

Migration corridors for alpine plants among the ‘sky islands’ of eastern Africa: do they, or did they exist?

Desalegn Chala¹, Niklaus E. Zimmermann², Christian Brochmann^{1*} and Vegar Bakkestuen^{1,3*}

¹Natural History Museum, University of Oslo, PO Box 1172 Blindern, NO-0318 Oslo, Norway

²Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

³Norwegian Institute for Nature Research, Gaustadalléen 21, N-0349 Oslo, Norway

*shared senior authorship

* **Corresponding author:** Desalegn Chala, Natural History Museum, University of Oslo, PO Box 1172 Blindern, NO-0318 Oslo, Norway, E-mail: d.c.gelete@nhm.uio.no or desdchala@gmail.com

Abstract

The tropical alpine ecosystem in eastern Africa is highly fragmented among biological ‘sky islands’, where populations of frost-tolerant organisms are isolated from each other by a ‘sea’ of tropical lowlands. One-third of the species in the afroalpine flora are exclusively alpine, but the other species can to varying degrees extend into grasslands and open forests of lower vegetation belts. A long-debated question is whether colonization of the alpine zone of these mountains and subsequent intermountain gene flow entirely depend on long-distance dispersal across unsuitable habitats, or whether suitable habitats shifted far enough downslope under past colder climates to form bridges enabling gradual migration. Here we address this question using a classification tree model. We mapped the extent of the current alpine habitat and projected it to the last glacial maximum (LGM) climate to assess whether gradual migration was possible for exclusively alpine taxa during this glacial period, and thus potentially also during earlier Pleistocene glaciations. Next, we modelled landcover under current and LGM climates to assess whether grassland and open forests could have served as migration corridors for alpine taxa that today extend into lower vegetation belts. We estimated that the LGM treeline was about 1,000 m lower and the alpine habitat was about eight times larger than today. At the LGM, we found that most of the currently fragmented alpine habitat of the Ethiopian highlands was interconnected except across the Great Rift Valley, whereas the solitary mountains of East/Central Africa remained isolated for exclusively alpine species. However, for drought-tolerant alpine species that today extend below the treeline, gradual migration through habitat corridors may have been possible among mountains during the dry glacial periods, and possibly also under the current climate before agriculture transformed the low-lying landscapes.

Keywords: Afroalpine, CART, gene flow, habitat connectivity, Last Glacial Maximum, treeline

Introduction

Fragmented landscapes and isolated islands have been a central topic in evolutionary and biogeographic research since Darwin (Darwin, 1859) and Wallace (Wallace, 1876). In eastern Africa, the tropical alpine habitat with its high degree of endemism and peculiar life forms is today restricted to the highest mountain peaks. It represents a highly fragmented biological 'sky island' system, where populations of frost-tolerant alpine organisms are isolated from each other by tropical lowlands. A long-standing question in African biogeography is whether colonization of the mountains and subsequent intermountain gene flow mainly depend on long-distance dispersal across unsuitable lowland habitat, or whether gradual migration has been possible through habitat bridges under colder climates in the past, when the afroalpine habitat extended to lower altitudes than today (Coetzee, 1964). Here we address this question based on a new analysis using alpine habitat and landcover type modeling, and by hindcasting these models to the climate of the Last Glacial Maximum (LGM) to represent the glacial periods during the Pleistocene.

Tropical alpine ecosystems are shaped by unique climatic conditions, characterized by warm days alternating with freezing nights and little seasonal variation in temperature (Hedberg, 1964; Rundel, 1994). At tropical latitudes, only mountains reaching above the treeline at about 3500 m present such climatic conditions and support true alpine floras (Hedberg, 1951). In eastern Africa, alpine ecosystems are found scattered in three major mountain regions, two of them located in Ethiopia and one in East/Central Africa. The contiguous and elevated Ethiopian Plateau is divided by the Great Rift Valley into one northwestern (NWE) and one southeastern (SEE) region, both characterized by high peaks and mountain ranges. The East/Central African (ECA) mountain region comprises several solitary high mountain peaks in Kenya, Tanzania, Uganda, Democratic Congo and Rwanda, separated from the Ethiopian mountains by the Turkana Lowlands (Fig. 1).

The high eastern African mountains harbor unique alpine floras and faunas. Endemism is remarkably high, 80% at the species level for vascular plants (Hedberg, 1957; Gehrke & Linder, 2014) and 64% for mosses (Spence & Pocs, 1988). In spite of the high degree of fragmentation, the mountain floras have many species in common and show close relationships to remote temperate regions in the northern and southern hemispheres (Hedberg, 1951; Hedberg, 1970; Smith & Cleef, 1988; Gehrke & Linder, 2009). It has been suggested that the isolation among these mountain floras may be stronger than that among oceanic islands, where ocean currents may play a role as an interconnecting dispersal agent (Hedberg, 1970). In particular, the Rift Valley is a prominent landscape feature which has long been considered to represent a strong barrier against gene flow (Hedberg, 1970; Kebede *et al.*, 2007; Gizaw *et al.*, 2016). The Rift Valley is a continuous valley system, 6000 km in length, and runs from northern Syria in Asia to Mozambique in southeastern Africa, cutting through mountains and lowlands in eastern Africa. It was formed by volcanic processes that started many million years ago and continued to be active well into the Pleistocene (Baker *et al.*, 1972; Gehrke & Linder 2014).

How the sky island biotas evolved and how similarities among mountain systems were established and maintained over time have been intensely discussed in the literature (Hedberg, 1951, 1957, 1970; Smith & Cleef, 1988; Harmsen *et al.*, 1991; Gottelli *et al.*, 2004; Koch *et al.*, 2006; Ehrich *et al.*, 2007; Umer *et al.*, 2007; Popp *et al.*, 2008; Gehrke & Linder, 2009; Gehrke & Linder, 2011; Gehrke & Linder, 2014; Sklenář *et al.*, 2014). The

proportion of shared species generally correlates with the geographic distance among the mountains (Gehrke & Linder, 2014; Sklenář *et al.*, 2014). The more geographically isolated the mountains are, the more endemic taxa they harbor and the poorer they are in species number (Smith & Cleef, 1988; Harmsen *et al.*, 1991; Gehrke & Linder, 2014). Colonization patterns thus seem to be closely associated with geographic distance. However, the patterns of genetic structuring and diversity within species seem to be more complex. Recent phylogeographic studies suggest that afroalpine species may have a dynamic history with periods of allopatric divergence often interrupted by repeated intermountain dispersal and hybridization, but the inferred histories differ considerably from species to species and among different mountains (Assefa *et al.* 2007, Ehrich *et al.* 2007; Gizaw *et al.* 2013, 2016; Masao *et al.* 2013; Wondimu *et al.* 2014). In some species, populations inhabiting distant mountains are genetically very similar, suggesting recent colonization or extensive gene flow. Two species that also extend to lower vegetation belts in the mountains show no geographic structuring of their genetic diversity at all (Gizaw *et al.* 2013; Masao *et al.* 2013), possibly indicating that habitat connectivity through lowland corridors has been more important than long-distance dispersal.

