Sweet vernal grasses (*Anthoxanthum*) colonized African mountains along two fronts in the Late Pliocene, followed by secondary contact, polyploidization and local extinction in the Pleistocene

Felly Mugizi Tusiime [1, 2], Abel Gizaw[2, 3], Tigist Wondimu[2, 3], Catherine Aloyce Masao[2, 4], Ahmed Abdikadir Abdi[2, 5], Vincent Muwanika[1], Pavel Trávníček[6], Sileshi Nemomissa[3], Magnus Popp[2], Gerald Eilu[1], Christian Brochmann[2#] and Manuel Pimentel[2, 7#]

[1]School of Forestry, Geographical and Environmental Sciences, Department of Forestry, Biodiversity and Tourism, Makerere University, P.O. Box 7062 Kampala, Uganda; [2]Natural History Museum, University of Oslo, PO Box 1172 Blindern, NO-0318 Oslo, Norway; [3]Department of Plant Biology and Biodiversity Management, Addis Ababa University, P. O. Box 3434 Addis Ababa, Ethiopia; [4]Sokoine University of Agriculture, Department of Forest Biology, P.O. Box 3010, Morogoro, Tanzania; [5]National Museums of Kenya, P. O. Box 40658-00100 Nairobi, Kenya; [6]Institute of Botany, Department of Flow Cytometry, CZ-252 43 Průhonice, Czech Republic; [7]CICA, Centro de Investigacións Científicas Avanzadas, Universidade da Coruña, Galicia, Spain.

Keywords: Africa, colonization, hybridization, polyploidization, tropical-alpine

*For correspondence: Manuel Pimentel. [7]*CICA, Centro de Investigacións Científicas Avanzadas, Universidade da Coruña, Galicia, Spain;* mailto: mpimentel@udc.es; telephone: 0034 981167000. Fax: 0034 981167065.

Running title: Grass jumping across the African sky islands

shared senior authorship

Abstract

High tropical mountains harbor remarkable and fragmented biodiversity thought to a large degree to have been shaped by multiple dispersals of cold-adapted lineages from remote areas. Few dated phylogenetic/phylogeographic analyses are however available. Here we address the hypotheses that the sub-Saharan African sweet vernal grasses have a dual colonization history and that lineages of independent origins have established secondary contact. We carried out range-wide sampling across the eastern African high mountains, inferred dated phylogenies from nuclear ribosomal and plastid DNA using Bayesian methods, and performed flow cytometry and AFLP (Amplified Fragment Length Polymorphism) analyses. We inferred a single Late Pliocene Eurasian origin of the eastern African taxa. The putative dodecaploid populations in one mountain group formed a distinct phylogeographic group and carried plastids that diverged from those of the currently allopatric southern African lineage in the Mid- to Late Pleistocene. We show that *Anthoxanthum* has an intriguing history in sub-Saharan Africa, including Late Pliocene colonization from southeast and north, followed by secondary contact, hybridization, allopolyploidization, and local extinction during one of the last glacial cycles. Our results add to a growing body of evidence showing that isolated tropical high mountain habitats have a dynamic recent history involving niche conservatism and recruitment from remote sources, repeated dispersals, diversification, hybridization, and local extinction.

Introduction

High tropical mountains house exceptional ecosystems with remarkable biodiversity, peculiar life forms and high levels of endemism (Gehrke & Linder 2014; Merckx et al. 2015), but our knowledge of the temporal and geographical origin and evolution of this diversity is limited. The highest peaks are inhabited by organisms adapted to extreme conditions such as nightly frosts and constitute highly fragmented 'sky archipelagos' interrupted by tropical lowland biotas. The upper parts of high tropical mountains appear to have been mostly colonized by long-distance dispersed organisms that already had been cold-adapted in remote areas of the world (Hedberg 1970, 1992; von Hagen & Kadereit 2001; Bell & Donoghue 2005; Assefa *et al*. 2007; Ehrich *et al*. 2007; Popp *et al*. 2008; Gehrke & Linder 2009; Merckx *et al*. 2015; Gizaw *et al*. 2016a). The patterns of recruitment of these long-distance dispersers are still poorly known (Gehrke & Linder 2009), but their high degree of endemicity has been used to suggest that they arrived to the tropics a long time ago (Hedberg, 1961). New evidence, however, points to a relatively recent origin of tropical-alpine lineages (Pliocene-Pleistocene; Linder 2014; Hughes & Atchison 2015; Merckx *et al*. 2015; Hughes 2016 but see Gizaw *et al*. 2016a), but only few dated phylogenies are available.

The high mountain systems in eastern Africa (i.e. East Africa and Ethiopia) provide an excellent model for the study of biogeographical questions such as colonization processes and routes, the role of climatic refugia in the preservation of genetic diversity, and the role of gene flow, hybridization and polyploidization in an extremely fragmented system (Sklenář *et al*. 2014; Wondimu *et al*. 2014; Gizaw *et al*. 2016b). Most of these mountains emerged in connection with the tectonic activity forming the East African Rift System (EARS; Baker *et al*. 1972; Ebinger *et al*. 2000), consisting of one western and one eastern branch, the latter extending through East Africa to Ethiopia. With a few exceptions such as Mt Ruwenzori,

which is a block mountain that emerged 8-3 million years ago (Mya), the mountains have volcanic origins and vary considerably in age (Gehrke & Linder 2014). Most formed during the Late Miocene to the Pleistocene, with the Ethiopian mountains (>40-7 Mya) and Mt Elgon being the oldest (23-12 Mya), and Mt Kilimanjaro (2.5-1 Mya) and Mt Meru (2.0-0.06 Mya) among the youngest (Gehrke & Linder 2014).

The alpine vegetation in eastern Africa mainly consists of endemic species (~80% in vascular plants; Hedberg 1957, 1969; Gehrke & Linder, 2014). Some endemics occur exclusively in a single mountain, but many of them are found along both branches of the Rift Valley and both in East Africa and Ethiopia, demonstrating that widespread intermountain dispersal has taken place after initial colonization of the region. The afro-alpine plant communities are mostly composed of C3 taxa that appear to have their closest relatives in montane and alpine areas in southern Africa, Eurasia and even the Americas, rather than in the surrounding afro-tropical, often C4-dominated plant communities (Hedberg 1970, 1992; Assefa *et al*. 2007; Ehrich *et al*. 2007; Popp *et al.* 2008; Gehrke & Linder 2009; Sikolia *et al.* 2009; Pimentel *et al*. 2013; Gehrke *et al*. 2016; Gizaw *et al*. 2016a).

A central question is whether the frost-tolerant plant lineages inhabiting the eastern African mountains colonized the mountains as soon as they formed, which could help explaining the high degree of afro-alpine endemism (cf. Hedberg 1961), or whether the initial colonization of these mountains primarily was facilitated by the Plio-Pleistocene climatic oscillations (e.g. Assefa *et al*. 2007). The dated phylogenies available do not point to a single answer. *Arabis alpina* was inferred to have colonized eastern Africa twice in the Pleistocene (Koch *et al*. 2006), and afro-alpine *Alchemilla* also appear to have evolved during the Pleistocene (Gehrke *et al*. 2016). Late Pliocene Eurasian origin was inferred for the afro-alpine *Anthoxanthum nivale* K.Schum. (Pimentel *et al*. 2013). A few broader phylogenies dated using fossil-

calibrated clocks have included at least one species occurring at high altitudes in eastern Africa, with origins ranging from the Miocene to the Pleistocene (Antonelli 2009; Jabbour & Renner 2012; Nürk *et al*. 2015; Gizaw *et al*. 2016a).

Although some lineages apparently colonized the first eastern African mountain from remote areas several millions of years ago, subsequent colonization of other mountains may have happened much later and/or repeatedly. In *Lychnis*, we estimated that several species originated in the Late Pliocene, but the terminal branches within species were often short, indicating that the contemporary populations in many mountains result from recent colonization (within the last 0.3 Myr; Gizaw *et al*. 2016a). In some species, it appears that a single Ethiopian population may have been the source for colonization of remote mountains in East and even West Africa during the last glacial cycles of the Pleistocene. A similar pattern with old (Miocene) species but recent (Pleistocene) intraspecific divergence has also been demonstrated in an afro-montane forest species (*Canarina eminii*; Mairal *et al*. 2015). Such patterns can be explained by extinction of intermediate populations (Antonelli & Sanmartin 2011; Mairal *et al*. 2015). Thus, it is possible that the eastern African mountains have experienced cycles of colonization, extinction, and recolonization of the same species during the Pleistocene climatic oscillations, and that many contemporary populations are recent re-colonists. Recent phylogeographic studies also suggest that afro-alpine species have a dynamic history involving intermountain divergence, repeated long-distance dispersals and hybridization, but the inferred histories differ considerably from species to species and among different mountains (Gizaw *et al*. 2013, 2016b; Masao *et al*. 2013; Wondimu *et al*. 2014).

Detailed studies of more plant groups that are suspected to have diverged at different times and with different biogeographic affinities are needed to better understand the history of the enigmatic afro-alpine habitat. Here we aim to reconstruct the history of the sub-Saharan sweet

126 vernal grasses (*Anthoxanthum* L.; Fig. 1a) by constructing dated phylogenies coupled with a
127 phylogeographic analysis of the eastern African populations. Our preliminary analyses
128 suggested that the sub-Saharan *Anthoxanthum* entered Africa at least twice in the Late
129 Pliocene, once to southern Africa from Southeast Asia (section *Ataxia*), and at least once to
130 eastern Africa from Eurasia (section *Anthoxanthum*; Pimentel *et al*. 2013). Two species of
131 *Anthoxanthum* occur in eastern Africa, the East African endemic *A. nivale* and the Ethiopian
132 endemic *Anthoxanthum aethiopicum* I.Hedberg. However, only two specimens from one
133 population of *A. nivale* and none of *A. aethiopicum* were included in our previous study.
134 *Anthoxanthum nivale* occurs in most East African mountains on moist ground in grassland
135 and moorland between 2400 and 4800 m (Fig. 1b; Clayton 1970). Two cytotypes are known,
136 tetraploids with $2n = 4x = 20$ and dodecaploids with $2n = 12x = 60$ (Hedberg 1957, 1976).
137 *Anthoxanthum aethiopicum* is tetraploid and only known from a few moist sites between 2700
138 and 4500 m in the Bale and Arsi Mountains (Hedberg 1976, Phillips 1995).

139 Here we address the origin and evolution of the sub-Saharan African *Anthoxanthum* by
140 including the Ethiopian *A. aethiopicum* and by extending our sampling of the East African *A.*
141 *nivale* to represent its entire distribution area. In particular, we address whether the eastern
142 African (i.e. East African and Ethiopian) sweet vernal grasses have been in secondary contact
143 with the southern African lineage of the genus, because both northern and southern affinities
144 have been suggested for *A. nivale* based on morphology. It has been suggested to be most
145 closely related to *A. aethiopicum* and European plants of the *Anthoxanthum odoratum* L. s.l.
146 complex (Hedberg 1976; Phillips 1995), whereas Clayton (1970) reported it as
147 morphologically variable with some specimens resembling the southern African endemic
148 *Anthoxanthum ecklonii* Stapf. (section *Ataxia*), which extends northwards to Malawi (Fig. 1a).
149 In addition, the occurrence of two cytotypes in *A. nivale* raises the intriguing possibility that

150 the two independently immigrated and currently allopatric lineages of sub-Saharan
151 *Anthoxanthum* once have been in secondary contact, resulting in hybridization and
152 polyploidization. To address these hypotheses we carried out range-wide field sampling
153 across the eastern African high mountains, sequenced nuclear ribosomal (ITS, ETS) and
154 plastid (*trn*T-L, *trn*L-F) DNA, reconstructed dated phylogenies using Bayesian methods, and
155 performed flow cytometry and AFLP (Amplified Fragment Length Polymorphism) analyses.

156

Materials and Methods

Plant materials

We sampled 35 *Anthoxanthum nivale* populations (150 plants; Fig. 1a,b; Appendix 1) from seven mountains/mountain systems in Uganda, Kenya and Tanzania. Because we found strong genetic differentiation between one eastern and one western mountain group, we will refer to 'eastern *A. nivale*' as the populations occurring on the four mountains east of the eastern branch of the Rift Valley (Mt Meru, Mt Kilimanjaro, Mt Aberdare, Mt Kenya), and 'western *A. nivale*' as the populations occurring on Mt Elgon west of the eastern branch and on two mountains along the western branch (Mt Ruwenzori, Mt Muhavura). The sampling covered the total distribution range of the species except for three small areas close to the sampled mountains (Hedberg 1957, Clayton 1970). Five (if possible) plants separated by at least 10 m were sampled to represent one population. Fresh leaf tissue was dried and three of the plants were pressed and deposited in the following herbaria: one at the National Herbarium of Ethiopia, Addis Ababa University, Ethiopia (ETH); one at the Natural History Museum, University of Oslo, Norway (O); and one in the country of collection: East African Herbarium, National Museum of Kenya, Nairobi, Kenya (EAH); National Herbarium of Tanzania, Arusha, Tanzania (NHT); or Makerere University Herbarium, Kampala, Uganda (MHU). We were not able to find any *Anthoxanthum aethiopicum* plants in the field. Thus, this species could not be included in the AFLP or flow cytometry analyses, but we used three samples from the Uppsala University Herbarium (UPS) successfully for DNA sequencing.

