**Supplementary Information: Environmental correlates of phylogenetic endemism in amphibians and the conservation of refugia in the Coastal Forests of Eastern Africa**

Christopher D. Barratt, Beryl A. Bwong, Renske E. Onstein, Dan F. Rosauer, Michele Menegon, Nike Doggart,Peter Nagel,W. Daniel Kissling, Simon P. Loader

*Phylogenetic tree*

Sequences were edited in Geneious 6 and aligned with MUSCLE before Bayesian analyses were performed in Beast 2.1.3 (Bouckaert et al., 2014) on the concatenated 1,271 bp alignment with a secondary calibration point of 350 million years (normal prior, sigma =10.0) as an approximate divergence date between caecilians (Gymnophiona) and frogs (Anura) following Wilkinson et al. (2011). We used the optimal models of evolution per partition (Table S2) according to the Bayesian Information Criterion determined by PartitionFinder 1.1.1 (Lanfear et al. 2013). For all partitions trees were linked, and models and clocks were unlinked. Markov Chain Monte Carlo (MCMC) simulations were run for 100 million iterations, with sampling every 5000 iterations. We checked for convergence using Tracer 1.6, and investigated parameter variations including strict, relaxed and log normal clocks, and Yule/coalescent priors, but found the relaxed clock log normal model with a Yule prior as the best parameter combination (adequate mixing and ESS values>200). We combined two independent runs in Logcombiner 2.1.3 and calculated the maximum clade credibility tree (Fig. 2) with Treannotator 2.1.2 which was used for further analyses, after the removal of 10% burn-in. Phylogenetic results are largely concordant with previous large-scale amphibian tree reconstructions in terms of topology and branch lengths (Roelants et al., 2007; Frost, 2016; Pyron & Wiens, 2011).

*Spatial data filtering*

All lineage data was compiled from our own and collaborators field work since 2001, relevant CFEA literature (Burgess and Clarke, 2000), and available distribution records from the Global Biodiversity Information Facility (GBIF). We utilized spatial records from amphibian collections housed at the University of Dar es Salaam, National Museums of Kenya, Natural History Museum London, Science Museum of Trento, Museum of Vertebrate Zoology Berkeley, Museum of Comparative Zoology Harvard, and Field Museum of Natural History Chicago. Due to uncertainty in some data sources such as GBIF and museum records we were conservative regarding which data points were retained, only keeping spatial data which we could be certain were accurate. All spatial records correspond approximately to the timeframe of the ecological variables used for species distribution modelling (i.e. between ca. 1928 and 2014), with the vast majority (ca. 98%) from between 1975 and 2014. To assess this, a rigorous filtering and correction procedure to these data prior to further analysis was made. We excluded records not present in our CFEA study region <1000m in altitude (i.e. specific EA restricted species and other species not considered as part of or interacting with the CFEA assemblage) and used only occurrence data that accurately matched the recorded locality names. Species names were corrected to reflect currently recognized taxonomy following Frost et al. (2016), excluding ambiguous records listed as ”sp.” or “indet.” that could not be confidently matched to a species tip on the tree.

*Forest modelling*

As no accurate historical vegetation maps of sub-Saharan Africa exist for the time-scales we investigate, we use modelling methods to estimate forest distribution, firstly in the present time and then projected back onto global circulation model climate data for the mid-Holocene (6 ka BP), Last Glacial Maximum (LGM; 21 ka BP) and Last Interglacial (LIG; 120 ka BP) time periods (Hijmans et al. 2005, Braconnot et al., 2007, Otto-Bliesner et al., 2006). Current forest extent was estimated by creating a forest distribution model using the global 1km consensus land cover product (Tuanmu and Jetz, 2015) to classify the study region based on evergreen forest presence (excluding mangroves which are not part of the CFEA). Firstly we reclassified their raster layer, taking only pixels representing the top 10% of all values for evergreen forest to ensure an accurate representation of current forest was maintained. We transformed the resulting raster into a single polygon and generated 500 random points within this area, with a constraint that the points had a buffer of at least 10 km apart to minimize bias (sensu Graham et al. 2010). We verified that the points covered an accurate representation of current forest habitat by overlaying them on satellite imagery from Google Earth and ground truthing based on first-hand knowledge of the region from recent fieldwork. We subsequently created a predictive model of current forest cover in MaxEnt 3.3.3k using bioclimatic variables (bio2, bio4, bio5, bio12, bio14, bio18 from the CCSM global circulation model) and a digital elevation model (GTOPO30; USGS, available from: http://csgtm.iscgm.org/dataset/gtopo30) at 30 arc-second resolution (approximately 1 km2) with the subsample modeling algorithm, and regularization multiplier of 1. After inspection of the predictive forest model to verify that it accurately represented currently forested habitats, we created historical projections based on past climate data, by projecting the model onto the three time slices with available historical climate data. We followed Graham et al. (2010) to derive a measure of forest stability over time by transforming the forest model at each time slice into its log value and averaging the values across the four time periods (i.e. from the present back to the LIG over 120k years). The exponent of this average value (EXP function in ARCGIS raster calculator) was then used to transform the resulting number into a continuous variable, representing a measure of forest stability in each grid cell (ranging from 0 to 1) following Rosauer et al. (2015) and Graham et al. (2010).