The Pleistocene ice ages were distinct climatic events during the past 700,000 years also in eastern Africa. The vegetation belts in the mountains were oscillating in response to the changing climate, but few details are known (Bonnefille *et al.*, 1990; deMenocal, 1995; Shanahan & Zreda, 2000; Thompson *et al.*, 2002). Pollen analyses show that the African tropics were colder and drier than today during the last glaciation (Bonnefille *et al.*, 1990; Mark & Osmaston, 2008) and that the alpine zone in East Africa may have descended as much as 1,000 m (Coetzee, 1964). The treeline started to rise soon after the start of the Holocene (11 200 cal BP) following the increase in temperature and moisture and soon reached its current elevation (Umer *et al.*, 2007). Globally, the natural treeline is associated with an isoline average temperature of 6.4 ± 0.7 °C during a growing season that lasts at least 90 days (Körner, 2012). In the tropics, there is only little seasonal variation in temperature and the growing season therefore lasts the whole year with respect to temperature (Rundel, 1994, Körner & Paulsen, 2004). In our study area, the annual temperature is highly correlated with seasonal temperatures (Fig. S1, table S1) and can therefore be used as a proxy for the temperature of the growing season. In tropical eastern Africa, the temperature was reconstructed to have been 4.0 ± 2.0 °C lower during the LGM than today (Bonnefille *et al.*, 1990). The LGM temperature reported by Schneider von Deimling *et al.* (2006) falls within the same range.

Despite shifting to lower altitudes during cold periods, the afroalpine belt in East/Central Africa has been thought to have been isolated since its origin (Hedberg, 1970; Harmsen *et al.*, 1991). This may be true for these solitary mountains, but it remains uncertain to what degree this also holds for the alpine habitat in Ethiopia. In Ethiopia, vast landmasses are found at high elevations and the current alpine habitat exists on punctuated volcanic uplifts. It is therefore possible that the alpine zone of most of the high Ethiopian mountains became interconnected and formed one continuous habitat, at least on each side of the Rift Valley under colder climates.

The possibility for gradual migration through habitat corridors between the eastern African mountains under colder climates may however have varied considerably among different afroalpine species, which show large variation in their lower elevation boundaries under the current climate. While some species are confined to areas above 4500 m, others extend into afroalpine grasslands and open montane forests, and some of them have

been observed as low as 1000 m (Hedberg, 1970; Gehrke & Linder, 2014). Only 169 (32%) of the species in the afroalpine flora are 'exclusively alpine' when defined by a lower elevation limit of 3200 m (Gehrke & Linder 2014). Thirty-six percent of the species not only occur below the afroalpine zone in eastern Africa but also in tropical countries that lack mountains characterized by alpine climate (Gehrke & Linder, 2014). Depending on their individual traits and habitat requirements, it may thus have been possible for some 'non-exclusively alpine' species to disperse among mountains through non-alpine habitat corridors under colder climates, and possibly even under the current climate before the landscape was strongly transformed by agriculture. Strong isolation and migration solely via long-distance dispersal may thus be restricted to the one-third of the afroalpine species that are exclusively alpine and only occur in areas experiencing nightly frosts. Since most alpine species are adapted to grow in full sunlight, grasslands at lower altitudes are candidate for serving as migratory corridor for the non-exclusively alpine species that are sufficiently drought-tolerant. For afroalpine species that today stretch to the lowermost altitudes even open forests may serve as migration corridor.

Here we explore the potential for past habitat connectivity under the colder climate of the LGM, to represent also earlier glacial maxima during the Pleistocene, and assess whether gradual migration corridors can help explain patterns in floristic and genetic similarity among the three mountain regions as well as among individual mountains. We used a classification tree model to project the treeline position as well as the land cover distribution based on paleoclimate data. We modelled the extent of the current alpine habitat and projected it to the LGM climate to assess whether the alpine habitat was widespread enough to allow gradual intermountain migration of exclusively alpine species. Next, we modelled a suite of eastern African landcover types both under current and LGM climates to assess whether grassland and open forest types may have served as migration corridors for non-exclusively alpine species, i.e. those that today extend downslope into the montane forest and other lower elevation vegetation types.

Methods

Defining the alpine habitat in the study region

The alpine habitat in tropical Africa covers the area situated above approximately 3500 m (Hedberg, 1951). We used this elevation as a cut-off value and classified the study area into alpine and non-alpine landcover units based on the 1 km resolution DEM (Hijmans *et al.*, 2005). We sampled both ‘alpine’ and ‘non-alpine’ habitat units randomly proportional to the log of their area (x 1,000) for generating a training and test dataset. We randomly split this dataset into 75% for training and 25% for testing.

Defining the landcover types in the study region

We used the 300 m resolution GlobCover dataset (Arino *et al.*, 2012) to represent landcover types in our study area after simplifying and reclassifying it into seven general landcover units: (i) closed forest, (ii) open forest I, (iii) open forest II, (iv) shrub/grassland I, (v) shrub/grassland II, (vi) flooded areas and (vii) croplands and other. Almost all landcover types of the GlobCover dataset cover wide altitudinal ranges. Especially the landcover units that are categorized as open forest and shrub/grassland are very heterogeneous in terms of their distribution across elevation gradients (Figs. S1& S2). Some of them are mainly confined to lower elevations, while others cover a large altitudinal span. To minimize heterogeneity, those covering a large altitudinal span and thus potentially most suited as migration corridors for alpine species were reclassified as type I (shrub/grassland I and open forest I), and those mainly confined to lower elevations and thus drier conditions were reclassified as type II (shrub/grassland II and open forest II; Table 1). We computed the area per landcover unit and sampled each unit randomly proportional to the log of its area (x 1,000) for generating a training and test dataset. The ‘croplands and other’ landcover unit was excluded from land cover reconstruction. Over the last 10,000 years, 50% of the ice-free terrestrial surface has been modified by humans, whereas there was little human influence prior to this period (Lambin *et al.*, 2003). Cropland is a landcover type that has been converted from other landcover units. We decided to model ‘potential land cover units’ by letting other landcover units colonize the areas now covered by cropland. All sample points were randomly split into 75% training and 25% test data.