Flow cytometry

All silica-dried *A. nivale* samples were analysed using flow cytometry. Five plants were analysed twice to check for errors. DNA ploidy level and relative nuclear DNA-content of

somatic cells were determined following Schönswetter *et al*. (2007a). *Bellis perennis* L. (2C = 3.38 pg; Schönswetter *et al*. 2007b) was used as an internal standard. Fluorescence intensity was measured following the modified two-step Otto procedure of Suda and Trávníček (2006). The relative fluorescence intensity of at least 5000 DAPI-stained particles was estimated using a Partec PA II flow cytometer (Partec GmbH, Münster, Germany) equipped with a HBO-100 mercury arc lamp. Results were expressed as high-resolution histograms of fluorescence intensity. We were not able to obtain chromosome numbers for our samples (the seeds did not germinate despite several attempts). We divided the data into different DNA-content groups (Supporting Information Figs. S1, S2) which likely represent different ploidy levels. Differences between groups were tested using a univariate analysis of variance (ANOVA). Statistical analyses were conducted using IBM SPSS (IBM, Armonk, USA).

DNA extraction

Total genomic DNA was extracted from the silica-dried leaves of *A. nivale* using an automated GeneMole® robot and the MoleStrips™ DNA Plant Kit following the manufacturer's protocol (Mole Genetics AS, Lysaker, Norway). About 1 cm^2 leaf tissue was ground in 2.0 ml crushing tubes with two tungsten carbide beads for 4 min at 23 Hz in a mixer mill (MM301, Retsch® GmbH & Co., Haan, Germany). DNA from the herbarium material of *A. aethiopicum* was extracted using the DNAeasy Plant Extraction Kit (Qiagen, Hilden, Germany) using the manufacturer's protocol. The quality of the extracted DNA was checked on 1% TBE-agarose gels and DNA was quantified using an UV-Vis spectrophotometer (Nanodrop, ThermoScientific, Wilmington, USA).

DNA sequencing and sequence alignment

For *A. nivale*, 1-3 plants from each of 17 populations representing its entire distribution range were selected for sequencing (Appendix 2). For *A. aethiopicum*, we obtained DNA sequences from all three available herbarium specimens. Amplification and sequencing of the four DNA regions (*internal transcribed spacer*, ITS, and *external transcribed spacer*, ETS of nuclear rDNA, and *trn*L-F and *trn*T-L of plastid DNA) followed Pimentel *et al.* (2013). Because we did not observe multiple bands for ITS or ETS in any accessions, no cloning was conducted. A total of 96 new sequences were generated for this study (27 ETS, 24 ITS, 24 *trn*L-F and 22 *trn*T-L; Appendices 2, 3). We also included sequences from 16 *Anthoxanthum* and *Hierochloë* R.Br. specimens from Pimentel *et al*. (2013), including representatives of southern African species. Nine additional sequences representing the main lineages in the subfamily Pooideae were retrieved from GenBank and used as outgroups (Appendix 3). The DNA regions were separately aligned using the MUSCLE algorithm (Edgar, 2004) as implemented in the software SeaView v4 (Gouy *et al*. 2010) and manually adjusted. Indels were excluded from all analyses.

Phylogenetic analyses and dating

We conducted Bayesian analyses using MrBayes v. 3.2.5 (Huelsenbeck & Ronquist 2001). The GTR+I+G substitution model was selected based on MrModelTest v. 2.3 (Nylander, 2004); indels were not coded. The plastid and nuclear datasets were analyzed separately for 15000000 generations (25% burn-in) with sampling every 1000 generations and allowing the program to estimate the likelihood parameters required. We assessed convergence using (i) the 'compare' function in the online application AWTY (Nylander *et al*. 2008) and (ii) TRACER v.1.5 (with the Effective Sample Size-ESS >200 for all parameters; Rambaut & Drummond 2007). Results collected prior to stationarity were discarded as burn-in. Results were presented as majority rule posterior probability consensus trees, summarised using

MrBayes. Indels were coded following the simple method by Simmons & Ochoterena (2000) and all phylogenetic analyses were conducted with and without considering the indels. No improvement in posterior probability support was obtained when indels were included so they were eliminated from the final analyses.

Bayesian divergence date analyses were conducted on the nuclear and plastid data sets using BEAST v. 1.8.1 (Drummond *et al*. 2012). Samples for which not all DNA regions had been successfully sequenced were excluded from this analysis. Input data for BEAST were compiled using BEAUTI v.1.7.2, and the strict clock was selected for all analyses following Drummond & Bouckaert (2015). A multispecies coalescence model as implemented in *BEAST (Heled & Drummond 2010) was applied since most species were represented by more than one sample. Other model priors were set as follows: (i) the date of divergence between tribe Meliceae and tribes Brachypodieae + Aveneae/Poaeae + Triticeae, normal prior distribution with mean 32.1 million years (My) and standard deviation 3.65; (ii) the date of divergence between tribes Aveneae/Poaeae and Triticeae, normal prior distribution with mean 23.4 My and standard deviation 3.10, and (iii) log-normal distributions and broad ranges spanning all biologically realistic values were established for substitution rates following Drummond & Bouckaert (2015). All secondary calibration ages (mean and standard deviation) were taken from Vicentini *et al*. (2008).

In the *BEAST analyses, a first test was conducted giving all *A. nivale* sequences the same taxonomic label, disregarding the geographic origin or their position in the plastid *vs* nuclear trees. Next, multi-labelled specimen trees were built in order to unravel the origin of the progenitor lineages of clades putatively affected by reticulation processes (Pirie *et al*. 2009; Popp *et al*. 2011; Blanco-Pastor *et al*. 2012). In this approach, species trees were reconstructed by assigning unique taxon labels to plastid and nuclear sequences from samples

that displayed incongruence between the plastid and nuclear phylogeny and therefore were suspected to be putative hybrids. "N" and "P" were added to the names of nuclear and plastid sequences, respectively, thus treating plastid and nuclear sequences as belonging to different taxa. The matrices were balanced by inserting "empty taxa" for uniquely labelled entries. Thus, each sequence in the nuclear matrix labelled "N" had a corresponding entry consisting of missing data labelled "N" in the plastid matrix, and each sequence in the plastid matrix labelled "P" had a corresponding entry consisting of missing data labelled "P" in the nuclear matrix (Pirie *et al*. 2009; Blanco-Pastor *et al*. 2012). Three MCMC analyses were run for 15 x 10^7 generations each with a sample frequency of 1 x 10^4. One extra analysis was run for 5 x 10^7 generations without data to test the influence of priors on posterior values following Heled & Drummond (2010). Log files were analysed using TRACER v1.5 to assess convergence. Maximum credibility trees were built using TreeAnnotator v.1.7.2 (Drummond & Rambaut, 2007).

Phylogenetic conflict analysis

Conflict between nuclear and plastid trees was assessed by comparing nodes with posterior probability support (PPS) ≥0.8 and illustrated by means of a tanglegram of the Bayesian consensus trees using Dendroscope 3 (Huson & Scornavacca 2012). We used coalescent simulations in Mesquite (Maddison & Maddison 2009) in order to test whether gene tree differences could be explained by coalescent stochasticity (Maureira-Butler *et al*. 2008; Mugrabi de Kuppler *et al*. 2015). This method assumes (i) known clade ages (ultrametric, time calibrated trees), (ii) constant generation time and effective population size, and (iii) panmixis within populations (Mugrabi de Kuppler *et al*. 2015). Two alternative species trees (plastid and nuclear) were constructed representing the two phylogenetic hypotheses for *A. nivale* by pruning the multilabelled Bayesian species tree built with *BEAST. Each species

was reduced to a single accession and all samples labelled as plastid were pruned from the nuclear species tree and vice versa. Only *Helictotrichon* was kept as outgroup. Two corresponding gene trees were also built from the multilabelled gene trees keeping all ingroup accessions and *Helictotrichon* (samples labelled as plastid were pruned from the nuclear gene tree and vice versa). Terminals were pruned using the R-based package APE v. 2.7-3 (Paradis *et al*. 2004) and the pruned trees were re-scaled using Mesquite. Divergence times were translated from millions of years to generations assuming a generation time of 1 year for *A. nivale* and *A. aethiopicum*, because closely related species such as *A. odoratum* and *Anthoxanthum amarum* Brot. usually flower in the first year (M. Pimentel, pers. obs.). We used the "Coalescent Contained within Current Tree" module of Mesquite to simulate samples of 1000 gene trees from the species trees and gene trees built with APE. Constant effective population sizes (N_e) were assumed to range from 1×10^4 to 1×10^6. We used the partition metric (Penny & Hendy 1985) implemented in PAUP 4.0b10 (Swofford 2002) as the symmetric distance to estimate the difference between the plastid and nuclear gene trees and the distribution of differences between the original gene trees pruned with APE and the simulated gene trees (Mugrabi de Kuppler *et al*. 2015). The null hypothesis of lineage sorting stochasticity should be rejected when the distance between the two gene trees is higher than 95% of the distribution of tree to tree differences of simulated trees from their respective gene trees (Maureira-Butler *et al*. 2008).

AFLP analysis of A. nivale

AFLP fingerprinting followed Gaudeul *et al*. (2000). A preliminary test was conducted using 15 primer combinations and 16 plants representing all mountains. Three primer combinations were selected for final analysis: [6-FAM (*Eco*RI-AGA/ *Mse*I -CCG), VIC (*Eco*RI-AGG/ *Mse*I -CTG) and NED (*Eco*RI-ACC/ *Mse*I -CAT)]. A total of 126 plants were retained in the AFLP

matrix after removal of samples that did not amplify (Fig. 1a,b; Appendix 1). For each sample, 2.0 µl 6-FAM, 2.0 µl VIC and 3.0 µl NED labelled selective PCR products were mixed with 11.7 µl formamide and 0.3 µl GeneScan ROX internal lane size standard (Applied Biosystems, Foster City, USA) and run on an ABI 3100 sequencer (Applied Biosystems, Foster City, USA). Thirteen samples (~10%) were duplicated (i.e. DNA extracted twice) for a reproducibility test (Bonin *et al.* 2004). Data analysis and scoring were conducted following Masao *et al*. (2013).

The final AFLP data matrix was transformed using the R-script AFLPdat (Ehrich, 2006). Genetic diversity (calculated as percentage of polymorphic markers; %PL), Nei's average gene diversity (D; Nei 1987; Kosman 2003), and Nei's unbiased expected heterozygosity (He; Nei 1987; Gaudeul *et al*. 2000) were estimated using Arlequin v.3.5 (Excoffier & Lischer 2010). Genetic rarity was calculated as down-weighed marker values (DW) following Schönswetter and Tribsch (2005) with modifications implemented in AFLPdat (Ehrich 2006). We calculated pairwise genetic similarity among AFLP phenotypes using Dice coefficient in NTSYSpc v.2.11a (Rohlf 2002), visualized using Principal Coordinate Analyses (PCoAs). A Neighbor Joining (NJ) tree was constructed based on Nei & Li's (1979) genetic distance using the software TREECON 1.3b (Van-de-Peer & De-Wachter 1994). The tree was midpoint rooted and support for branches was estimated from 1000 bootstrap replicates.

Genetically homogenous groups were inferred from Bayesian clustering analysis using STRUCTURE version 2.3.3 (Pritchard *et al*. 2000). We used the recessive allele model taking into account the dominant nature of AFLP markers (Falush *et al*. 2007). Both the admixture and the non-admixture models with uncorrelated allele frequencies were tested. Based on the result of the preliminary analysis we selected the admixture model with correlated allele frequencies for the final analysis. The analyses were performed at the Lifeportal, University

of Oslo (http://www.lifeportal.uio.no) with number of genetic groups (*K*) ranging from 1 to 10 with 10 replicate runs for each *K* and a burn-in period of 2 x 10^5 and 10^6 iterations. Separate tests were performed for each main genetic group to assess additional structure. We used the STRUCTURE-SUM R-script (Ehrich *et al*. 2007) to summarize the results and to infer the optimal value of *K* based on the estimated posterior log likelihood of the data *L*(*K*), the similarity among different runs for the same *K* (Nordborg *et al.* 2005), and the rate of change in probability between successive runs, Δ*K*, as a function of *K*, calculated following Evanno *et al*. (2005). The average estimate of individual admixture values among the replicated runs for the selected optimal *K* was calculated using the program CLUMPP (Jakobsson & Rosenberg 2007) and the result was visualized using the program DISTRUCT (Rosenberg 2004).

