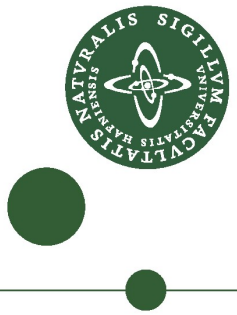
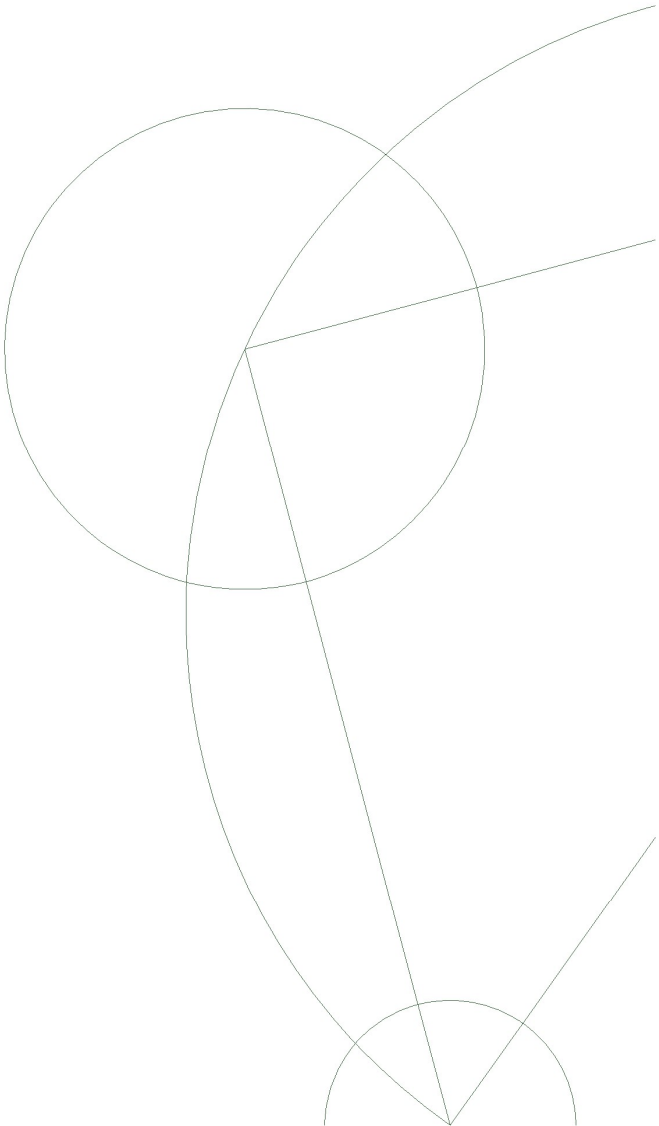
FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN



Population Genetics Project

Stefano Pellegrini `mlq211`

Splitting and admixture in the Grant's gazelle species complex



Supervisor: Genís Garcia-Erill

Saturday 28th March, 2020

1. Abstract

Recent studies [1] on mtDNA revealed that Grant's gazelle consists of three species (*N. granti, N. notata* and *N. petersii*) rather than one. In this study, RAD sequencing data [2] from 92 individuals of Grant's gazelle and 3 individuals of *Eudorcas thomsonii* were analyzed. Genetic admixture and TreeMix [3] analysis supported the subdivision, proposed by Siegismund et al. (2013) [4], of *N. granti* species into *N. g. granti* and *N. g. robertsii* subspecies. PCA, genetic admixture and TreeMix analysis supported the hyphotesis that, the Mkomazi population currently classified as *granti*, might be a hybrid population that resulted from the contact between *granti* and *petersii*. Fixation index analysis showed that this population is genetically more similar to *petersii* than to *g. granti*. These results raised the necessity to question the validity of the current classification of the Mkomazi population, and, eventually, its implication in the conservation management of the different groups.

2. Introduction

Background

The species of the genus *Nanger* belonged to the genus *Gazella* until recently, when *Nanger* was elevated to full genus status [1]. The supergenus comprising *Nanger* and *Gazella* is considered one of the most taxonomically complex groups within the family *Bovidae* [1]. Its species show extensive intraspecific variation in horn size and shape, body size, coat color and chromosome number, which has hindered past conservation effort [1]. The taxonomic classification within the *Nanger* genus is puzzling and the literature is inconsistent. The Grants gazelles (*Nanger granti*) distributed in East Africa were originally considered a single species, but was later regarded to consist of several subspecies. Kingdon (1982) [5] described the 4 subspecies *petersii, brighti, robertsi and granti*, while Grubb (2005) [6] described *notata* but did not include *granti*. Studies [1] based on mtDNA have shown significant genetic differences that suggest three distinct species groups, where *granti* and *robertsii* are clustered in one group, *notata* and *bright* are clustered in a second group, and *petersii* are clustered in a third group. Several subspecies have been described and characterized based on phenotypic variation such as coloration and horn size. But the variability of traits within populations, the little to no genetic support [4], and the difficulty of distinguishing subspecies in the field, have questioned the validity of the taxonomic divisions [1].