Predictor variables

In order to model both alpine/non-alpine vegetation and the sampled landcover types, we used 19 bioclimatic variables of the current climate (~1 km resolution) obtained from the WorldClim database (Hijmans *et al.*, 2005). We also used a DEM of ~1 km resolution from the same data source and generated two topographic variables, slope angle and slope aspect. The slope aspect was converted into eight categorical classes (N, NE, E, SE, S, SW, W, NW). We stacked the bioclimatic layers and the two topographic variables and overlaid the stacked raster datasets with the randomly sampled points. These sample points now contained information about alpine vs. non-alpine classes for alpine habitat modelling and the six land cover units for landcover types as dependent variables and the 19+2 variables as independent predictors. Pairwise Pearson correlations were computed between all independent variables. For further analyses, we only retained predictor variables with a Pearson correlation threshold below |0.7| by selecting only one variable based on assumed biological importance among those correlated higher than this threshold. This resulted in ten variables selected for building both models (alpine habitat and landcover types; Fig. S3): slope angle, slope aspect, Annual Mean Temperature (BIO1),

Mean Diurnal Range (BIO2), Isothermality (BIO3), Precipitation Seasonality (Coefficient of Variation; BIO15), Precipitation of Wettest Quarter (Bio16), Precipitation of Driest Quarter (Bio17), precipitation of warmest quarter (BIO18) and precipitation of coldest quarter (BIO19).

In order to allow both alpine habitat and landcover type reconstruction of the LGM climate, we used the same eight bioclimatic variables that we selected from three different climate models of the LGM (~5 km resolution) from the WorldClim database (www.worldclim.org): CCSM4 (CC, hereafter); MICRO-ESM (MR) and MPI-ESM-P (ME). All the bioclimatic layers of the three LGM climate models were resampled to the same (1 km) resolution as the current climate variables using nearest neighbor assignment technique in a GIS.

Statistical model calibration of alpine habitat extent and landcover types

The randomly selected training data points and the selected predictor variables were used as input to fit a classification tree model to reconstruct both the alpine habitat extent and the landcover types in the rpart package (Therneau & Atkinson, 2015). We first fitted a full rpart model with low values for minimal complexity gain per additional split ($cp = 0.001$), which can be considered a relaxed parameter set that allows for overfitting the resulting tree. Next, we optimized the complexity parameter (cp) of a statistically optimized tree model through crossvalidation by identifying the cp value after which the improvement was no longer significant. The full tree was pruned at this identified cp value of the optimized model. Finally we computed misclassification error rates and Cohen's kappa values on both the training and test data sets to evaluate model performance. Finally, we projected these models (alpine habitat extent and land cover types) to the three LGM climate models.

In addition to using the reconstructed LGM climates, we also generated an additional environmental layer in order to build an even simpler model of treeline shift under LGM climate. We derived an LGM mean annual temperature map by deducting 4.0 °C from each pixel of the current mean annual temperature values relying on information from Bonnefille et al. (1990), i.e., in tropics the temperature was expected to be 4.0 ± 2.0 °C lower during LGM than today. We then built an additional, simpler model of alpine habitat based only on two variables: 1) annual mean temperature and 2) slope angle. This conforms to the general theory of treeline position by Körner (2012), but modified by adding one more variable, slope, which had high variable importance in the more complex model (Table 2). We used the same training and test dataset and the same procedure as in the more complex models and evaluated it accordingly. Then we hindcasted the model to the computed average annual temperature of the LGM, which had temperatures reduced by 4.0 °C. Finally we investigated how much the treeline descended and the extent of the alpine habitat was enlarged during the LGM under the more simple treeline model applied to the current temperature map that was reduced by 4 °C. This means that we had a total of 4 treeline reconstructions for LGM, three originating from the more complex treeline model that was applied to the three LGM climate models, and one originating from the simpler treeline model.

Results

Habitat connectivity among exclusively alpine taxa

Hindcasting the alpine habitat based on all selected predictor variables to the three LGM climate models showed that there was a vast extent of alpine habitat both in Ethiopia and in East/Central Africa (Fig. 2). The simpler model based only on slope and mean annual temperature of the LGM gave similar results (Fig. 3). The models were highly and equally robust, with very low misclassification error rates (< 0.001) and gave very high kappa values (> 0.99) when evaluated on both the training and test datasets. In the more complex treeline model, annual mean temperature had the highest variable importance followed by slope and precipitation of the warmest quarter (bio18; Table 2). In the simpler treeline model, mean annual temperature contributed 60% and slope contributed 40% to the predictive power.

The simpler model predicted a larger extent of alpine habitat area both under current and LGM climate than did the more complex model. Although the hindcasted LGM extent of the alpine habitat area varied among climate models, the predicted lower altitudinal limit of the alpine zone was quite similar, approximately 1000 m lower than today (Figs. 4, 5). The simpler model predicted larger alpine extent under LGM climate (Fig. 5, 7). Averaged over all four treeline reconstructions, the afroalpine habitat was 8.3 times larger during the LGM as compared to today.

In Ethiopia, most of the currently fragmented afroalpine habitat areas were united during the LGM, except across the Rift Valley. Especially in the Shewa-Wallo massif in the NWE mountain region, almost all currently alpine fragments became interconnected following the downward shift of the treeline. In contrast, the alpine habitat areas in East/Central Africa remained disconnected during LGM, despite expansion on each of the solitary mountains.

Under both current and LGM climates, the alpine habitat area in Ethiopia was more than two times larger than that in the East/Central Africa (Figs. 6 & 7). The Ethiopian alpine habitat represented 64% and 69% of the total alpine habitat under current and LGM climates, respectively. The Ethiopian mountains in the SEE region alone harbored a current alpine habitat almost as large as that in East/Central Africa. The modelling based on current climate did not predict presence of a proper alpine habitat in Yemen, but, notably, all models suggested the presence of an araboalpine habitat during the LGM.

Habitat connectivity among alpine taxa that extend to lower altitudes

The landcover model had comparably high misclassification rates (0.45 on training data and 0.43 on test data) and a fair (Monserud & Leemans, 1992) kappa value (0.48 on training and 0.46 on test data), so the results were interpreted with caution. Even under the current climate, the land cover data and the model predicted extensive grasslands interconnecting the three main mountain regions. The same was found for two of the three LGM climate models (MR and ME; Fig. 8). In all these models, the grassland habitat of type I interconnected individual Ethiopian mountains both in the NWE and SEE regions, which also were bridged by grasslands extending across the Rift Valley. The East/Central African solitary mountains were also interconnected by grasslands of type I in the northern and western parts, and by savannah open forest (open forest II) in the southeastern part. The third LGM model (CC) showed the lowland (savannah) grassland (type II) to be more widespread. Although this model also showed grassland habitat connectivity among individual mountains both in Ethiopia and East/Central Africa, the Turkana lowlands were mapped to be fully covered by lowland (savannah) grasslands (type II), which might only have served as a dispersal corridor for very drought-adapted alpine taxa. The GlobCover (Arino *et al.*, 2012) dataset currently shows very limited forest cover in eastern Africa with much open space that can serve as a migration corridor for alpine taxa (Fig. 8e).