Partitioning of genetic variation in the total dataset was explored using analyses of molecular variance (AMOVAs, Excoffier *et al*. 1992) in Arlequin version 3.5 (Excoffier & Lischer 2010). For each analysis, 10000 permutations were performed to assess the significance of the results. Genetic differentiation was also estimated as the unbiased F_{st} (θw) of Weir and Cockerham (1984) using the same software. The 95% confidence interval for the estimator was obtained by bootstrapping 1000 replicates over loci.

Results

Nuclear DNA-content

The five plants duplicated in the flow cytometry analysis did not reveal any errors. A total of 110 samples (of 130 attempted) were successfully analysed. These were tentatively divided into three DNA-content groups (G1-G3) except for ten plants that were considered as outliers (Supporting Information Figs. S1, S2). The mean relative DNA content values (as compared to unit value of standard plant *Bellis perennis*) of the three groups were significantly different in the analysis of variance: G1: 1.60 ± 0.09 (7 plants), G2: 2.89 ± 0.23 (26 plants), and G3: 4.78 ± 0.23 (67 plants). G1 plants were rare, only occurring in the two westernmost mountains, G2 plants occurred in all three western mountains, and G3 plants were restricted to the four eastern mountains (Fig. 1b).

Phylogenetic reconstruction, divergence dating and phylogenetic conflict analysis

We obtained congruent topologies for the two plastid markers and for the two nuclear markers, but conflict between the plastid and nuclear topologies affecting section *Ataxia* (Pimentel *et al*. 2013) and the eastern high polyploid *A. nivale*. The plastid and nuclear datasets were therefore not combined. The aligned plastid DNA matrix comprised 43 terminals (Appendices 2, 3) and 1851 characters (1-1094, *trn*L-F; 1095-1851, *trn*T-L). Monophyly was rejected for *A. nivale* in the plastid tree (Fig. 2). Two major plastid clades were recovered. One contained the western low-ploid (DNA-content groups G1 and G2) samples of *A. nivale* together with *A. aethiopicum* and the European species *A. alpinum* and *A. odoratum*. The other major clade contained the eastern high-ploid (G3) samples of *A. nivale* together with the southern African.

The aligned nuclear DNA matrix included 42 terminals (Appendix 1) and 1385 characters (1-747, ETS; 748-1385, ITS). In the nuclear tree, the southern African species formed a group sister to all eastern African (*A. nivale* and *A. aethiopicum*) and European accessions (Fig. 3). The eastern African accessions formed a fully supported subclade as sister to a poorly supported group consisting of the European species *A. alpinum* and *A. odoratum*. *Anthoxanthum aethiopicum* was recovered as a monophyletic group in a trichotomy with the two fully supported subclades corresponding to the western low-ploid (G1 and G2) accessions and the eastern high-ploid (G3) accessions of *A. nivale*.

In the *BEAST analysis, treating plastid and nuclear sequences of the eastern *A. nivale* accessions as different taxa resulted in a more resolved tree than when treating them as the same taxon. The plastid and nuclear topologies from the *BEAST analyses (Supporting Information Figs. S3, S4) were congruent with those from MrBayes (Figs. 2, 3). In the multilabelled species tree (Fig. 4), the plastid sequences of the high-ploid eastern *A. nivale* accessions were recovered in a southern African clade as sister to *A. dregeanum*, from which they diverged in the Mid- to Late Pleistocene [mean 0.125 Mya, 95% highest posterior density (HPD) 0.0-0.7 Mya; Fig. 4, A]. In contrast, the nuclear sequences of these eastern *A. nivale* accessions were recovered in a fully supported clade with all other *A. nivale* and *A. aethiopicum* sequences. This eastern African clade was estimated to have diverged from its European sister during the Late Pliocene (mean 2.882 Mya, HPD 2.0-3.9 Mya; Fig. 4, B). The eastern accessions of *A. nivale* were estimated to have diverged from the western ones in the Late Pliocene or Early- to Mid Pleistocene (mean 1.37 Mya, HDP 0.4-2.6 Mya; Fig. 4, C). The phylogenetic conflict analysis further demonstrated the incongruence between the plastid and nuclear topologies (Supporting Information Fig. S5). The coalescent simulations resulted in a symmetric distance of 48 between the plastid and nuclear gene trees constructed. The null

386 hypothesis of lineage sorting stochasticity was rejected in all analyses with Ne below 400000
387 based on plastid data and below 250000 based on nuclear data (Fig. 5).

388 *AFLP variation in* A. nivale

389 The final AFLP dataset contained 424 polymorphic markers scored in 126 plants from 35
390 populations (Appendix 1). Reproducibility was 97.7%. Two genetic groups, one containing
391 all western populations (DNA-content groups G1 and G2) and one containing all eastern
392 populations (G3), were inferred in the STRUCTURE analyses (K = 2; Figs. 1b, 6; Supporting
393 Information Fig. S6). In the western group, two subgroups corresponding to the two DNA-
394 content groups (G1 and G2) were inferred in separate analyses. In the eastern group,
395 subdivisions reflected geography but the optimal number of subgroups was unclear. The
396 PCoA and NJ analyses revealed a similar structuring of the dataset (Figs. 6B, 6C; Supporting
397 Information Fig. S7). The first two axes in the PCoA analysis separated the plants into three
398 distinct groups corresponding to the three DNA-content groups (Fig. 6B). The three DNA-
399 content groups were also recovered in the NJ tree (Fig. 6C).

400 Nei's genetic diversity (D) across all AFLP phenotypes was 0.185 ± 0.089, mean
401 heterozygosity (H_e) was 0.195 ± 0.162, and mean number of polymorphic loci (PL) was
402 51.2% ± 37.5%. Intrapopulation genetic diversity ranged from 0.031 to 0.190 (Appendix 1).
403 Genetic diversity (D) and rarity (DW) were highest in the two western Rift mountains
404 (Ruwenzori and Muhavura). When calculated separately for the three DNA-content groups,
405 G3 contained highest gene diversity (D = 0.139 ± 0.067; H_e = 0.215 ± 0.159). The combined
406 G1-G2 group had even higher diversity (D = 0.172 ± 0.084; H_e = 0.269 ± 0.17; Table 1). In a
407 non-hierarchical AMOVA, 60% of the total AFLP variation was attributed to variation among
408 populations (Table 2). Hierarchical AMOVAs showed high variation between the two genetic

groups (33.9%, F_{st} = 0.509), among the three DNA-content groups (47.2%, F_{st} = 0.734) and among individual mountains (37.3%, F_{st} = 0.624).

Discussion

Double colonization followed by secondary contact, allopolyploidization and extinction

Our results imply that the genus *Anthoxanthum* has an intriguingly dynamic biogeographic history in eastern Africa and in sub-Saharan Africa as a whole, initiated by double Late Pliocene colonization from two different sources and followed by expansion of the two distinctly divergent lineages, one from the south and one from the north. Our extended analyses thus corroborate the double colonization hypothesis presented in our previous study (Pimentel *et al*. 2013). Furthermore, based on the extension of our sampling to cover the entire range in East Africa and combined phylogenetic, phylogeographic, and DNA content inference, we conclude that the two currently allopatric lineages of sub-Saharan *Anthoxanthum* once must have met and hybridized in East Africa to produce an allopolyploid during the climatic oscillations of the later parts of the Pleistocene.

Our nuclear phylogeny (Figs 3, 4) shows that the eastern African taxa originated after a single immigration of a Eurasian lineage in the Late Pliocene, in agreement with the morphology-based hypothesis of their close relationship to the *A. alpinum/odoratum* complex (Hedberg 1976). The eastern African nuclear sequences were recovered in three well-supported allopatric subclades (with unresolved relationships), one with the Ethiopian *A. aethiopicum*, one with eastern East African *A. nivale*, and one with western East African *A. nivale*. The subclades were inferred to have diverged during the Late Pliocene/Middle Pleistocene. The southern African taxa also formed a distinct clade in our nuclear phylogeny, consistent with a single colonization of southern Africa from Southeast Asia in the Late Pliocene (Pimentel *et al*. 2013).

In contrast, our plastid phylogeny (Figs 2, 4) recovered eastern *A. nivale* in a clade with the

southern African taxa, whereas western *A. nivale* was recovered in an eastern African/Eurasian clade in agreement with the nuclear tree. Coalescent simulations rejected lineage sorting stochasticity as explanation for the incongruence between the plastid and nuclear data (Fig. 5) when effective populations sizes were below 400000 (plastid data) and 250000 (nuclear data). Large effective population sizes are unlikely for most species in the phylogeny (narrow endemics with pronounced clonal reproduction); only the Eurasian *A. odoratum* and *A. alpinum* have large distribution areas. However, estimates for other widespread, outcrossing taxa have recovered values well below our numbers (Maureira-Butler *et al*. 2008). We therefore conclude that even though the southern and eastern African lineages are currently allopatric (Fig. 1a), they once met and hybridized. Based on plastid divergence we inferred this event to have taken place in the Middle/Late Pleistocene. Based on morphology (see below) and our AFLP analysis, however, the hybridization event seems not solely to have resulted in plastid capture. The eastern samples formed a highly distinct genetic group, separated from the western samples by 20.2% along the first PCoA axis and by 33.9% (F_{st} = 0.509) in an AMOVA analysis (Fig. 6, Table 2). As AFLPs are known to mainly represent genome-wide nuclear markers (Ridout & Donini 1999), this high level of divergence may reflect that also the nuclear genome of eastern *A. nivale* contains DNA from the southern African lineage. The placement of eastern *A. nivale* in the eastern African/Eurasian clade in our nuclear phylogeny (Fig. 3), which was inferred from ribosomal sequences, is thus most likely caused by elimination of southern African rDNA from the hybrid via concerted evolution.

Our hybridization hypothesis thus resolves the conflicting morphology-based opinions of whether *A. nivale* has northern affinities (Hedberg 1976; Phillips 1995) or southern affinities (Clayton 1970). It has both, and our inference that eastern *A. nivale* not only contains plastid

DNA but also nuclear DNA from the southern African lineage is strengthened by the observation that some specimens of *A. nivale* show morphological similarities with the southern African endemics, which currently only extend northwards to Malawi (Fig. 1a; Clayton 1970). As the eastern East African mountains are well explored, the southern African lineage seems to have gone extinct in this area after the hybridization event. We also found that whereas the western samples of *A. nivale* were low-ploid (DNA-content groups G1 and G2), the eastern samples were high-ploid (G3; Figs 1b; Supporting Information Figs. S1, S2). This finding suggests that the hybridization event between the northern and southern lineages of *Anthoxanthum* in sub-Saharan Africa involved allopolyploidy, which is common in this genus (Chumová *et al*. 2015). The putative allopolyploid nature of these eastern *A. nivale* populations, together with their high genetic differentiation in the AFLP study would merit them to be recognized as a new species following the evolutionary or the phylogenetic species concept (reviewed in Soltis *et al*. 2007). However Soltis *et al*. (2007) suggested that the definition of new taxa in polyploid complexes should be subjected to strict criteria including morphological differentiation, so we await a final taxonomic conclusion until an in-depth morphological analysis of *A. nivale* is available.

The Pleistocene hybridization event detected in this study is particularly intriguing in light of the deeper history of the two involved lineages. The southern African lineage belongs to section *Ataxia*, a mostly SE Asian, tropical-alpine group which itself originated by a Miocene hybridization event between the genera *Anthoxanthum* and *Hierochloë* (Pimentel *et al*. 2013). Our results therefore imply that the East African allopolyploid has a double hybrid background spanning millions of years and widely different geographic affinities. This hybridization event between cold-adapted SE Asian (through southern Africa) and Eurasian

lineages in East Africa constitutes one of the very few examples of a connection between two tropical-alpine habitats in the Old World (Gehrke & Linder 2009; Linder 2014).