The Grants gazelles range from southern Sudan and Ethiopia to Central Tanzania and from the Kenya/Somali coast to Lake Victoria (Siegismund et al, 2013) [4]. The *notata* species is distributed in the northern area, the *granti* to the South-West and the *petersii* to the East. The *petersii* distribution is restricted by the Galana and Tana Rivers and the latter prevents its contact to *notata* [4]. The contact between *petersii* and *granti* was prevented by dense vegetation barriers of Acacia-Commiphora woodlands but during recent decades, due to environmental changes, the woodlands have been transformed into open grassland [1]. This allowed the expansion of the *granti* toward East, in Tsavo, which was previously occupied only by *petersii* [7].

The Grants gazelles are currently categorized as least concern by the International Union for Conservative Nature (IUCN), but their numbers are decreasing [8]. Having a meaningful characterization of subspecies is crucial for conservation biology. If a presumed population consists of several genetically distinct populations, the degree of endangerment could be misleading. Genetically, a small effective

population size can cause several problems for a population [9], mainly because the effect of genetic drift in a small population is much stronger and can lead to loss of genetic variability. This will, in turn, lead to a decrease in heterozygosity, an increase in deleterious alleles and vulnerability to environmental changes. Knowing the genetic relationship is necessary to guide the preservation of genetic variability within a population.

Geographical distribution

In this study, we used data collected from 95 individuals, where 92 of these were Grant's Gazelles (*N. granti, N. notata* and *N. petersii*) and 3 were Thomson's gazelles (*Eudorcas thomsonii*). Three individuals from the species *Eudorcas thomsonii* were included as an outgroup. The samples were collected from 13 different localities distributed in Kenya and Tanzania (Figure 1).