Our results from the landcover type modelling under both current and LGM climates showed that the Rift Valley in Ethiopia may have allowed gradual migration of alpine species that are able to grow below the treeline (Figs. 8a, b, c & d). In contrast, the Turkana lowlands appeared to have remained a considerable barrier between the Ethiopian and the East/Central African mountains also during the LGM. The savannah open forest mapped between the Ethiopian NWE region and East/Central Africa on the western side may however have served as migration corridor for some species.

Discussion

Using alpine habitat and landcover type modeling and hindcasting these models to the LGM climate, we have here provided insights into a long-debated issue in African biogeography: to what degree did colonization of the currently strongly isolated sky islands and subsequent intermountain gene flow depend on long-distance dispersal across unsuitable habitat, and to what extent did suitable habitat corridors occur under LGM and earlier colder climates to enable gradual migration? We have shown that the answer to these questions may strongly depend on the species and the mountain region considered. Exclusively alpine species, which currently only occur above the treeline in these mountains, remained isolated in each of the solitary East/Central African mountains even during the LGM, when we estimated the treeline to be 1,000 m lower and the alpine habitat about eight times larger than today. In this region, exclusively alpine species seem to have been entirely dependent on long-distance dispersal for colonization and intermountain gene flow (except for a few mountains in Kenya and along the Uganda/Rwanda/Congo borders; Figs. 2 & 3). In contrast, the alpine fragments in Ethiopia probably fused during the LGM except across the Rift Valley, allowing for migration of such species. We also show that drought-tolerant alpine species that today extend below the treeline may have had opportunities for gradual migration through grassland and open forest corridors among mountains in all regions during the dry glacial periods, possibly also under the current climate before agriculture transformed the low-lying landscapes.

Exclusively alpine species: isolated in East/Central Africa, partly connected in Ethiopia

The results from both treeline models and all LGM climate data sources showed that the afroalpine habitat was considerably expanded during this (and probably earlier) glaciation(s), on average about 8.3 times larger than today. Although the spatial extent of the alpine habitat predicted by different LGM climate models differed somewhat, they consistently predicted that the LGM treeline was about 1,000 m lower than today, in agreement with the results of a pollen analysis from Mount Kenya (Coetzee, 1964).

Our finding of a lack alpine habitat in the low-lying landscapes separating the East/Central African mountains during the LGM is consistent with Hedberg (1970) and Harmsen *et al.* (1991), who considered their alpine zones to have been isolated since their origin. This would explain why these mountains harbor more single-mountain endemics than the Ethiopian mountains (Hedberg, 1964; Hedberg, 1970; Knox & Palmer, 1998), which our analyses showed to have been more interconnected during colder climates. Considerable intermountain gene flow among the East/Central mountains has nevertheless been inferred in several phylogeographic studies, even across the vast Ugandan gap separating the mountains along the western and eastern branches of the Rift Valley, demonstrating the importance of long-distance dispersal and hybridization (Gizaw *et al.* 2013, 2016; Masao *et al.* 2013; Wondimu *et al.* 2014, and unpublished data).

In Ethiopia, the Rift Valley appears to have presented a strong barrier against migration of exclusively alpine species even during the LGM (Figs. 2&3). Although the northwestern and southeastern mountain regions share a considerable number of alpine species (Gehrke & Linder 2014), the Rift Valley represent a distinct genetic break

in several of those examined to date (Gizaw *et al.* 2013, 2016; Masao *et al.* 2013; Wondimu *et al.* 2014, and unpublished data). However, the phylogeographic studies have mainly focused on the geographically most separated mountain ranges in Ethiopia, the Simen Mts in the northwest and the Bale Mts in the southeast (Fig. 1), which often harbor distinct genetic groups (e.g. Gizaw *et al.* 2016). The study of Wondimu *et al.* (2014) also included Mt Choke, which is situated in the northwestern region but much closer to the Rift Valley than the Simen Mts and showed that the populations in this mountain represented admixtures of both the Simen and Bale Mts. These results call for more extensively sampled phylogeographic studies to assess the relevance of our finding of the fusion of LGM alpine habitat on each side of the Rift but not across it for gene flow in exclusively alpine species in Ethiopia.

Interestingly, our models predicted the existence of an arboalpine LGM habitat in Yemen, an area that lacks proper alpine zones today. This is a significant finding since Yemen and the Arabian Peninsula have been suggested as a main migratory pathway for Eurasian lineages colonizing the mountains of eastern Africa (Koch *et al.* 2006; Assefa *et al.* 2007, Ehrich *et al.* 2007, Popp *et al.* 2008, Gehrke & Linder 2009).

In eastern Africa, we found that the alpine habitat seems mainly to be shaped by temperature-related bioclimatic variables, which is in agreement with Körner & Paulsen (2004). Mean annual temperature strongly correlates with seasonal temperature variables and was found to be the most important variable in our model. The seasonal variation in temperature is minimal (Rundel, 1994). Because of the high correlation with seasonal temperature variables, mean annual temperature can be used as a proxy to the temperature of the growing season. We found that relying only on annual temperature and slope to model the treeline and the extent of the alpine habitat in eastern Africa provided a model almost as robust as the model in which we used more variables to discriminate the alpine habitat from non-alpine lowlands.

Alpine species that extend below the treeline: potential grassland corridors

Because our landcover model had comparably high misclassification error rates and a fair kappa value, we interpret the results with caution. Many of the landcover types in eastern Africa are today influenced by anthropogenic activities, and they cover large altitudinal spans (Figs. S2, S3). This indicates that they are quite heterogeneous and that the thematic resolution of GlobCover is not very precise, and may partly explain the comparably high misclassification in our habitat model.

From the GlobCover data, it can clearly be seen that the forest cover in eastern Africa is small and highly fragmented (Fig. 8e). In accordance, even under the current climate (when agricultural landscapes were replaced by potential natural landcover types in our models), we predicted extensive grasslands interconnecting individual mountains as well as the three main mountain regions. The same was found for two of the three LGM climate models. Grasslands and open forests probably interconnected all the Ethiopian mountains, also across the Rift Valley. In East/Central Africa, the mountains were also interconnected by grassland extending from the Virunga

and Rwenzori mountains in the west to Mt Elgon and Mt Kenya in the northeast, with further extensions towards northwestern Ethiopia on the western side of the Turkana Lowlands (Figs. 1 & 8). The Turkana Lowlands separating the East/Central African mountains from Ethiopia seem to present the strongest barrier for alpine species in our study region; this area was mapped to be fully covered by the driest type of grasslands (type II), which might have served as a dispersal corridor only for very drought-adapted taxa.