The eastern African lineage: immigration and polyploid evolution

We were not able to directly verify ploidy levels corresponding to the three DNA-content groups (G1 and G2 in the western mountains and G3 in the eastern mountains; Figs 1b; Supporting Information Figs. S1, S2) observed in this study because our attempts to germinate seeds failed. A comparison of our dataset with the extensive chromosome counts published by Hedberg (1976) suggests however that our groups correspond to tetraploids ($2n = 4x = 20$), octoploids ($2n = 8x = 40$) and dodecaploids ($2n = 12x = 60$). This ploidy level assignment is largely consistent with the DNA content values observed, but without chromosome counts it must be regarded as tentative. Hedberg (1976) only found two ploidy levels, tetraploids both on the western and eastern mountains and dodecaploids on the eastern mountains. Because dodecaploids by far dominated her counts from the eastern mountains, our DNA-content group G3 clearly corresponds to dodecaploids. She detected a few tetraploids only on one eastern mountain (Mt Aberdare), where they co-occurred with dodecaploids, but low-ploids were not found in our limited sampling from this mountain. From the western mountains, she mainly examined plants from Mt Elgon and found only tetraploids. On this mountain we only observed putative octoploids (DNA-content group G2 + one transitional specimen between G1 and G2) based on quite extensive sampling. Octoploids were not detected by Hedberg (1976) who made all her counts in plants collected at 3550 m or below, whereas all the populations we sampled in Mt Elgon grew above 3800 m. A direct relationship between ploidy and altitude was observed by Hedberg (1976) in other mountain systems and sharp altitudinal limits between the ranges of different cytotypes are common in other *Anthoxanthum* polyploid complexes (Felber-Girard *et al*. 1996). From the westernmost

mountains (Ruwenzori and Muhavura), where we found both low-ploid DNA content groups (G1 and G2) based on extensive sampling, she only reported a single tetraploid count. We therefore conclude that the plants with the lowest DNA-content (G1) in our sampling most likely are tetraploids while the less common octoploids went undetected in her study.

The lineage colonizing eastern Africa from Eurasia in the Late Pliocene was thus probably tetraploid and closely related to the Eurasian diploid/tetraploid complex constituted by *A. alpinum* and *A. odoratum* (Figs 2, 3). We found tetraploids to be dominant in the westernmost mountains in East Africa (G1, Appendix 1). Because the block mountain Ruwenzori was in its final uplift stage at this time (Gehrke & Linder 2014) and harbors tetraploids as well as the highest levels of genetic (AFLP) diversity and rarity in East Africa (Table 1), it is possible that the ancestral lineage first arrived in this area. Here tetraploids and octoploids co-occur, suggesting that the octoploids may have formed *in situ* via autopolyploidy (Fig. 1b), similar to the well-documented instances of autopolyploidy in the *A. alpinum/odoratum* complex in Europe (Chumová *et al*. 2015). The western East African tetraploids and/or octoploids may later have spread eastwards to the eastern mountains as well as northwards to Ethiopia in the Late Pliocene/Early Pleistocene (Fig. 4). In the eastern East African mountains, they apparently came into contact with sweet vernal grasses belonging to the expanding southern African lineage, hybridized and formed allopolyploids, followed by local extinction of the parental southern lineage. We are not aware of any published chromosome counts for the southern African taxa, but according to J. Loureiro (pers. comm.), all but one of them are tetraploids, which is consistent with our hypothesis.

Sweet vernal grasses are today absent from the Arabian Peninsula and most of the Middle East, but they grow in coastal and mountainous areas of North Africa from Morocco to Tunisia (Maire, 1931; Tutin 1980). This might suggest that the Eurasian lineage migrated to

western East Africa via North Africa, not via the Arabian Peninsula as suggested for other afro-alpine species (e.g. Assefa *et al*. 2007; Popp *et al*. 2008). However, all specimens of North African sweet vernal grasses we have examined so far belong to a more distantly related Mediterranean lineage (the *A. aristatum/ovatum* lineage; Pimentel *et al*. 2013), so further consideration of this alternative must await clarification of the taxonomy and phylogenetic relationships of the North African *Anthoxanthum*.

Distribution of genetic diversity in A. nivale

We observed high differentiation and virtually no introgression in *A. nivale* across the Rift Valley (Figs. 6A, B, C), a well-known barrier to gene flow (e.g. Assefa *et al*. 2007; Masao *et al.* 2013). In our study, however, the relative effects of geography and ploidy in explaining the absence of gene flow are difficult to disentangle. The AMOVA analyses (Table 2) showed that DNA content explained the highest percentage of genetic variation (47.2%) followed by the division between the western and eastern mountains (33.9%). Introgression was detected within each region, but never across DNA-content groups (Figs. 6B and C). These results indicate a long history of isolation between ploidy levels as well as between the eastern and western groups of populations.

The overall genetic diversity in *A. nivale* (D = 0.155) is lower than expected for an outcrossing, perennial plant taxon (Nybom 2004), as also observed in many other afro-alpine species (Ehrich *et al*. 2007; Geleta & Bryngelsson 2009; Masao *et al.* 2013). This finding is consistent with the hypothesis that afro-alpine species may have experienced severe bottlenecks during cycles of colonization-extinction-recolonization, and that the current populations have established after recent long-distance colonization. Our finding of higher diversity (D = 0.172 ± 0.084 and H_e = 0.269 ± 0.171) and rarity (5.3) in the western than in

553 the eastern mountains (D = 0.139 ± 0.067, H_e = 0.215 ± 0.159; 2.9) supports the hypothesis of
554 longer persistence of the species in the west, which may have been the first area to be
555 colonized by the Eurasian lineage.

556 *Dynamic history of afro-alpine plant communities*

557 Our results add to a growing body of evidence showing that tropical high mountain habitats
558 have a dynamic recent history involving niche conservatism, recruitment from remote
559 sources, repeated dispersals, diversification, hybridization, and extinction (Hughes 2016;
560 Lagomarsino *et al*. 2016). The alpine zone of isolated tropical high mountains seems mainly
561 to have been colonized via long-distance dispersal of lineages that already were preadapted to
562 cold conditions in other areas, as recently shown for several groups of organisms on Mt
563 Kinabalu on Borneo (Merckx *et al*. 2015). This scenario also holds true for the sweet vernal
564 grasses in sub-Saharan Africa, which show an exceptionally dynamic history with expansion
565 of two independently immigrated lineages, secondary contact resulting in hybridization and
566 allopolyploidization, and local extinction of one parental lineage after the hybridization event.
567 The extinction event was possible to trace because the now locally extinct lineage left its
568 footprint in an allopolyploid derivative. Our results are thus consistent with a hypothesis of
569 cycles of local colonization, extinction, and recolonization during the Pleistocene climatic
570 oscillations as drivers shaping afro-alpine and afro-montane plant communities, as also
571 suggested by recent documentation of old species that show recent interpopulational
572 divergence (Mairal *et al*. 2015; Gizaw *et al*. 2016a). Our study also adds to the emerging
573 evidence suggesting that long-distance-dispersed frost-tolerant plant lineages colonized
574 eastern Africa successively over a long time period, some possibly before the formation of the
575 current high mountains (Late Miocene/Early Pliocene; Gizaw *et al*. 2016a), some possibly at
576 the time of their final uplift phase (Ruwenzori in the Late Pliocene; Pimentel *et al*. 2013 and

577 this study), and some long after their formation (Ethiopian mountains in the Pleistocene; Koch
578 *et al*. 2006).

579

580 **Acknowledgements**

581 This work was supported by the Norwegian Programme for Development, Research and
582 Higher Education (NUFU; project no. 2007/1058: AFROALP-II - Afro-alpine 'sky islands':
583 genetic versus taxonomic biodiversity, climate change, and conservation) to CB and SN.
584 Partial support was provided by the Academy of Sciences of the Czech Republic (Project
585 RVO 67985939) to PT. We thank the Uganda National Council for Science and Technology,
586 Uganda Wildlife Authority, Tanzanian Commission for Science and Technology, Tanzanian
587 National Parks Authority, and National Museums of Kenya for permission to conduct
588 fieldwork. We thank the staff at ETH, O, EA, MHU, SUA and NHT for curation of our
589 specimens.

590

591 **References**

592 Antonelli A (2009) Have giant lobelias evolved several times independently? Life form shifts
593 and historical biogeography of the cosmopolitan and highly diverse subfamily
594 Lobelioideae (Campanulaceae). *BMC Biology*, **7**, 82.

595 Antonelli A, Sanmartín I (2011) Why are there so many plant species in the Neotropics?.
596 *Taxon*, **60**, 403-414.

597 Assefa A, Ehrich D, Taberlet P, Nemomissa S, Brochmann C (2007) Pleistocene colonisation
598 of afro-alpine 'sky islands' by the arctic-alpine *Arabis alpina*. *Heredity* **99**: 133-142.

599 Baker BH, Mohr PA, Williams LAJ (1972) Geology of the eastern rift system of Africa.
600 *Geological Society of America Special Papers*, **136**, 1-68.

601 Bell CD, Donoghue MJ (2005) Phylogeny and biogeography of Valerianaceae (Dipsacales)
602 with special reference to the South American valerians. *Organisms, Diversity and*
603 *Evolution*, **5**, 147-159.

604 Blanco-Pastor JL,Vargas P, Pfeil BE (2012) Coalescent simulations reveal hybridization and
605 incomplete lineage sorting in Mediterranean *Linaria*. *PLOS One*, **7**, e39089.

606 Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004) How to
607 track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**,
608 3261-3273.

609 Chumová Z, Krejčíková J, Mandáková T, Suda J, Trávníček P (2015) Evolutionary and
610 taxonomic implications of variation in nuclear genome size: lesson from the grass genus
611 *Anthoxanthum* (Poaceae). *PLOS One*, **10**, e0133748.

612 Clayton WD (1970) Gramineae (Part 1). In: *Flora of Tropical East Africa* (eds Milne-
613 Redhead E, Polhill RM). Crown Agents for Overseas Governments and Administration,
614 London, UK.

615 Drummond AJ, Rambaut A (2007) Beast: Bayesian evolutionary analysis by sampling trees.
616 *BMC Evolutionary Biology*, **7**, 214.
617 Drummond AJ, Suchard A, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUTi
618 and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969-1973.
619 Drummond AJ, Bouckaert RR (2015) Bayesian evolutionary analysis with BEAST.
620 Cambridge University Press, Cambridge, UK.
621 Ebinger CJ, Yemane T, Harding DJ, Tesfaye S, Kelley S, Rex DC (2000) Rift deflection,
622 migration and propagation: Linkage of the Ethiopian and Eastern Rifts, Africa. *Geological*
623 *Society of America Bulletin*, **112**, 163-176.
624 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high
625 throughput. *Nucleic Acids Research*, **32**, 1792-1797.
626 Ehrich D (2006) AFLPdat: a collection of R functions for convenient handling of AFLP data.
627 *Molecular Ecology Notes*, **6**, 603-604.
628 Ehrich D, Gaudeul M, Assefa A, Koch MA, Mummenhoff K, Nemomissa S, Intrabiodiv
629 Consortium, Brochmann C (2007) Genetic consequences of Pleistocene range shifts:
630 contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*,
631 **16**, 2542-2559.
632 Evanno G, Regnaut S, Goudet J (2005) Detecting the numbers of clusters of individuals using
633 the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.
634 Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform
635 population genetics analyses under linux and windows. *Molecular Ecology Resources*, **10**,
636 564-567.

637 Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from
638 metric distances among DNA haplotypes: application to human mitochondrial DNA
639 restriction data. *Genetics*, **131**, 479-491.
640 Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus
641 genotype data: dominant markers and null allele. *Molecular Ecology Notes*, **7**, 574-578.
642 Felber-Girard M, Felber F, Buttler A (1996) Habitat differentiation in a narrow hybrid zone
643 between diploid and tetraploid *Anthoxanthum alpinum*. *New Phytologist*, **133**, 531-540.
644 Gaudeul M, Taberlet P, Till-Bottraud I (2000) Genetic diversity in an endangered alpine
645 plant, *Eryngium alpinum* L. (Apiaceae), inferred from amplified fragment length
646 polymorphism markers. *Molecular Ecology*, **9**, 1625-1637.
647 Gehrke B, Linder HP (2009) The scramble for Africa: pan-temperate elements on the African
648 high mountains. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**,
649 2657-2665.
650 Gehrke B, Linder HP (2014) Species richness, endemism and species composition in the
651 tropical Afroalpine flora. *Alpine Botany*, **124**, 165-177.
652 Gehrke B, Kandziora M, Pirie MD (2016) The evolution of dwarf shrubs in alpine
653 environments: a case study of *Alchemilla* in Africa. *Annals of Botany*, **117**, 121-131.
654 Geleta M, Bryngelsson T (2009) Inter simple sequence repeat (ISSR) based analysis of
655 genetic diversity of *Lobelia rhynchopetalum* (Campanulaceae). *Hereditas*, **146**, 122-130.
656 Gizaw A, Kebede M, Nemomissa S, Ehrich D, Tessema B, Mirré V, Popp M, Brochmann C
657 (2013) Phylogeography of the heathers *Erica arborea* and *E. trimera* in the afro-alpine
658 "sky-islands" inferred from AFLPs and plastid DNA sequences. *Flora*, **208**, 453-463.
659 Gizaw A, Brochmann C, Nemomissa S, Wondimu T, Masao CA, Mugizi TF, Abdi AA,
660 Oxelman B, Popp M, Dimitrov D (2016a) Colonization and diversification in the African

661 "sky-islands": insights from fossil-calibrated molecular dating of *Lychnis*
662 (Caryophyllaceae). *New Phytologist*, **211**, 719-734.
663 Gizaw A, Wondimu T, Mugizi TF, Masao CA, Abdi AA, Popp M, Ehrich D, Nemomissa S,
664 Brochmann C (2016b) Vicariance, dispersal and hybridisation in a naturally fragmented
665 system: the afroalpine endemics *Carex monostachya* and *C. runssoroensis* (Cyperaceae).
666 *Alpine Botany*, **126**, 59-71.
667 Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user
668 interface for sequence alignment and phylogenetic tree building. *Molecular Biology and*
669 *Evolution*, **27**, 221–224.
670 Hedberg I (1976) A cytotaxonomic reconnaissance of tropical African *Anthoxanthum* L.
671 (Graminaceae). *Botaniska Notiser*, **129**, 85-90.
672 Hedberg O (1957) Afroalpine vascular plants. A taxonomic revision. *Symbolae Botanicae*
673 *Upsalienses*, **15**, 1-411.
674 Hedberg O (1961) The phytogeographic position of the afroalpine flora. *Recent Advances in*
675 *Botany*, **1**, 914-919.
676 Hedberg O (1969) Evolution and speciation in a tropical high mountain flora. *Biological*
677 *Journal of the Linnean Society*, **1**, 135-148.
678 Hedberg O (1970) Evolution of the afroalpine flora. *Biotropica*, **2**, 16-23.
679 Hedberg O (1992) Afroalpine vegetation compared to páramo: convergent adaptations and
680 divergent differentiation. In: *Páramo: an Andean ecosystem under human influence* (eds
681 Balslev H, Luteyn JL), pp. 15-29. Academic Press, London, UK.
682 Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data.
683 *Molecular Biology and Evolution*, **27**, 570–580.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754-755.

