovett et al., 2000). Another drawback is that it is harder to detect centres of neo-endemism when using generic-level data (Thornhill et al., 2016). This makes the 11 SUs detected as centres of neo-endemism (Fig. 6) strong cradles of diversity. The completeness of the plant distribution dataset as well as the resolution of the phylogeny may also affect our estimate of PE. Here, even if the RAINBIO dataset represents c. 89% of the known diversity in tropical Africa, it is the best-quality data that are available to date (Dauby et al., 2016). The relatively high proportion of nonmonophyletic genera (36%) that are present in the unfiltered angiosperm tree is inherent and unavoidable when making use of large-scale phylogenetic approaches. On one hand, this is caused by the compromise between the amount of markers used and the number of species included in the dataset (see, e.g., Qian & Jin, 2016; Smith & Brown, 2018). On the other hand, there are still some taxonomical uncertainties regarding the generic status of certain angiosperm African genera (e.g. Rubiaceae, Convolvulaceae, Cyperaceae) in which several genera have not been taxonomically revised (Muasya et al., 2009; De Block et al., 2015; Simões & Staples, 2017). Despite this, we are confident that given the broad scale of our analysis, our results will be robust to improved phylogenetic studies at the genus level.

In tropical Africa, documented regions with high degrees of diversity and endemism, and postulated to act as both cradles and museums of species diversity are predominantly mountainous (e.g. Cameroon Volcanic Line, the Albertine Rift and the Eastern Arc Mountains) (Fjeldså & Lovett, 1997; Lovett et al., 2005; Burgess et al., 2007; Plumptre et al., 2007). These mountain blocks are generally small in terms of size (e.g. c. 70 × 70 km for the Eastern Arc Mountains; Burgess et al., 2007). Thus, the resolution of our analyses (100 × 100 km SUs) may have diluted the montane endemism effect with the high diversity and potentially widespread genera of adjacent lowlands regions. Using smaller SUs gave a similar overall CANAPE result, but with less coverage of the study zone because of a higher number of poorly sampled SUs that were excluded from the analyses (Fig. S7). Moreover, in mixed-endemism regions, neo-endemics and paleo-endemics may be locally concentrated in different places within a mountain block (Bitencourt & Rapini, 2013).

**Conservation implications**

Identifying and conserving areas of evolutionary potential, harbouring important processes leading to diversification and/or lineage persistence, are of crucial conservation importance (Fjeldså, 1994; Mace et al., 2003; Faith et al., 2010; Kraft et al., 2010). In eastern Africa, most of the centres of PE are centres of mixed-endemism (Fig. 6), somewhat bypassing the need to choose between conserving regions of active diversification or persistence (Cowling & Pressey, 2001).

The coverage of the identified centres of PE by the PA network is high, as 85% of the CANAPE SUs contain at least a portion of PA, and 42% contain at least a portion of the most restrictive PAs (Ia, Ib or II types). This means that important regions of diversification and persistence are already represented in the tropical African PA network. In terms of surface, only 18% of the CANAPE SUs’ surface is covered by PAs (Fig. 9; Tables S6, S7). This is as expected, as PAs are generally small areas and are unlikely to cover a complete 100 × 100 km SU. Significant CANAPE areas that are not well represented by PAs are mainly situated in South Malawi and, to a lesser extent, in Ethiopia (Fig. 9). Still, it is possible that the actual occurrences of endemics and the PAs within a SU do not coincide, so it might be worthwhile analysing diversity patterns at a fine scale when adapting or adding PAs.

The regions containing significant high PE, particularly regions of mixed-endemism that are simultaneously ‘cradles’ and ‘museums’ of diversity, should be considered in future conservation planning or PA extension. Considerable increases in the coverage of multiple facets of biodiversity (taxonomic, phylogenetic and functional) are possible with small expansions of protected areas, if the planning is achieved thoroughly and with an explicit consideration of these multiple facets (Pollock et al., 2015, 2017).