Regardless of the uncertainty associated with our model, it is difficult to assess to what degree grasslands may have served as migration corridors for ‘non-exclusively’ alpine species. Firstly, such grasslands may be partly too dry to be suitable for most alpine species, especially during glacial periods. Secondly, it is possible that most of the alpine species that today also occur below the treeline have their primary populations in the alpine zone, functioning as sources for occasional sink populations extending downslope along open patches and along watercourses. However, we tentatively explored occurrence data for some alpine taxa from the Global Biodiversity Information Facility (GBIF), providing some support for the Rift Valley in Ethiopia, but not the Turkana Lowlands, as a potential migration corridor for non-exclusively alpine plants (Chala et al., unpubl. data). We found some occurrence points located in the current grasslands in the Ethiopian Rift Valley, bridging those from the northwestern and southeastern mountain regions. Our findings are also in line with phylogeographic studies of some alpine species, such as *Erica arborea* (Gizaw et al., 2013) and *Koeleria capensis* (Masao et al., 2013). These species are drought-tolerant, extend to lower vegetation belts in the eastern African mountains, and show no geographic structuring of their genetic diversity, possibly indicating that habitat connectivity through grassland and open forest corridors has been more important than long-distance dispersal.

Although the Turkana Lowlands appear as a virtually impenetrable barrier against gradual migration of alpine species, the genetic structuring in several species points to higher levels of gene flow across this extensive lowland gap than across the Rift Valley in Ethiopia. In these species, the populations in southeastern Ethiopia belong to the same genetic group as those in East Africa, whereas another, distinct genetic group is found in northwestern Ethiopia (Gizaw et al., 2013; Wondimu et al., 2014; Gizaw et al., 2016). The most likely explanation is that this unexpected pattern is driven by long-distance dispersal directed by prevailing winds and bird migration routes.

Our results call for more detailed comparative studies of species with different traits, habitat requirements, and altitudinal ranges using a combined phylogeographic and modeling approach, to further explore the relative importance of migrational corridors and long-distance dispersal in the fascinating sky island system of eastern Africa.

Acknowledgements

We acknowledge financial support by the Norwegian State Loan Fund (Lånekassen); Natural History Museum, University of Oslo and the PET and GEco research groups at the Museum. We thank Achilleas Psomas, Swiss Federal Research Institute (WSL), and all members of PET and GEco for discussions on the manuscript.

References

- Arino, O., Ramos Perez, J.J., Kalogirou, V., Bontemps, S., Defourny, P. & Van Bogaert, E. (2012) Global Land Cover Map for 2009 (GlobCover 2009). In. PANGAEA - Data Publisher for Earth & Environmental Science.
- Assefa, A., Ehrich, D., Taberlet, P., Nemomissa, S. & Brochmann, C. (2007) Pleistocene colonization of afro-alpine "sky islands" by the arctic-alpine *Arabis alpina*. *Heredity*, **99**, 133-142.
- Baker, B.H., Mohr, P.A., Williams, L.A.J. (1972). Geology of the Eastern Rift System of Africa. *The Geological Society of America. Special Paper* **136**: 67.
- Bonnefille, R., Roeland, J.C. & Guiot, J. (1990) Temperature and rainfall estimates for the past 40,000 years in equatorial Africa. *Nature*, **346**, 347-349.
- Coetzee, J.A. (1964) Evidence for a considerable depression of the vegetation belts during the upper Pleistocene on the East African Mountains. *Nature*, **204**, 564-566.
- Darwin, C. R. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London..
- deMenocal, P.B. (1995) Plio-Pleistocene African climate. *Science*, **270**, 53-59.
- Ehrich, D., Gaudeul, M., Assefa, A., Koch, M.A., Mummenhoff, K., Nemomissa, S. & Brochmann, C. (2007) Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, **16**, 2542-59.
- Gehrke, B. & Linder, H.P. (2009) The scramble for Africa: pan-temperate elements on the African high mountains. *Proceedings of the Royal Society Biological Sciences*, **276**, 2657-2665.
- Gehrke, B. & Linder, H.P. (2011) Time, space and ecology: why some clades have more species than others. *Journal of Biogeography*, **38**, 1948-1962.
- Gehrke, B. & Linder, H.P. (2014) Species richness, endemism and species composition in the tropical Afroalpine flora. *Alpine Botany*, **124**, 165-177.
- Gizaw, A., Kebede, M., Nemomissa, S., Ehrich, D., Bekele, B., Mirré, V., Popp, M. & Brochmann, C. (2013) Phylogeography of the heathers *Erica arborea* and *E. trimera* in the afro-alpine 'sky islands' inferred from AFLPs and plastid DNA sequences. *Flora Morphology, Distribution, Functional Ecology of Plants*, **208**, 453-463.
- Gizaw, A., Wondimu, T., Mugizi, T., Masao, C., Abdi, A., Popp, M., Ehrich, D., Nemomissa, S. & Brochmann, C. (2016) Vicariance, dispersal, and hybridization in a naturally fragmented system: the afro-alpine endemics *Carex monostachya* and *C. runssoroensis* (Cyperaceae). *Alpine Botany*, 1-13.
- Gottelli, D., Marino, J., Sillero-Zubiri, C. & Funk, S.M. (2004) The effect of the last glacial age on speciation and population genetic structure of the endangered Ethiopian wolf (*Canis simensis*). *Molecular Ecology*, **13**, 2275-2286.
- Harmsen, R., Spence, J.R. & Mahaney, W.C. (1991) Glacial interglacial cycles and development of the afroalpine ecosystem on East-African Mountains .2. Origins and development of the biotic component. *Journal of African Earth Sciences*, **12**, 513-523.
- Hastie T, Tibshirani R & Friedman J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. 2nd edition, Springer, New York, 746 pp.