Hughes CE, Atchison GW (2015) The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist*, **207**, 275-282.

Hughes CE (2016) The tropical Andean plant diversity powerhouse. *New Phytologist*, **210**, 1152-1154.

Huson D, Scornavacca C (2012) Dendroscope 3 : an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, **61**, 1061-1067.

Jabbour F, Renner SS (2012) A phylogeny of Delphinieae (Ranunculaceae) shows that *Aconitum* is nested within *Delphinium* and that Late Miocene transitions to long life cycles in the Himalayas and southwest China coincide with bursts in diversification. *Molecular Phylogenetics and Evolution*, **62**, 928-942.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801-1806.

Koch MA, Kiefer C, Ehrich 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.

Kosman E (2003) Nei's gene diversity and the index of average differences are identical measures of diversity within populations. *Plant Pathology*, **52**, 533-535.

Lagomarsino LP, Condamine FL, Antonelli A, Mulch A, Davis CC (2016) The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist*, **210**, 1430-1442.

Linder HP (2014) The evolution of African plant diversity. *Frontiers in Ecology and Evolution*, **2**, 38.

Maddison WP, Maddison DR (2009) *Mesquite: a modular system for evolutionary analysis*. URL http://mesquiteproject.org. [accessed 15 October 2015].

Mairal M, Pokorny L, Aldasoro JJ, Alarcón M, Sanmartín I (2015) Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). *Molecular Ecology*, **24**, 1335-1354.

Maire R (1931) *Anthoxanthum* L. In: *Flore de l'Afrique du Nord* (eds. Maire R, Guinochet M, Fanel L), pp. 25-29. P Lechevalier, Paris, France.

Masao CA, Gizaw A, Piñeiro R, Tussiime FM, Wondimu T, Abdi AA, Popp M, Gussarova G, Lye KA, Munishi P *et al.* (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.

Maureira-Butler IJ, Pfeil BE, Muangprom A, Osborn TC, Doyle JJ (2008) The reticulate history of *Medicago* (Fabaceae). *Systematic Biology*, **57**, 466-482.

Merckx VSFT, Hendriks KP, Beentjes KK, Mennes CB, Becking LE, Peijnenburg KTCA, Afendy A, Arumugam N, de Boer H, Biun A *et al*. (2015) Evolution of endemism in a young tropical mountain. *Nature*, **524**, 347-350.

Mugrabi de Kupler AL, Fagundez J, Bellstedt DU, Oliver EG, Leon J, Pirie MD (2015) Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (Ericeae, Ericaceae). *Molecular Phylogenetics and Evolution*, **88**, 121-131.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University, New York, USA.

730 Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of
731 restriction endonucleases. *Proceedings of the National Academy of Sciences of the United*
732 *States of America*, **76**, 5269-5273.
733 Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P,
734 Gladstone J, Goyal R *et al*. (2005) The pattern of polymorphism in *Arabidopsis thaliana*.
735 *PLoS Biology*, e196.
736 Nürk NM, Uribe-Convers S, Gehrke B, Tank DC, Blattner FR (2015) Oligocene niche shift,
737 Miocene diversification – cold tolerance and accelerated speciation rates in the St. John's
738 Worts (*Hypericum*, Hypericaceae). *BMC Evolutionary Biology*, **15**, 80.
739 Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific
740 genetic diversity in plants. *Molecular Ecology*, **13**, 1143-1155.
741 Nylander JAA. 2004. *MrModeltest v2*. URL https://github.com/nylander/MrModeltest2.
742 [accessed 15 September 2012].
743 Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (are we there
744 yet?): a system for graphical exploration of MCMC convergence in Bayesian
745 phylogenetics. *Bioinformatics*, **24**, 581–583.
746 Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R
747 language. *Bioinformatics*, **20,** 289–290.
748 Penny D, Hendy MD (1985) The use of tree comparison metrics. *Systematic Zoology*, **34**, 75-
749 82.
750 Phillips S (1995) Poaceae (Gramineae). In: *Flora of Ethiopia and Eritrea* (eds Hedberg I,
751 Edwards S). The National Herbarium, Biology Department, Science Faculty Addis Ababa
752 University, Addis Ababa, Ethiopia.

753 Pimentel M, Sahuquillo E, Torrecilla Z, Popp M, Catalán P, Brochmann C (2013)
754 Hybridization and long-distance colonization at different time scales: towards resolution
755 of long-term controversies in the sweet vernal grasses (*Anthoxanthum*). *Annals of Botany*,
756 **112**, 1015-1030.
757 Pirie MD, Humphreys AM, Barker NP, Linder HP (2009) Reticulation, data combination, and
758 inferring evolutionary history: an example from Danthonioideae (Poaceae). *Systematic*
759 *Biology*, **58**, 612–628.
760 Popp M, Gizaw A, Nemomissa S, Suda J, Brochmann C (2008) Colonization and
761 diversification in the African 'sky islands' by Eurasian *Lychnis* L. (Caryophyllaceae).
762 *Journal of Biogeography*, **35**, 1016-1029.
763 Popp M, Mirré V, Brochmann C (2011) A single mid-Pleistocene long distance dispersal by a
764 bird can explain the extreme bipolar disjunction in crowberries (*Empetrum*). *Proceedings*
765 *of the National Academy of Sciences of the United States of America*, **108**, 6520–6525.
766 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
767 multilocus genotype data. *Genetics*, **155**, 945-959.
768 Rambaut A, Drummond A (2007) Tracer v. 1.4. University of Edinburgh, Edinburgh, UK.
769 Ridout CJ, Donini P (1999) Use of AFLP in cereals research. *Trends in Plant Science*, **4**, 76-
770 79.
771 Rohlf FJ (2002) *NTSYSpc, Numerical Taxonomy and Multivariate Analysis System. Version*
772 *2.11 a, User guide*. [WWW document] URL
773 www.exetersoftware.com/downloads/ntsysguide21.pdf. [accessed 24 May 2014].
774 Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population
775 structure. *Molecular Ecology Notes*, **4**, 137-138.

776 Schönswetter P, Lachmayer M, Lettner C, Prehsler D, Rechnitzer S, Reich DS, Sonnleitner
777 M, Wagner I, Huelber K, Schneeweiss GM *et al*. (2007a) Sympatric diploid and hexaploid
778 cytotypes of *Senecio carniolicus* (Asteraceae) in the Eastern Alps are separated along an
779 altitudinal gradient. *Journal of Plant Research*, **120**, 721-725.
780 Schönswetter P, Suda J, Popp M, Weiss-Schneeweiss H, Brochmann C (2007b) Circumpolar
781 phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA
782 sequences, nuclear DNA content and chromosome numbers. *Molecular Phylogenetics and*
783 *Evolution*, **42**, 92-103.
784 Schönswetter P, Tribsch A (2005) Vicariance and dispersal in the alpine perennial *Bupleurum*
785 *stellatum* L. (Apiaceae). *Taxon*, **54**, 725-732.
786 Sikolia S, Onyango JC, Beck E, Kinyamario JI (2009) The distribution of C_3 and C_4
787 photosynthetic species of the Centrospermeae along an altitudinal gradient in Western
788 Kenya. *International Journal of Botany*, **5**, 47-57.
789 Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic
790 analyses. *Systematic Biology*, **49**, 329-381.
791 Sklenář P, Hedberg I, Cleef AM (2014) Island biogeography of tropical alpine floras. *Journal*
792 *of Biogeography*, **41**, 287-297.
793 Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS
794 (2007) Autopolyploidy in angiosperms: have we grossly underestimated the number of
795 species?. *Taxon*, **56**, 13-30.
796 Suda J, Trávníček P (2006) *Estimation of relative nuclear DNA content in dehydrated plant*
797 *tissues by flow cytometry*. John Wiley & Sons, New York, USA.
798 Swofford DL (2002) *PAUP* Version 4. Phylogenetic analysis using parsimony (*and Other*
799 *Methods)*. Sinauer Associates, Sunderland, Massachusetts.

800 Tutin TG (1980) *Anthoxanthum* L. In: *Flora Europaea, Vol. 5* (eds. Tutin TG, Heywood VH,
801 Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA), pp. 229-230. Cambridge
802 University Press, Cambridge, UK.
803 Van-De-Peer Y, De-Wachter Y (1994) TREECON for windows: a software package for the
804 construction and drawing of evolutionary trees for the Microsoft Windows environment.
805 *Computer Applications in the Biosciences*, **10**, 569-70.
806 Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA (2008) The age of the
807 grasses and clusters of origins of C_4 photosynthesis. *Global Change Biology*, **14**, 2963–
808 2977.
809 Von Hagen KB, Kadereit JW (2001) The phylogeny of *Gentianella* (Gentianaceae) and its
810 colonization of the southern hemisphere as revealed by nuclear and chloroplast DNA
811 sequence variation. *Organisms, Diversity and Evolution*, **1**, 61-79.
812 Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population
813 structure. *Evolution*, **38**, 1358-1370.
814 Wondimu T, Gizaw A, Tusiime FM, Masao CA, Abdi AA, Gusarova G, Popp M, Nemomissa
815 S, Brochmann C (2014) Crossing barriers in an extremely fragmented system: two case
816 studies in the afro-alpine sky island flora. *Plant Systematics and Evolution*, **300**, 415-430.
817

818 **Data Accessibility**

819 DNA sequences have been deposited in GenBank (for accession numbers see Appendices 1, 2
820 and 3). Sequence alignments (concatenated sequences) have been deposited in FigShare
821 entries DOI: XXXXX (nuclear DNA) and DOI: XXXXXX (plastid DNA)

822 The AFLP dataset has been deposited in FigShare entry DOI: XXXXXX
823

Author Contributions Box

CB and MPi conceived the study. CB, SN, GE and VM obtained funding. FMT, AG, TW, CAM, AAA, VM, SN, MPo, GE and CB collected samples. FMT, AG, PT and MPi performed most lab and computer analyses. FMT, AG, MPo, CB and MPi drafted the manuscript. All authors contributed to the writing, read, and approved the final manuscript.

842 Table 1. Genetic diversity and rarity in the East African *A. nivale* based on 424 AFLP markers
843 scored in 125 plants (35 populations), calculated separately for each mountain and for each of
844 the two genetic groups inferred in the STRUCTURE analyses. *D*, Nei's average gene
845 diversity [Nei, 1987, estimated as the average proportion of pairwise differences among
846 genotypes (Kosman, 2003)]; DW, frequency-down-weighted marker values (Schönswetter &
847 Tribsch, 2005) as a measure of genetic distinctivity or rarity; He, Nei's unbiased expected
848 heterozygosity (Nei, 1987; Gaudeul *et al*., 2000).