**Acknowledgements**

This work was co-funded by the French Foundation for Research on Biodiversity (FRB) and the Provence-Alpes-Côte d’Azur région (PACA) region through the Centre for Synthesis and Analysis of Biodiversity data (CESAB) programme, as part of the RAINBIO research project (http://rainbio.cesab.org/) and the Agence Nationale de la Recherche (grant number ANR-15-CE02-0002-01), both to TLPC. JCS considers this work a contribution to his VILLUM Investigator project ‘Biodiversity Dynamics in a Changing World’ funded by VILLUM FONDEN (grant 16549) and his TREECHANGE project funded by Independent Research Fund Denmark|Natural Sciences (grant 6108-00078B). GD was partly funded by the Belgian Fund for Scientific Research (F.R.S.-FNRS).

**Author contributions**

L-PMJD and TLPC conceived the study; L-PMJD and SBJ undertook the analyses; L-PMJD, SBJ, GD, AB-O, BAM, VD, J-CS, MSMS, TS, DJH, BS, JJW, OJH and TLPC contributed
data and ideas; L-PMJD and TLPC read and approved the final version. L-PMJD and SBJ contributed equally to this work.

ORCID
Anne Blach-Overgaard https://orcid.org/0000-0002-0200-1547
Thomas L. P. Couverte https://orcid.org/0000-0002-8509-6587
Léo-Paul M. J. Dagallier https://orcid.org/0000-0002-3270-1544
Gilles Dauby https://orcid.org/0000-0002-9498-413X
Vincent Droissart https://orcid.org/0000-0001-9798-5616
Olivier J. Hardy https://orcid.org/0000-0003-2052-1527
David J. Harris https://orcid.org/0000-0002-6801-2484
Steven B. Janssens https://orcid.org/0000-0003-0603-7587
Barbara A. Mackinder https://orcid.org/0000-0003-1966-2544
Anne Blach-Overgaard ORCID OJH and TLPC read and approved the final version. L-PMJD

References


Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Relationship between weighted endemism (WE) and genus richness (GR).

Fig. S2 Relationship between phylogenetic diversity (PD) and genus richness (GR).

Fig. S3 Map of observed phylogenetic diversity.

Fig. S4 Boxplot of the distributions of the mean ruggedness (in metres) of the SUs depending on their phylogenetic diversity (PD) significance class.

Fig. S5 Relationship between phylogenetic endemism (PE) and weighted endemism (WE).

Fig. S6 Maps of phylogenetic endemism.

Fig. S7 Maps of CANAPE results with different SU sizes.

Fig. S8 Boxplot of the distribution of the percentage of montane plots of the sampling units depending on their CANAPE category.

Fig. S9 Boxplot of the distribution of the percentage of overlap of the SUs with ‘montane areas’ depending on their CANAPE category.

Table S1 P-values resulting from the pairwise comparisons of mean ruggedness per sampling unit in each PD significance class.

Table S2 P-values resulting from the pairwise comparisons of mean elevation per sampling unit in each PD significance class.

Table S3 P-values resulting from the pairwise comparisons of mean ruggedness per sampling unit in each CANAPE significance class.
Table S4   *P*-values resulting from the pairwise comparison of the percentage of montane plots of the sampling units depending on their CANAPE category.

Table S5   *P*-values resulting from the pairwise comparison of the distribution of the percentage of overlap of the SUs with ‘montane areas’ depending on their CANAPE category.

Table S6   Number of significant CANAPE SUs overlaid by at least one protected area.

Table S7   Surface intersection between significant CANAPE SUs and protected areas.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

---

**About *New Phytologist***

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a *not-for-profit organization* dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.

- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication ‘as ready’ via *Early View* – our average time to decision is <26 days. There are no *page or colour charges* and a PDF version will be provided for each article.

- The journal is available online at Wiley Online Library. Visit [www.newphytologist.com](http://www.newphytologist.com) to search the articles and register for table of contents email alerts.

- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)

- For submission instructions, subscription and all the latest information visit [www.newphytologist.com](http://www.newphytologist.com)