- Hedberg, O. (1951) Vegetation belts of the East African mountains. *Svensk Botanisk Tidskrift*, **45**, 140-202.
- Hedberg, O. (1957) Afroalpine vascular plants. A taxonomic revision. *Symbolae Botanicae Upsalienses*, **15**, 1-411.
- Hedberg, O. (1964) Features of afroalpine plant ecology. *Acta Phytogeographica Suecica*, **49**, 1-144.
- Hedberg, O. (1970) Evolution of the Afroalpine flora. *Biotropica*, **2**, 16-23.
- Kebede, M., Ehrlich, D., Taberlet, P., Nemomissa, S. & Brochmann, C. (2007) Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and Tropical East African mountains. *Molecular Ecology*, **16**, 1233-43.
- Knox, E.B. & Palmer, J.D. (1998) Chloroplast DNA evidence on the origin and radiation of the giant lobelias in eastern Africa. *Systematic Botany*, **23**, 109-149.
- Koch, M.A., Kiefer, C., Ehrlich, D., Vogel, J., Brochmann, C. & Mummenhoff, K. (2006) Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology*, **15**, 825-839.
- Körner, C. & Paulsen, J. (2004) A world-wide study of high altitude treeline temperatures. *Journal of Biogeography*, **31**, 713-732.
- Körner, C. (2007) Climatic Treelines: Conventions, Global Patterns, Causes (Klimatische Baumgrenzen: Konventionen, globale Muster, Ursachen). *Erdkunde*, **61**, 316-324.
- Körner, C. (2012) *Alpine Treelines*. Springer, Basel.
- Lambin, E.F., Geist, H.J. & Lepers, E. (2003) Dynamics of land-use and land-cover change in tropical regions. *Annual Review of Environment and Resources*, **28**, 205-241.
- Mark, B.G. & Osmaston, H.A. (2008) Quaternary glaciation in Africa: key chronologies and climatic implications. *Journal of Quaternary Science*, **23**, 589-608.
- Masao, C., Gizaw, A., Piñeiro, R., Tusiime, F., Wondimu, T., Abdi, A., Popp, M., Gussarova, G., Lye, K., Munishi, P., Nemomissa, S. & Brochmann, C. (2013) Phylogeographic history and taxonomy of some afro-alpine grasses assessed based on AFLPs and morphometry: *Deschampsia cespitosa*, *D. angusta* and *Koeleria capensis*. *Alpine Botany*, **123**, 107-122.
- Monserud, R.A. & Leemans, R. (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, **62**, 275-293.
- Popp, M., Gizaw, A., Nemomissa, S., Suda, J. & Brochmann, C. (2008) Colonization and diversification in the African 'sky islands' by Eurasian *Lychnis* L. (Caryophyllaceae). *Journal of Biogeography*, **35**, 1016-1029.
- Rundel, P.W. (1994) Tropical alpine climates. *Tropical Alpine Environments: Plant Form and Function* (ed. by P.W. Rundel, A.P. Smith and F.C. Meinzer), pp. 21-44. Cambridge University Press, Cambridge.
- Shanahan, T.M. & Zreda, M. (2000) Chronology of Quaternary glaciations in East Africa. *Earth and Planetary Science Letters*, **177**, 23-42.
- Sklenář, P., Hedberg, I. & Cleef, A.M. (2014) Island biogeography of tropical alpine floras. *Journal of Biogeography*, **41**, 287-297.
- Smith, J.M.B. & Cleef, A.M. (1988) Composition and origins of the worlds tropical alpine floras. *Journal of Biogeography*, **15**, 631-645.

- Spence, J.R. & Pocs, T. (1988) Distribution patterns in the Afroalpine moss flora of East Africa. *Quaternary and Environmental Research on East African Mountains* (ed. by W.C. Mahaney), pp. 291-307. Balkema, Rotterdam.
- Therneau, T.M. & Atkinson, E.J. (2015) An introduction to recursive partitioning using the rpart routines. Available at: <https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>.
- Thompson, L.G., Mosley-Thompson, E., Davis, M.E., Henderson, K.A., Brecher, H.H., Zagorodnov, V.S., Mashiotta, T.A., Lin, P.N., Mikhalenko, V.N., Hardy, D.R. & Beer, J. (2002) Kilimanjaro ice core records: evidence of holocene climate change in tropical Africa. *Science*, 298, 589-93.
- Umer, M., Lamb, H.F., Bonnefille, R., Lezine, A.M., Tiercelin, J.J., Gibert, E., Cazet, J.P. & Watrin, J. (2007) Late Pleistocene and Holocene vegetation history of the Bale Mountains, Ethiopia. *Quaternary Science Reviews*, 26, 2229-2246.
- Wallace, A.R. (1876) *The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface*. Harper and brothers, New York.
- Wondimu, T., Gizaw, A., Tusiime, F., Masao, C., Abdi, A., Gussarova, G., Popp, M., Nemomissa, S. & Brochmann, C. (2014) Crossing barriers in an extremely fragmented system: two case studies in the afro-alpine sky island flora. *Plant Systematics and Evolution*, 300, 415-430.

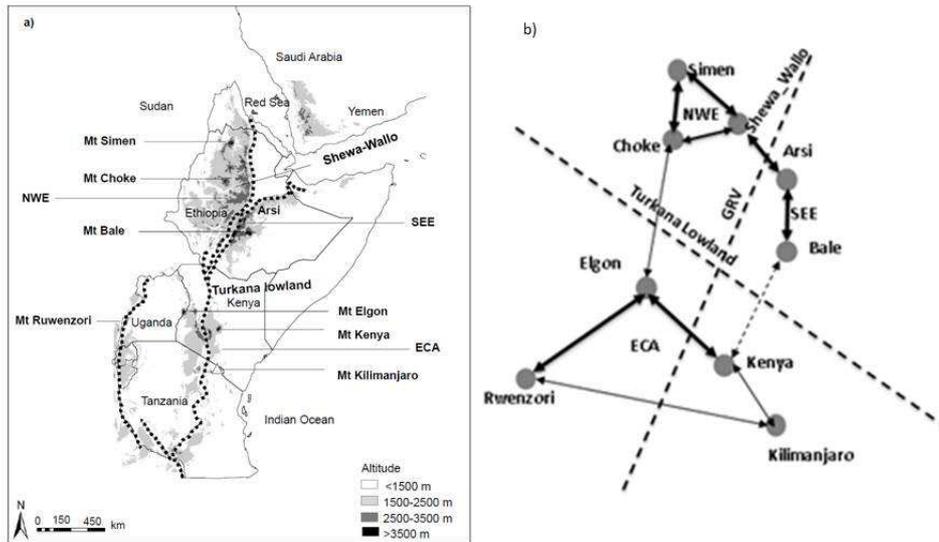


Figure 1: **a** Mountains and low-lying landscapes of eastern Africa. The eastern and western branches of the Great Rift Valley are indicated with *stippled lines*. **b** Potential habitat corridors for drought-tolerant alpine species that today extend below the treeline, assessed based on landcover modelling (*line thickness indicates potential corridor suitability*)