	D ± sd	DW	H_e± sd
Mt Aberdare	0.095 ± 0.053	2.38	0.398 ± 0.121
Mt Kenya	0.102 ± 0.051	2.49	0.257 ± 0.149
Mt Kilimanjaro	0.119 ± 0.059	3.34	0.282 ± 0.155
Mt Meru	0.118 ± 0.060	3.01	0.319 ± 0.143
Mt Elgon	0.085 ± 0.043	3.73	0.297 ± 0.144
Mt Muhavura	0.172 ± 0.088	5.63	0.358 ± 0.126
Mt Ruwenzori	0.200 ± 0.131	9.63	0.555 ± 0.079
Western Mountain group	0.172 ± 0.084	126.1	0.269± 0.171
Eastern Mountain group	0.139 ± 0.067	125.8	0.215 ± 0.159

849

850

851

852

853 Table 2. Analyses of molecular variance (AMOVA) and F_{st} values based on 424 AFLP
854 markers scored in 125 plants (35 populations) of the East African *A. nivale*. All *P*-values were
855 <0.001.

	Source of variation	**d.f**	**% of variation**	**F statistics**		
				F_{st}	F_{sc}	F_{ct}
All populations	Among populations	34	60.1	0.60		
	Within populations	87	39.9			
Three DNA-content groups (G1-G3)	Among groups	2	47.2	0.73	0.50	0.47
	Among populations within groups	35	26.2			
	Within populations	84	26.6			
Two genetic groups (EM vs WM)	Between groups	1	33.9	0.68	0.51	0.34
	Among populations within groups	33	33.7			
	Within populations	87	32.4			

856

Appendix 1: Collection data and genetic diversity of 125 *A. nivale* plants in 35 populations analysed with 424 AFLP markers; n = number of individual plants, D = Nei's average gene diversity, DW = genetic rarity (calculated as measure of frequency down-weighed marker value), He ± s.d. = expected heterozygosity ± standard deviation, PL = number of polymorphic loci, %PL = percentage polymorphism, Alt = Altitude, lat = latitude, long = longitude. * indicates populations where the flow cytometry analyses failed for one or more individuals, and # indicates populations containing individuals with transitional cytotypes (G1+, transition between G1 and G2; G2+, transition between G2 and G3 and G3+, higher DNA-content than G3). The number in parentheses under 'cytotype' correspond to the order code given to the sample during flow cytometry analyses and correspond to the number on the horizontal axis of Fig. S1.

No.	Database No.	Population	Country	Locality	Altitude (m)	Lat	Long	*n*	D ± s.d.	DW	He ± s.d.	PL	% PL	Cytotype (FCM order code)
1	O-DP-34849, 34851	KN0031	Kenya	Mt Elgon: S of Mt Koitobos	3920	1.1057	34.6018	3	0.042 ± 0.032	2,59	0.667 ± 0.000	28	6,29	G1+ (10), G2 (11, 15)
2	O-DP-34896 -34900	KN0044	Kenya	Mt Elgon: Mt Koitobos	3860	1.1025	34.6058	5	0.057 ± 0.035	3,51	0.476 ± 0.098	53	11,90	G2 (19, 24-26, 28)
3	O-DP-35109	KN0101	Kenya	Mt Elgon: Mt Koitobos	3950	1.1240	34.5903	1	_	_	_	_	_	G2 (16)
4	O-DP-35454, 35455	KN0176	Kenya	Mt Elgon: Caldera	4030	1.1185	34.5857	2	0.038 ± 0.039	3	1.000 ± 0.000	17	3,82	G2 (17, 20)
5	O-DP-35545, 35547, 35548	KN0202	Kenya	Mt Elgon: Caldera	4040	1.1180	34.5867	3	0.042 ± 0.032	4,9	0.667 ± 0.000	28	6,29	G2 (12, 22, 30)
6	O-DP-35805, 35806, 35808	KN0259	Kenya	Mt Elgon: E of Mt Koitobos	3860	1.1083	34.6061	3	0.198 ± 0.148	7,68	1.000 ± 0.000	39	8,76	G2 (13, 14, 23)
7	O-DP-35871 -35873	KN0272*	Kenya	Mt Elgon: NE of Mt Koitobos	3800	1.1029	34.6131	3	0.045 ± 0.034	2,39	0.667 ± 0.000	30	6,74	G2 (31, 33)
8	O-DP-27894 - 27898	KN0583*	Kenya	Aberdare Mts: at the end of the car road towards Satima	3620	-0.3350	36.6510	5	0.006 ± 0.037	1,58	0.463 ± 0.094	57	12,80	G3 (50, 58, 82, 96)
9	O-DP-28220, 28222, 28223	KN0662*	Kenya	Aberdare Mts: at the end of the car road towards Satima	3660	-0.3372	36.6503	3	0.084 ± 0.064	3,83	0.667 ± 0.000	56	12,50	G3 (97, 100)
10	O-DP-28629, 28630	KN0795	Kenya	Mt Kenya: Near Shipton's Camp	4340	-0.1395	37.3092	2	0.072 ± 0.073	3,26	1.000 ± 0.000	32	7,19	G3 (46, 85)

11	O-DP-28719, - 28721	KN0829*	Kenya	Mt Kenya: Near Shipton's Camp	4270	-0.1416	37.3139	3	0.039 ± 0.030	5,14	0.667 ± 0.000	26	5,84	G3 (41, 92)
12	O-DP-28802 - 28805	KN0851	Kenya	Mt Kenya: N of Shipton's Camp	4230	-0.1392	37.3143	4	0.091 ± 0.061	3,96	0.535 ± 0.068	76	17,08	G3 (42, 45, 53, 63)
13	O-DP-28917 -28921	KN0876	Kenya	Mt Kenya: NE of Batian Peak	4050	-0.1214	37.2956	5	0.053 ± 0.033	2,28	0.463 ± 0.094	51	11,46	G3 (71, 75, 77, 86, 99)
14	O-DP-29079 - 29081	KN0912	Kenya	Mt Kenya: NE of Batian Peak	4040	-0.1214	37.2956	3	0.037 ± 0.029	1,33	0.667 ± 0.000	25	5,62	G3 (48, 101, 104)
15	O-DP-29159 - 29163	KN0935	Kenya	Mt Kenya: Shipton's Cave	4190	-0.1336	37.2765	5	0.005 ± 0.031	2,24	0.487 ± 0.100	46	10,30	G3 (81, 84, 94)
16	O-DP-36427 - 36431	KN0980	Kenya	Mt Kenya: SE of Point Lenana	4390	-0.1486	37.3321	5	0.052 ± 0.032	1,62	0.097 ± 0.127	49	11,01	G3 (72, 80, 95, 107, 109)
17	O-DP-36702 - 36706	KN1052	Kenya	Mt Kenya	4020	-0.1461	37.3480	5	0.058 ± 0.036	1,97	0.099	53	11,91	G3 (70, 91, 105)
18	O-DP-36865	KN1109#	Kenya	Mt Kenya: Teleki Valley	4120	-0.1693	37.2753	1	_	_	_	_	_	G3 (108)
19	O-DP-37044, 37045	TZ0031	Tanzania	Mt Kilimanjaro: Shira Plateau	3410	-2.9866	37.2224	2	0.031 ± 0.033	1,99	1.000 ± 0.000	14	3,15	G3 (102, 103)
20	O-DP-37271 - 37275	TZ0092	Tanzania	Mt Kilimanjaro: Shira Plateau	3970	-3.0523	37.2752	5	0.106 ± 0.065	4,72	0.470 ± 0.0959	100	22.47	G3 (47, 52, 59, 65, 68)
21	O-DP-42690 - 42692, 37301, 37302	TZ0106*	Tanzania	Mt Kilimanjaro: Shira Plateau	3900	-3.0628	37.2782	5	0.091 ± 0.056	2,64	0.495 ± 0.101	82	18.43	G3 (62, 74, 83, 89)
22	O-DP-37451 -37455	TZ0136*	Tanzania	Mt Kilimanjaro: Barranco	4160	-3.0862	37.3234	5	0.11 ± 0.065	3,94	0.460 ± 0.092	103	23.15	G3 (60, 79, 93)
23	O-DP-38087 - 38090	TZ0278#	Tanzania	Mt Kilimanjaro: Mawenzi	3820	-3.1467	37.4420	4	0.092 ± 0.061	3,92	0.547 ± 0.075	75	16.85	G3 (55, 90, 98); G3+ (106)
24	O-DP-38121 - 38124	TZ0291#	Tanzania	Mt Kilimanjaro: Horombo	3820	-3.1350	37.4337	4	0.052 ± 0.035	1,87	0.535 ± 0.069	43	9,66	G2+ (38), G3 (69, 88), G3+ (110)
25	O-DP-38496 - 38500	TZ0380*	Tanzania	Mt Meru: Saddle Hut area	3600	-3.2170	36.7690	5	0.058 ± 0.036	3,27	0.474 ± 0.097	54	12,13	G3 (39,76,78)
26	O-DP-38609 - 38613	TZ0404*	Tanzania	Mt Meru: Saddle Hut area	3640	-3.2180	36.7668	5	0.038 ± 0.024	2,65	0.467 ± 0.096	36	8,09	G3 (40, 43, 56)
27	O-DP-38644 - 38648	TZ0412*	Tanzania	Mt Meru: Saddle Hut area	3640	-3.2180	36.7668	5	0.097 ± 0.059	2,99	0.500 ± 0.101	86	19,30	G3 (44, 57, 64, 87)
28	O-DP-38748 - 38751	TZ0430	Tanzania	Mt Meru	3740	-3.2186	36.7596	4	0.105 ± 0.069	3,43	0.569 ± 0.083	82	18.43	G3 (49, 51, 54, 66)
29	O-DP-42984 - 42987	UG2043#	Uganda	Virunga Mts: Mt Muhavura, along trail to summit	3550	-1.3763	29.6715	4	0.189 ± 0.124	7,35	0.532 ± 0.066	158	35.51	G1 (6), G2 (27), G2+ (37)
30	O-DP-42965 - 42968	UG2068*	Uganda	Virunga Mts: Mt Muhavura, summit	4140	-1.3827	29.6780	4	0.061 ± 0.041	5,27	0.551 ± 0.078	49	11,01	G1 (3)
31	O-DP-53146	UG2082	Uganda	Virunga Mts: Mt Muhavura, betw. 2nd Hut	4000	-1.3827	29.6780	1	_	_	_	_	_	G2 (21)

				and summit										
32	O-DP-43615, 43617 - 43619	UG2117	Uganda	Virunga Mts: Mt Muhavura, betw. 2nd Hut and summit	4020	-1.3820	29.6767	4	0.161 ± 0.106	5,7	0.523 ± 0.058	137	30.79	G1 (1, 2,7); G2 (29)
33	O-DP-40316, 40318, 40319	UG2206	Uganda	Virunga Mts: Mt Muhavura, betw. 1st Hut and 2nd Hut	3530	-1.3758	29.6710	3	0.055 ± 0.042	4,94	0.667 ± 0.000	37	8,31	G1 (4), G1+ (9)
34	O-DP-43086 - 43088	UG2265*	Uganda	Rwenzori Mts: Lower Bigo Valley	3430	0.3850	29.9273	3	0.190 ± 0.143	11	0.667 ± 0.000	127	28.54	G1 (5), G1+ (8), G2 (34)
35	O-DP-40696	UG2303	Uganda	Rwenzori Mts: Upper Bigo Valley	3570	0.3852	29.9137	1	_	_	_	_	_	G2 (36)

877 **Appendix 2**: *Anthoxanthum nivale* specimens used in the phylogenetic analyses. DNA-C.G., DNA content group. For population details see

878 Appendix 1.

Specimen id	Population id	Mountain	DNA-C.G.	ETS	ITS	*trn*L-F	*trn*T-L	Bayesian analyses	*BEAST
ODP43087	UG2265	Mt Ruwenzori	G1	KX650653	KX650749	KX650706	KX650693	+	+
SANT65955/1	-	Mt Ruwenzori	-	KC897974	KC897913	KC897725	KC897848	+	+
SANT65955/2	-	Mt Ruwenzori	-	KC897975	KC897914	KC897726	KC897849	+	+
ODP34897	KN0044	Mt Elgon	G2	KX650654	KX650750	KX650708	KX650699	+	+
ODP35872	KN0272	Mt Elgon	G2	KX650655	KX650745	KX650711	KX650694	+	+
ODP35547	KN0202	Mt Elgon	G2	KX650656	KX650746	KX650709	KX650695	+	+
ODP35454	KN0176	Mt Elgon	G2	KX650657	KX650747	KX650707	KX650696	+	+
ODP53146	UG2082	Mt Muhavura	G2	KX650659	KX650731	KX650705	KX650692	+	+
ODP34900	KN0044	Mt Elgon	G2	KX650661	KX650748	-	KX650697	+	-
ODP35805	KN0259	Mt Elgon	G2	KX650658	-	-	-	+	-
ODP34899	KN0044	Mt Elgon	G2	KX650660	KX650751	KX650710	KX650698	+	+
ODP38613	TZ0404	Mt Meru	G3	KX650662	KX650739	KX650714	-	+	-
ODP38611	TZ0404	Mt Meru	G3	KX650663	KX650740	KX650713	-	+	-
ODP37272	TZ0092	Mt Kilimanjaro	G3	KX650664	KX650741	KX650722	-	+	-