The forest model performed well, with a high AUC values (0.811). Variable contribution was highest for bio14 (53.73%) with each other variable contributing between 6.33 and 13.69% with the exception of bio5 (Table S6). Forest projections at 6 ka BP, 21 ka BP, 120 ka BP and the calculated forest stability surface measure over this time period show that several areas have retained high stability throughout. These areas correspond to lowland parts of the EA and surrounding areas, parts of central coastal Tanzania, with additional areas scattered along the coastline to the east of our map. Remaining low elevation areas and those extending inland have mainly been unstable in terms of forest cover, probably affected by sea level inundation and periods of severe climate change (Fig. S3).

*Spatial autocorrelation*

To test for spatial autocorrelation we examined the correlation structure of the log transformed response variable (PE) and residuals of the best GLM, visualizing correlograms of Moran’s *I* across distance classes of 10 km increments. Model residuals of the GLM were spatially autocorrelated at small distance classes (i.e. Moran’s *I* values > 0.7 up to 20 km distance between grid cells), suggesting the need for a spatial model. To define a spatial weights matrix we used the knearneigh function (with k = 1) to derive the minimum distance connecting each cell to at least one neighbor. Using this minimum distance we then defined the neighborhood structure using the dnearneigh function, and created a spatial weights matrix using the nb2listw function, for details see Kissling & Carl (2008). We then used the best GLM and ran a spatial autoregressive error model with the spatial weights matrix to account for spatial autocorrelation using the errorsarlm function in R package ‘spdep’ (Bivand & Piras, 2015). This spatial autoregressive error model accounted for spatial autocorrelation in GLM residuals across all distance classes (Tables S6, S7).

**References**

Bivand, R. & Piras, G. (2015) Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software,* 63, 1-36. Available from: http://www.jstatsoft.org/v63/i18/.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, M.A., Rambaut, A., & Drummond, A.J. (2014). BEAST 2: A Software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10, e1003537. doi:10.1371/journal.pcbi.1003537.

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, T., Hewitt, C.D., Kageyama, M., Kitoh, A., Laîné, A., Loutre, M.-F. , Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, S.L., Yu, Y., Zhao, Y. (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum - Part 1: experiments and large-scale features. *Climate of the Past*, **3**, 279–276.

Calcagno, V. & de Mazancourt, C. (2010) glmulti : An R package for easy automated model selection with (generalized ) linear models. *Journal of Statistical Software*, **34**, 1-29.

Frost, D.R. (2016) Amphibian Species of the World: an Online Reference. Version 6.0 (*1 October 2016*). Electronic Database available from: http://research.amnh.org/herpetology/amphibia/index.html. American Museum of Natural History, New York, USA.

Graham, C.H., VanDerWal, J., Phillips, S.J., Moritz, C. & Williams, S.E. (2010) Dynamic refugia and species persistence: Tracking spatial shifts in habitat through time. *Ecography*, **33**, 1062-1069.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Kissling, W.D. & Carl, G. (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, **17**, 59-71.

Lanfear, R., Calcott, B., Ho, S.Y.W., & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution,* **29,** 1695–1701.

Otto-Bliesner., B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H., Hu, A., & CAPE Last Interglacial Project members (2006) Simulating Arctic climate warmth and icefield retreat in the last interlglaciation. *Science*, **311**, 1751-1753.

Pyron, R.A. & Wiens, J.J. (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, **61**, 543-583.

Roelants, K., Gower, D.J., Wilkinson, M., Loader, S.P., Biju, S.D., Guillaume, K., Moriau, L. & Bossuyt, F. (2007) Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences USA*, **104**, 887-892.

Tuanmu, M.N. & Jetz, W. (2014) A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, **23**, 1031-1045.

Wilkinson, M., San Mauro, D., Sherratt, E. & Gower, D.J. (2011). A nine-family classification of caecilians (Amphibia: Gymnophiona). Zootaxa, **2874**, 41-64.