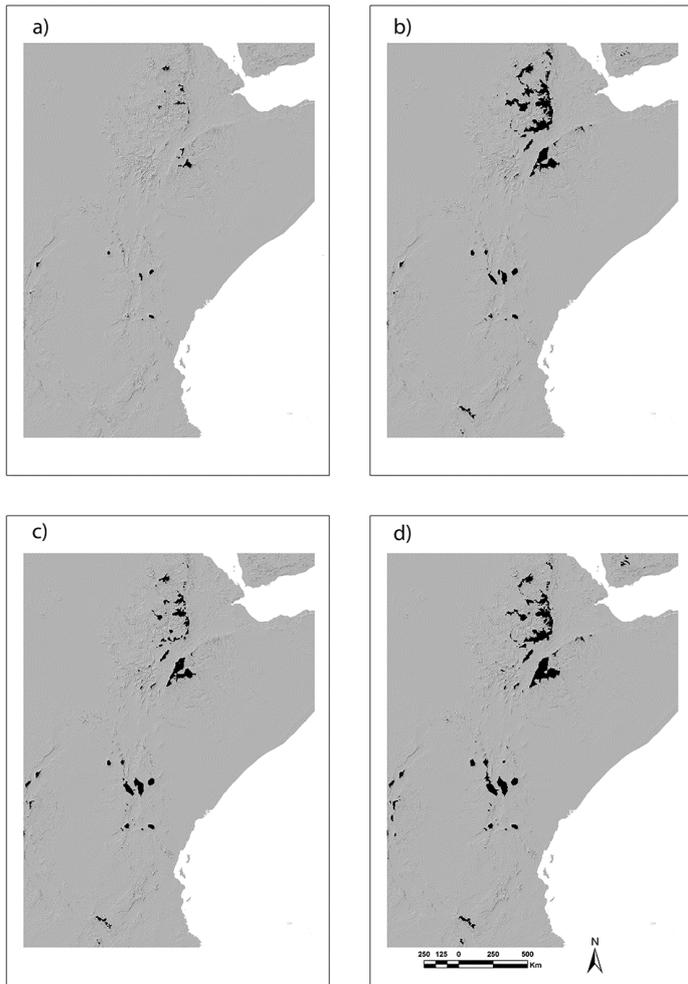


Figure 2: Extent of the afroalpine habitat in eastern Africa modelled based on all selected predictor variables under the current climate (**a**) and hindcasted to three different last glacial maximum climate models: CCSM4 (CC; **b**); MICRO-ESM (MR; **c**) and MPI-ESM-P (ME; **d**). The alpine habitats are indicated in *black*

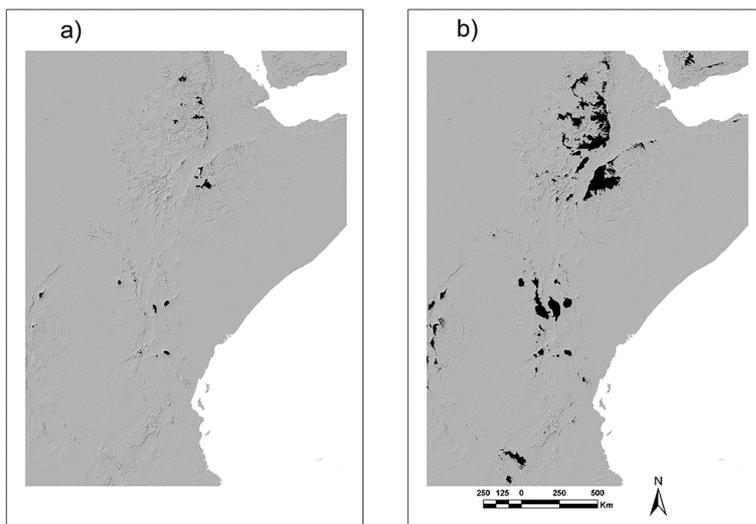


Figure 3: Extent of the afroalpine habitat in eastern Africa modelled based on current mean annual temperature and slope (**a**) and hindcasted to mean LGM temperature (**b**). The alpine habitats are indicated in *black*

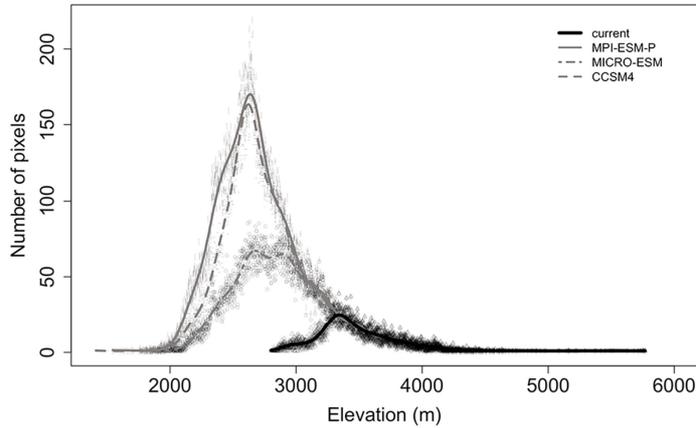


Figure 4: Altitudinal range of the alpine habitat in eastern Africa modelled based on all selected variables under current climate and hindcasted to three different LGM climate models: CCSM4; MICROESM and MPI-ESM-P. The y-axis represents a 1 km × 1 km grid cell count and the x-axis represents elevation (m)

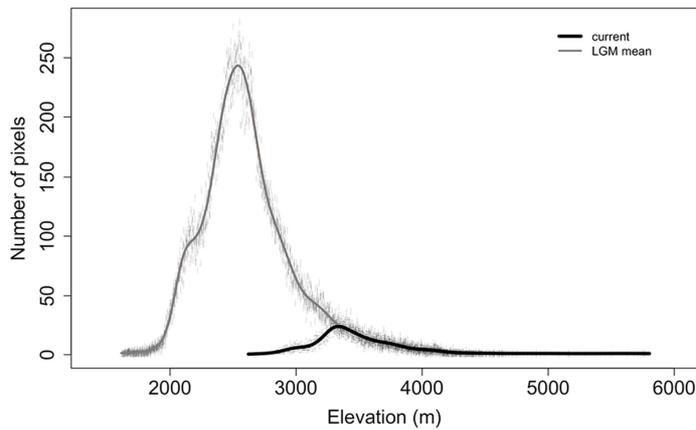


Figure 5: Altitudinal range of the alpine habitat in eastern Africa modelled based on mean annual temperature and slope (current) and hindcasted to mean annual LGM temperature (LGM mean). The y-axis represents a 1 km × 1 km grid cell count and the x-axis represents altitude (m)