ODP28803	KN0851	Mt Kenya	G3	KX650665	KX650736	KX650716	KX650688	+	+
ODP28223	KN0662	Mt Aberdare	G3	KX650666	-	KX650723	KX650689	+	-
ODP36428	KN0980	Mt Kenya	G3	KX650668	KX650735	-	KX650685	+	-
ODP28804	KN0851	Mt Kenya	G3	KX650672	KX650734	KX650720	KX650686	+	+
ODP29159	KN0935	Mt Kenya	G3	KX650674	KX650732	KX650715	KX650682	+	+
ODP27894	KN0583	Mt Aberdare	G3	KX650673	KX650737	KX650721	KX650691	+	+
ODP27895	KN0583	Mt Aberdare	G3	-	-	KX650718	KX650690	+	-
ODP28802	KN0851	Mt Kenya	G3	KX650675	KX650738	KX650717	KX650687	+	+
ODP38122	TZ0291	Mt Kilimanjaro	G3	KX650669	KX650733	KX650719	KX650681	+	+
ODP38644	TZ0412	Mt Meru	G3	KX650667	KX650743	KX650712	-	+	-
ODP37273	TZ0092	Mt Kilimanjaro	G3	KX650670	KX650742	KX650724	KX650684	+	+
ODP37453	TZ0136	Mt Kilimanjaro	G3	KX650671	KX650744	KX650725	KX650683	+	+

885 **Appendix 3**: Specimens of other taxa used in the phylogenetic analyses (for *A. nivale*, see Appendix 2). The sequences of *A. aethiopicum* and

886 *Hierochloë altissima* were produced in this study. The sequences of *A. dregeanum*, *A. ecklonii*, *A. madagascariense*, *A. gracile*, *A. odoratum* and

887 *A. alpinum* were obtained from Pimentel *et al*. (2013), and the remaining sequences were retrieved from GenBank (different populations).

Taxon	Specimen id	Locality	ETS	ITS	*trn*L-F	*trn*T-L	Bayesian analyses	*BEAST
A. aethiopicum	UPS 234211	Bale Region, Ethiopia. 2400 m. 02/11/1986	KX650676	KX650752	KX650726	KX650700	+	+
A. aethiopicum	UPS 234213	Mt Boruluccu, Ethiopia. 3700 m. 04/12/1965	KX650677	KX650753	KX650727	KX650701	+	+
A. aethiopicum	US3289331	Bale Region, Ethiopia. 2750 m. 29/10/1984	KX650679	KX650755	KX650729	KX650703	+	+
A. dregeanum	SANT65938	Swartberg Pass, Western Cape, South Africa	KC897961	KC897900	KC897712	KC897836	+	+
A. dregeanum	SANT65939	Jonkershoek, Western Cape, South Africa	KC897962	KC897901	KC897713	KC897837	+	+
A. ecklonii	SANT65940	Stutterheim, Eastern Cape, South Africa	KC897963	KC897902	KC897714	KC897838	+	+
A. ecklonii	SANT65943	Dohne Swamp, Amatole, Eastern Cape, South Africa	KC897964	KC897903	KC897715	KC897839	+	+
A. madagascariense	SANT65953	Tambunana, Tsiafajavjona, Madagascar	KC987972	KC897911	KC897723	KC897846	+	+

A. gracile	SANT65965	Lago Corsi, Iglesias, Sardinia, Italy	KC897967	KC897906	KC897718	KC897842	+	+
A. odoratum	SANT65957	Little Sugar Loaf, Wicklow, Ireland	KC897980	KC897921	KC897733	KC897856	+	+
A. odoratum	SANT65959	Carrigoona, Wicklow, Ireland	KC897977	KC897916	KC897728	KC897851	+	+
A. odoratum	SANT65958	Carrickgollogan, Dublin, Ireland	KC897976	KC897915	KC897727	KC897850	+	+
A. odoratum	SANT53424	Marei, O Corgo, Galicia, Spain	KC897983	KC897924	KC897736	KC897859	+	+
A. odoratum	SANT52208	Jehlanka chalet, Rokytnice, Czech Republic	KC897992-KC897994	KC897933-KC897935	KC897745-KC897747	KC897868-KC987870	+	+
A. odoratum	SANT53394	Norrbotten, Sweden	-	-	KC897696	KC897945	+	-
A. alpinum	SANT52191	Brévent, Chamonix, France	KC897944	KC897883	KC897695	KC897819	+	-
Hierochloë altissima	SANT72671	Caleta el Molinar, Niebla, Valdivia, Chile	KX650678	KX650754	KX650730	KX650704	+	+
Helictotrichon sp.	GenBank	GenBank	GQ324269	DQ336820.1	DQ336840.1	DQ336865.1	-	+
Festuca pratensis	GenBank	GenBank	JN187582	JN187604	JN187652	JQ973011.1	-	+
Lolium perenne	GenBank	GenBank	JN187583	JN187605	JN187653	KC897880	-	+
Puccinellia sp.	GenBank	GenBank	GQ283196.1	AF532934.1	AF533024.1	DQ336859	-	+
Poa sp.	GenBank	GenBank	GQ324369.1	JF904811.1	JN030969.1	JN030969.1	-	+

Dasypyrum villosum	GenBank	GenBank	AJ315031.1	JQ972933.1	JQ972965.1	-	-	+
Secale cereale	GenBank	GenBank	AJ315034.1	AF803400.1	EU013658.1	DQ336856.1	-	+
Melica sp.	GenBank	GenBank	KC897882	JN187651	JN187602	JN187580	-	+
Glyceria declinata	GenBank	GenBank	JN187851	JN187602	JN187651	KC897881	-	+

888

889 **Figure legends**

890 **Fig. 1.** Distribution area of Sub-Saharan species and map of sampled populations. (a) Total
891 distribution of *Anthoxanthum* in sub-Saharan Africa, representing five largely allopatric
892 endemic species. Dots represent sampling sites for material included in the study. (b)
893 Sampling sites covering the entire geographic distribution of the East African endemic *A.*
894 *nivale*, showing the three DNA-content groups (G1: 1.6; G2: 2.89; G3: 4.78) identified using
895 flow cytometry and the two genetic groups (grey and white) inferred from STRUCTURE
896 analyses of 424 AFLP markers scored in 125 plants (35 populations). The stippled lines
897 represent the Great Rift Valley system.

898 **Fig. 2.** Majority rule consensus tree inferred from Bayesian analysis (MrBayes) of plastid
899 DNA sequences (*trnT-L* and *trnL-F*). Forty-four samples representing eight *Anthoxanthum*
900 taxa and one outgroup (*Hierochloë altissima*) are represented in the tree. The symbol //
901 denotes branches that were shortened to simplify presentation. Dashed lines indicate branches
902 with PP<0.8. For each terminal, the species name is followed by accession number
903 (Appendix 1), DNA-content group (G1-G3; for *A. nivale* only) or ploidy level (only known
904 for Eurasian taxa), and geographic origin (for *A. nivale*: EM - Eastern Mountains, WM -
905 Western Mountains).

906 **Fig. 3.** Majority rule consensus tree inferred from Bayesian analysis (MrBayes) of nuclear
907 ribosomal DNA sequences (ITS and ETS). Forty-seven samples representing eight
908 *Anthoxanthum* taxa and one outgroup (*Hierochloë altissima*) are represented in the tree. The
909 symbol // denotes branches that were shortened to simplify presentation. Dashed lines
910 indicate branches with PP<0.8. For each terminal, the species name is followed by accession
911 number (Appendix 1), DNA-content group (G1-G3; for *A. nivale* only) or ploidy level (only

known for Eurasian taxa), and geographic origin (for *A. nivale*: EM - Eastern Mountains, WM - Western Mountains).

Fig. 4. Multilabelled maximum clade credibility species tree obtained in the *BEAST analysis. In *A. nivale*, the Eastern Mountains (EM) group is inferred to be hybridogenous (represented twice in the tree). Dashed lines indicate branches with PP<0.8. Bars show the 95% confidence interval for the age of the divergence. A, B, and C are nodes discussed in detail in the text.

Fig. 5. Distribution of distances between each gene tree and each of 1000 simulated gene trees from coalescence simulations. (a) distances of simulated gene trees from the plastid gene tree. Black bars, Ne (effective population size)=10000; dark gray bars, Ne=250000; light gray bars, Ne=1000000. (b) distances of simulated gene trees from the nuclear gene tree. Black bars, Ne=10000; dark gray bars, Ne=4000000; light gray bars, Ne=1000000. The arrow indicates the distance between the original plastid and nuclear gene trees.

Fig. 6. Genetic diversity and structuring in the East African *A. nivale* based on 424 AFLP markers (126 plants, 35 populations). The three DNA-content groups are indicated as G1-G3, the two genetic groups as grey and white, and the different mountains as symbols. (a) Genetic groups inferred from STRUCTURE analyses (K = 2). (b) Principal Coordinates Analysis (PCoA) based on Dice's coefficient of similarity among the 125 AFLP genotypes. (c) Neighbour-Joining dendrogram computed from pairwise F_{st} values as measures of distance among the 35 populations. Branch support was estimated using 1000 bootstrap replicates (only values >50% shown for major branches).

938 **Short legends for SI figures**

939 **Fig. S1** DNA-content in 125 plants of *Anthoxanthum nivale* inferred from flow cytometry
940 analysis.

941 **Fig. S2** Boxplots showing variation in the three DNA-content groups detected in *A. nivale*.

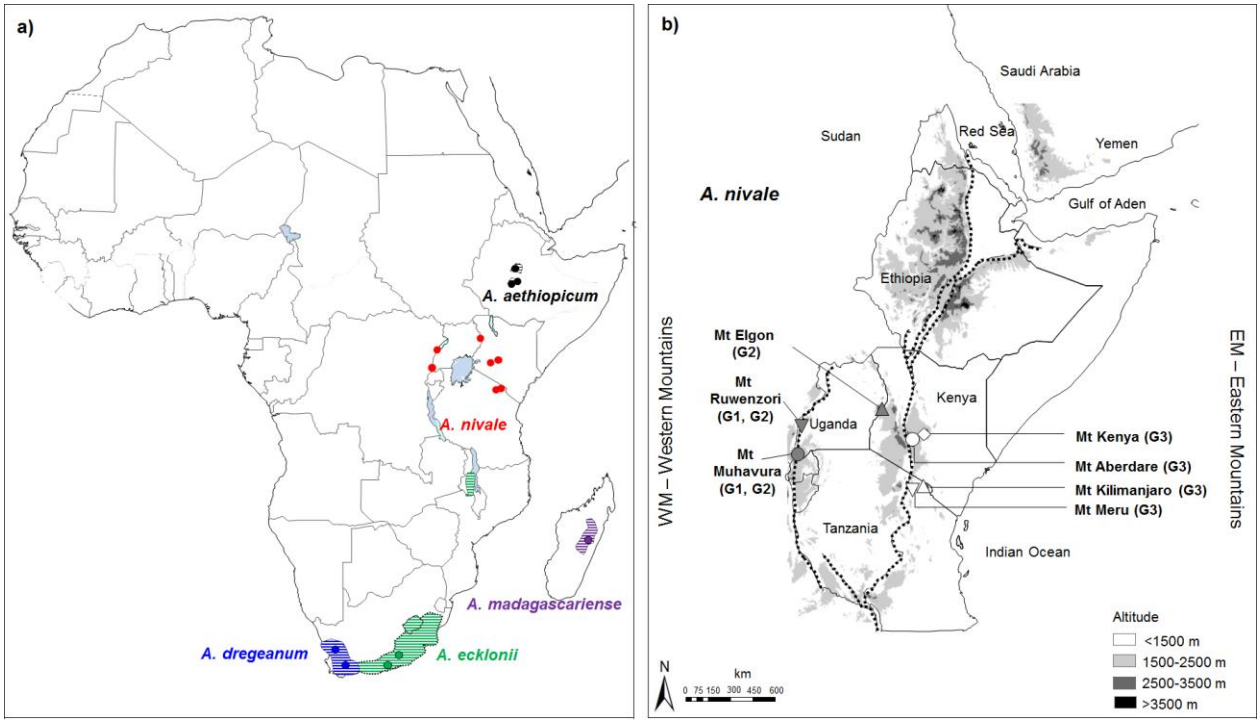
942 **Fig. S5** Tanglegram of Bayesian maximum clade credibility trees based on plastid (right) and
943 nuclear (left) data.

944 **Fig. S6** Results of STRUCTURE analyses based on AFLP data used to determine optimal
945 number of genetic groups (*K*) in *A. nivale* (left: total dataset; middle: Eastern Mountain
946 group; right: Western Mountain group).