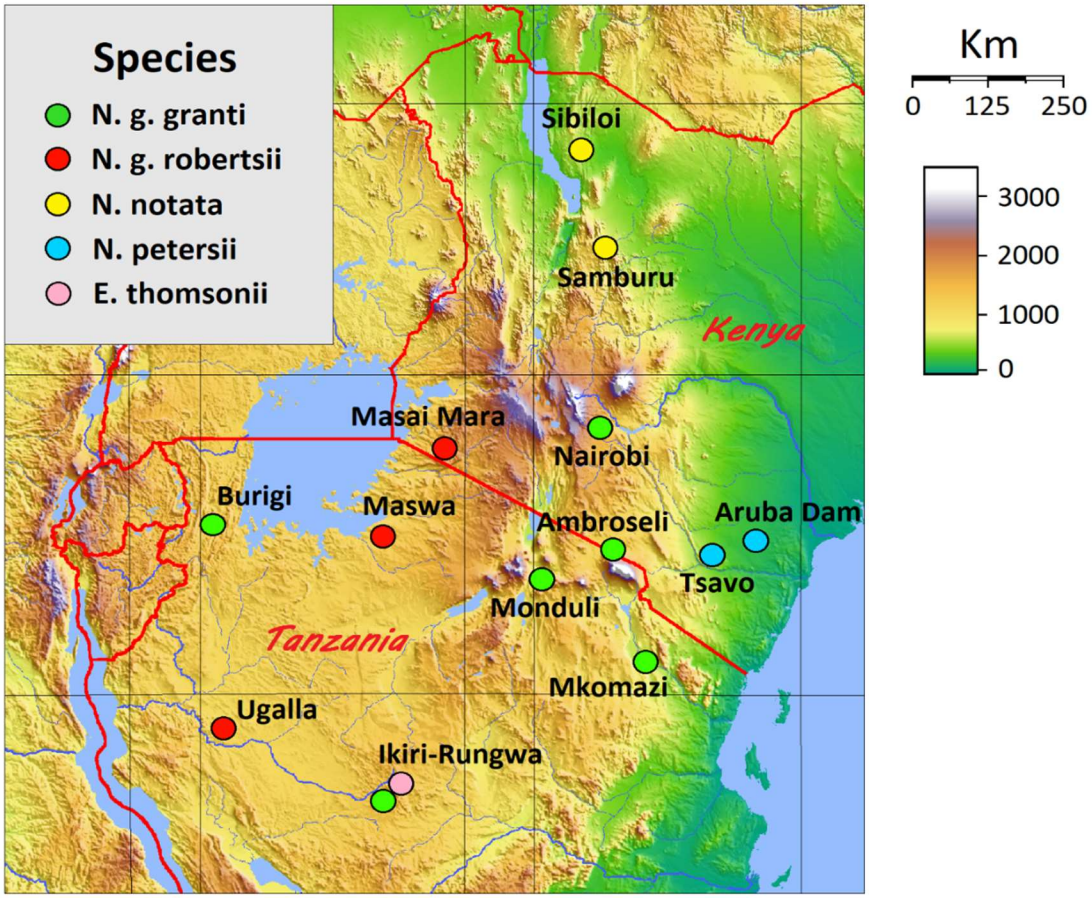


Figure 1. Map of Kenya and Tanzania showing the sampling localities [10].

The *N. granti* (*g. granti* and *g. robertsii*) samples were collected in the South-West part of the region (Ugalla, Maswa, Masai Mara, Burigi, Monduli, Mkomazi, Amboseli, Nairobi), *N. notata* in the north (Sibiloi and Samburu), *N. petersii* in the east (Tsavo and Aruba Dam) and *E. thomsonii* in the south (Ikiri-Rungwa) (Table 1). One individual from *g. granti* subspecies were collected from Ikiri-Rungwa and it was excluded from all the analysis where the individuals were clustered by localities.

Region	Species	Subspecies	Locality	Samples
SW	*granti*	*granti*	Amboseli	3
			Nairobi	5
			Monduli	15
			Mkomazi West	4
			Mkomazi East	8
			Burigi	1
			Ikiri-Rungwa	1
			Total	37
		robertsii	Masai Mara	11
			Maswa	14
			Ugalla	1
			Total	26
			Total (g. species)	63
E	*petersii*		Tsavo	11
			Aruba Dam	8
			Total	19
N	*notata*		Samburu	7
			Sibiloi	3
			Total	10
S	*thomsonii*		Ikiri-Rungwa	3
	Total			95

Table 1. Information about sampling localities and number of samples for each species and population.

3. Materials and methods

3.1. Data sampling

Tissue samples of 95 individuals were collected between 1991 and 1998 from 13 different localities in Kenya and Tanzania. The data was generated by a technique called Restriction site Associated DNA sequencing (RAD sequencing) [2]. RAD sequencing cut the genome with a restriction enzyme (in this case Sbf1) and the final DNA sample was reduced to 10 megabases for each individual. The data consisted of PLINK files [11] with SNP genotypes from the variable positions in the DNA subsampled by the restriction enzyme. It has been pre-filtered removing sites with more than 20% missing data across individuals, that reduced the number of SNPs to 35174. An additional file (grants_popInfo.txt) with samples' information (species and localities) was also provided.

3.2. Data analysis

Principal Component Analysis

We started our analysis by performing a principal component analysis (PCA). It is a powerful tool for data visualization, and it can be useful to get a general picture of the data and to plan following analysis. The idea of PCA is to give a location to each individual data-point on each of a small number of principal component axes. These PC axes are chosen to reflect major axes of variation in the data [12].

We performed three PCAs (two eigendecompositions, plus one additional clustering view), in the first one the individuals were clustered by species and subspecies (*g. granti, g. robertsii, notata, petersii* and *thomsonii*) and all individuals were included. Since the PCA is very sensitive to outliers, in the second one we used the same group division but this time we removed the outgroup (*E. thomsonii*). In the last one, the outgroup was removed and the individuals were clustered by both localities and species. As a common preprocessing step for all the performed PCAs, we removed all SNPs that presented at least a missing value across individuals, resulting in 195 variable sites left. Also, we removed each SNP where all individuals were homozygous for the same allele, which reduced the number of the variable sites to 173. Finally, we converted the genotypes 1, 2, 3 (homozygous reference allele, heterozygous and homozygous derived allele) to 0, 1 and 2 by subtracting 1 to all genotypes. Next, we computed the allele frequency, standardized the resulting data, and performed the PCA by eigendecomposition of the covariance matrix. To divide the data points in clusters we retrieved the species and localities information from the samples' information file (grants_popInfo.txt). We used this information to sort the plink files by species, and to color the clusters projected into the first two principal components obtained in the previous step.

Nucleotide diversity

After PCA, we estimated the nucleotide diversity within the different species and populations (divided by localities). The nucleotide diversity estimation is a measure of genetic variation that, if associated with other genetic variability measures (e.g. fixation index), can be useful to understand the diversity within or between ecologically related populations [13].