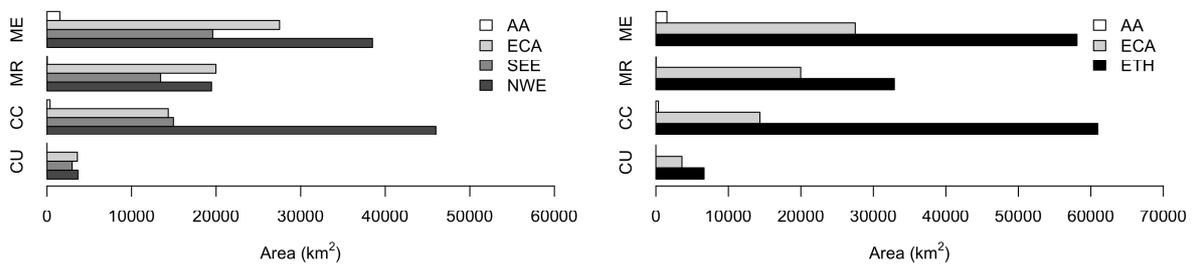


Figure 6: Spatial extent of alpine habitat modelled with all selected environmental variables under current (Cu) and hindcasted to three last glacial maximum climate models (CCSM (CC), MICRO-ESM (MR) and MPI-ESM-P (ME)). AA Yemen (araboalpine), ECA East/ Central Africa, SEE southeastern Ethiopia, NWE northwestern Ethiopia, ETH Ethiopia

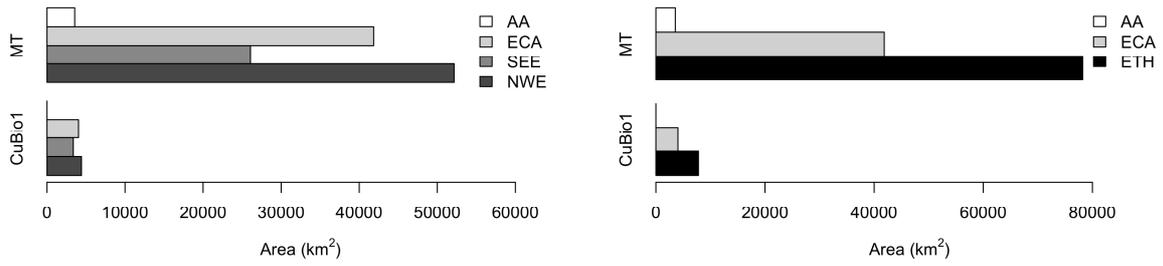


Figure 7: Spatial extent of alpine habitat modelled based on mean annual temperature (CuBio1) and slope and hindcasted to mean annual LGM temperature (MT). AA Yemen (arboalpine), ECA East/Central Africa, SEE southeastern Ethiopia, NWE northwestern Ethiopia, ETH Ethiopia

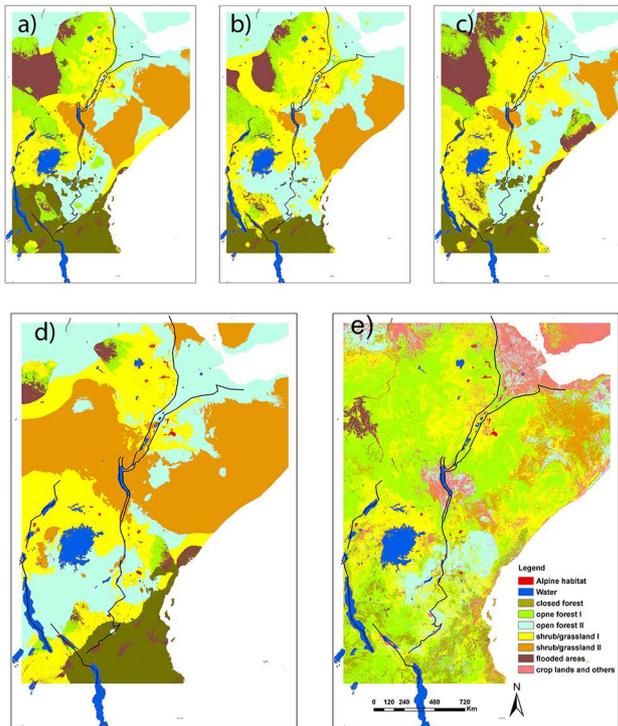


Figure 8: Spatial extent of the landcover types in eastern Africa under current (a) and three LGM climate models: MICRO-ESM (MR; b), MPI-ESM-P (ME; c) and CCSM4 (CC; d) and after reclassification of the GlobCover in to major landcover units (e). Mountains with alpine zones, water bodies and the Great Rift Valley are superimposed

Table 1: Reclassification of GlobCover into eight classes of generalized landcover units used in the present paper

Generalized legend	GlobCover	Labels
Closed forest	Closed (>40%) broadleaved deciduous forest (>5m)	50
	Closed (>40%) needleleaved evergreen forest (>5m)	70
Open forest I	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)	40
	Open (15-40%) broadleaved deciduous forest/woodland (>5m)	60
	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	100
	Mosaic forest or shrubland (50-70%) / grassland (20-50%)	110
	Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)	130
Open forest II	Open (15-40%) needleleaved deciduous or evergreen forest (>5m)	90
	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)	140
Shrub/grassland I	Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)	20
	Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)	30
Shrub/grassland II	Mosaic grassland (50-70%) / forest or shrubland (20-50%)	120
	Sparse (<15%) vegetation	150
Flooded areas	Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily) - Fresh or brackish water	160
	Closed (>40%) broadleaved forest or shrubland permanently flooded - Saline or brackish water	170
	Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil - Fresh, brackish or saline water	180
Croplands and others	Post-flooding or irrigated croplands (or aquatic)	11
	Rainfed croplands	14
	Artificial surfaces and associated areas (Urban areas >50%)	190
	Bare areas	200
	Water bodies	210

Table 2: Contribution of variables to the models

Models					
alpine vs. non-alpine				landcover	
complex		simple			
variables	Importance (%)	Variables	Importance (%)	Variables	Importance (%)
Bio1	29	Bio01	60	Bio16	19
Slope	20	slope	40	Bio01	15
Bio18	16			Bio19	14
Bio16	13			slope	13
Bio19	12			bio18	9
Bio3	10			bio03	9
				bio15	9
				bio17	6

Table 3: Model performance

Evaluation method	Data	Models		
		alpine vs. non-alpine		landcover
		simple model	complex model	
misclassification error rate	training	0.00045	0.00042	0.4362
	test	0.00094	0.00093	0.4506
Kappa value	training	0.999	0.999	0.477
	test	0.998	0.999	0.46