947 **Fig. S7** Principal Coordinates Analyses (PCoA) based on Dice's coefficient of similarity
948 among AFLP genotypes of *A. nivale.*

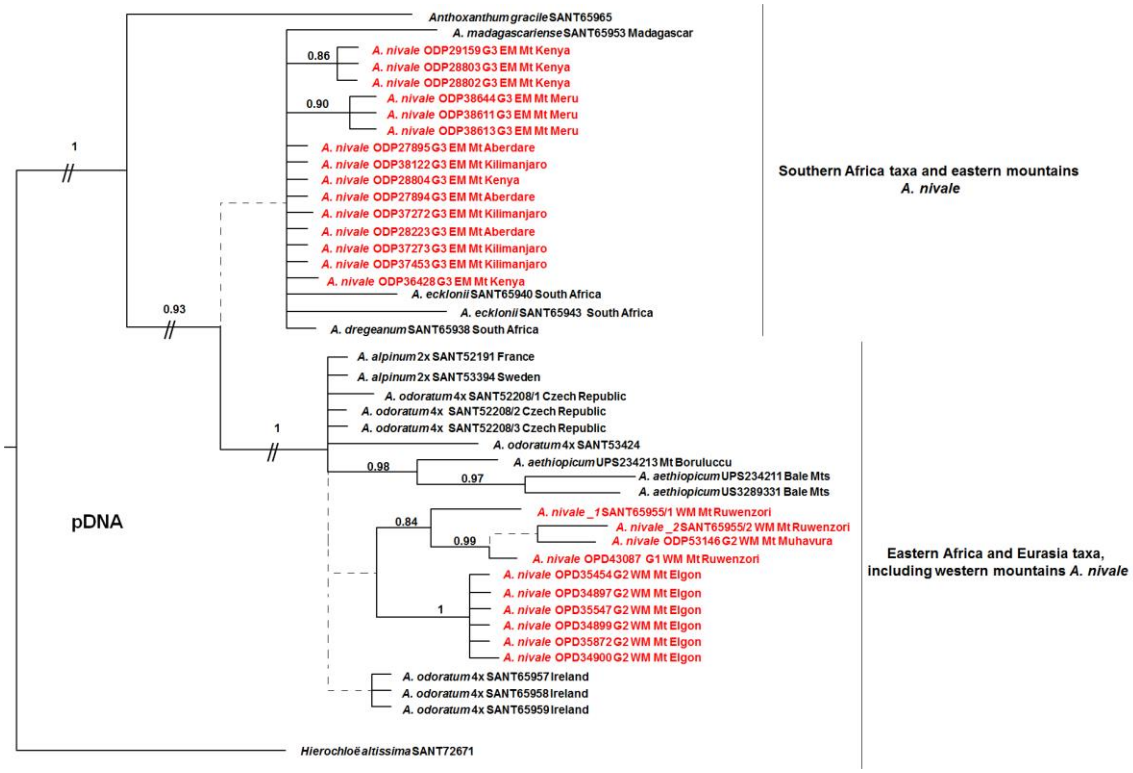
949

950 Fig. 1



951

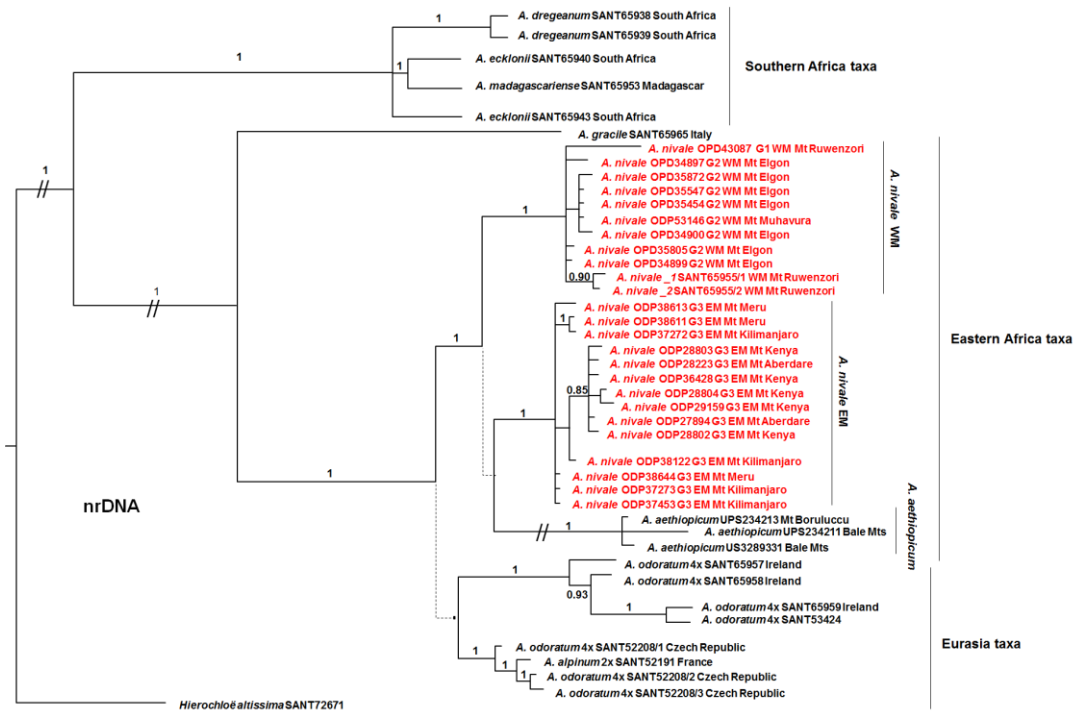
952 Fig 2



953

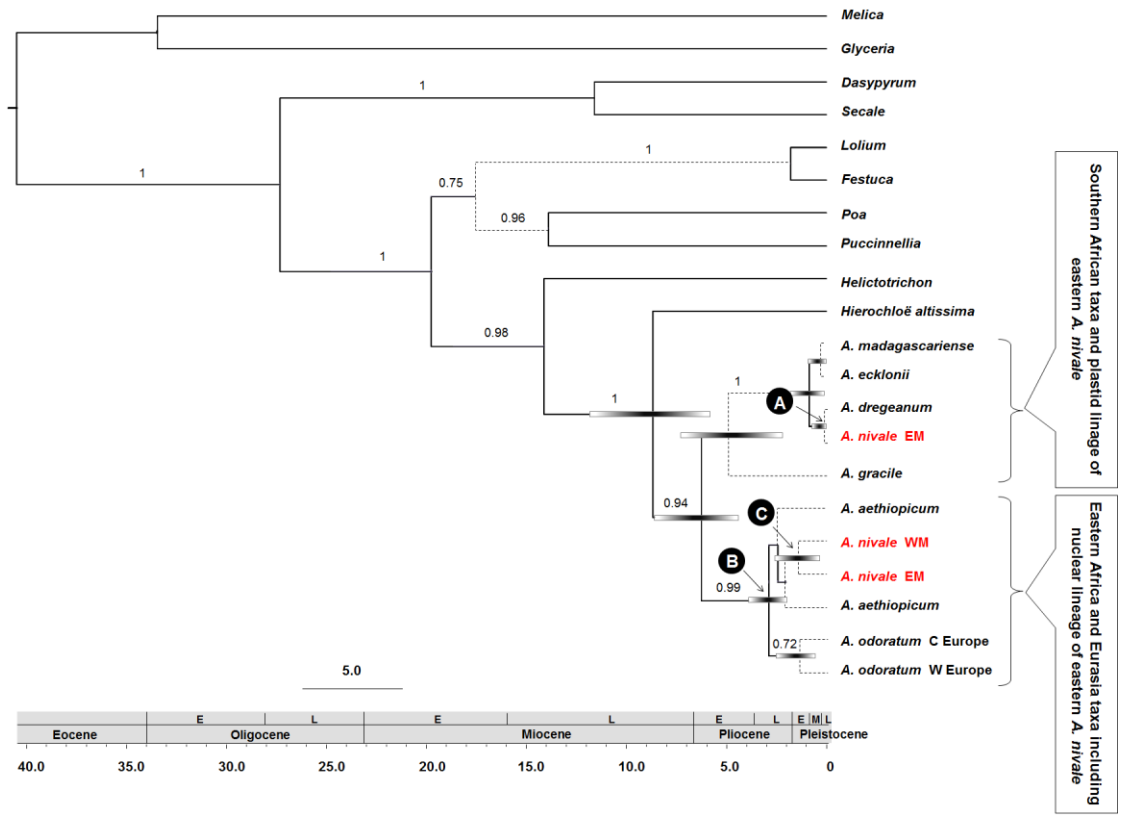
954

955 Fig 3



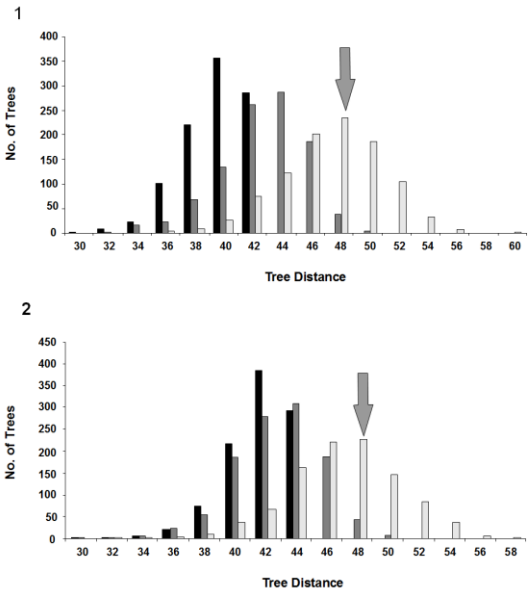
956

957 Fig 4



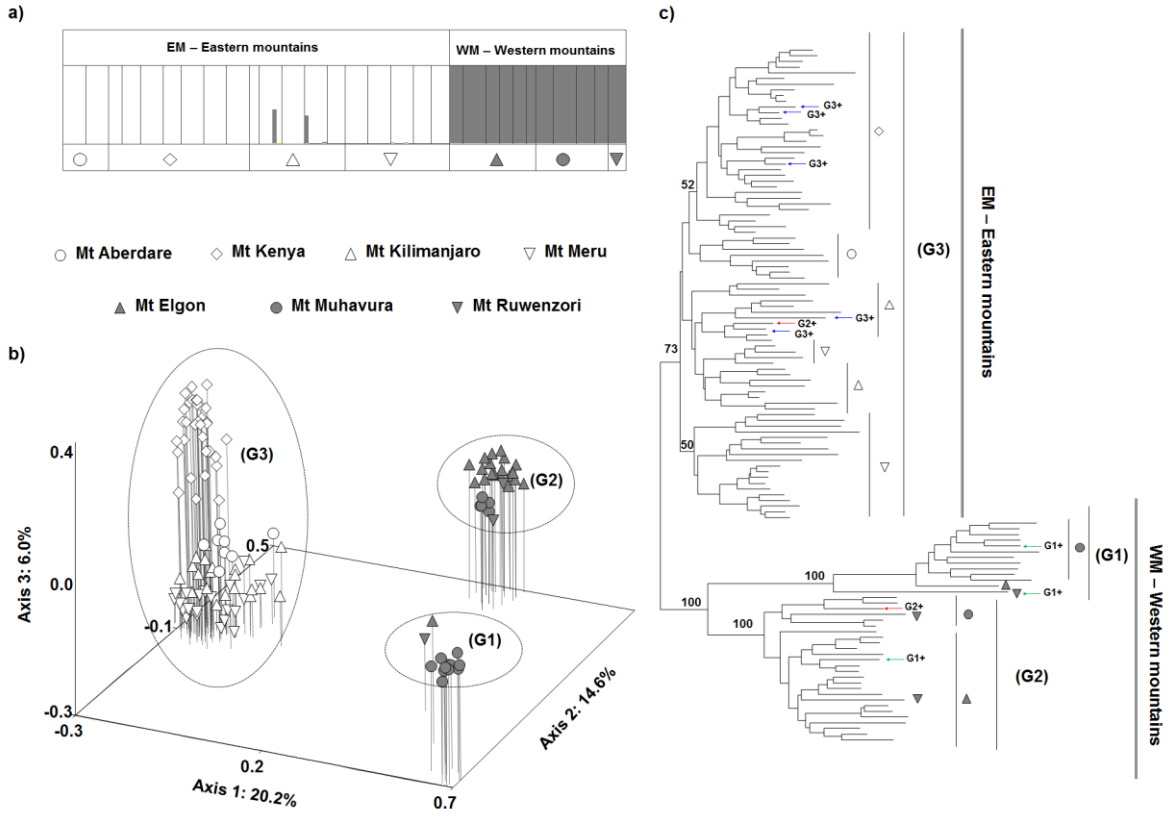
958

959 Fig 5



960

961 Fig 6



962