As we did in the PCA, we preprocessed the data by retrieving information from the samples' information file. This time we used information about species and localities to generate id.txt files that were used as input for subsetting the genotypes in the plink files. We generated a subset of the genotypes data for each species and population, including a subset of *granti* species and one of *g. granti* subspecies without Mkomazi. Also, in the populations' subsets, we did not include the localities where we only had one sample: Ugalla (*g. robertsii*) and Burigi (*g. granti*). For this analysis, we removed the *g. granti* sample from Ikiri-Runwa, because all the other 3 members of this population were *thomsonii* (outgroup). Next, we used the genotypes to calculate the allele frequency for each variable site in each cluster (species or populations) and consequently we removed the fixed allele. Finally, we computed the average of the frequencies of the remaining variable sites, and we estimated the average expected heterozygosity assuming Hardy-Weinberg proportion by applying the formula

$$H_e = 2p(1-p)$$

Fixation index

Another measure of genetic variability is the fixation index (FST), we used it to investigate the genetic variation between species and populations. The fixation index can be considered as the reduction in heterozygosity due to population structure [14]. There are different fixation index estimators, Wright's fixation index can be calculated by computing the difference between the genetic variation between populations (H_T, expected heterozygosity of the combined populations) and the genetic variation within populations (H_S, average of the subpopulations expected heterozygosity), standardized by the genetic variation between population (H_T). If the FST value is small, it signifies that the allele frequency between the populations are similar, if it is large, it signifies that they are different [15].

$$F_{\mathrm{ST}} = \frac{\text{variation between populations} - \text{variation within populations}}{\text{variation between populations}}$$

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Wright's estimator [16].

In our analysis we used the Weir and Cockerham fixation index estimator, it uses the analysis of variance (ANOVA) to estimate the components of the variance within and between populations [17]. We used this estimator because, compared to Wright's estimator, it is less biased with respect to the sample size [18].

As for most of the preceding steps, we grouped the genotypes by species and subspecies using the samples' information file, and we removed all SNPs that presented at least a missing value across the individuals. For this analysis, we excluded the samples from Mkomazi from the *g. granti* subspecies, and we considered them as an individual population. Finally, we calculated the FST pairwise between the six groups (*g. granti, g. robertsii,* Mkomazi*, petersii, notata and thomsonii*). We calculated the Weir and Cockerham estimator [17] by applying the formula

$$\Theta = \frac{a}{a + b + c}$$

where *a* is the variance between population, *b* is the variance between individuals and *c* is the variance within individuals. Finally, we computed the weighted average of θ.

Admixture

After measuring the genetic variation between and within populations we performed an admixture analysis. Since at least three species live in neighboring ranges, the purpose of this test was to check if there has been any admixture event between the populations, and so if there are any hybrids populations within Grant's gazelles.

We started by sorting the genotypes by species, then we used the ADMIXTURE [19] tool to estimate the individual ancestry from SNP data. We wanted to know which was the best clustering given our data, so we performed different cross-validation tests with values of K (distinct number of populations) ranging from 3 to 8. Next, we plotted the admixture obtained by the five K-values, reporting the species and the populations of each sample.

Treemix

In the last step of our analysis, we used the TreeMix [3] software to further investigate the admixture hypothesis and the presence of gene flow between populations. TreeMix is a genetic drift-based program, used for understanding patterns of population split and migration events [3]. The software takes a set of allele frequencies from different populations as input, and it returns the drift tree which has the maximum likelihood for the given set. As an additional feature, it can also infer a different number of admixture events [3].

As a preprocessing step, we generated a .fam file as a family ID file for each species and location. Next, we calculated the allele frequency for each SNP and we converted our files into treemix.frq files by the plink2treemix.py script [3]. Finally, we used the TreeMix software to build the maximum likelihood tree based on the allele frequencies. We generated two trees, in one we clustered the samples by species and in the other we clustered them by locations. In both trees we included the outgroup, we used it as root (*thomsonii* for the species and Ikiri-Rungwa for locations), and we allowed two migrations events.

3. Results

Principal Component Analysis

In the first PCA we divided the data by species, and we included the *thomsonii*. In the PCA plot (Figure 3) can be noted that most of the variance captured by the principal components (74.9% by the first one and 4.2% by the second one) is explained by the differences between the outgroup and the Grants gazelles, for this reason it is not possible to distinctly observe differences between the *Nanger granti* species. That was expected because the PCA technique is very sensitive to outliers [20], so for the next principal component analysis we decided to remove the outgroup.

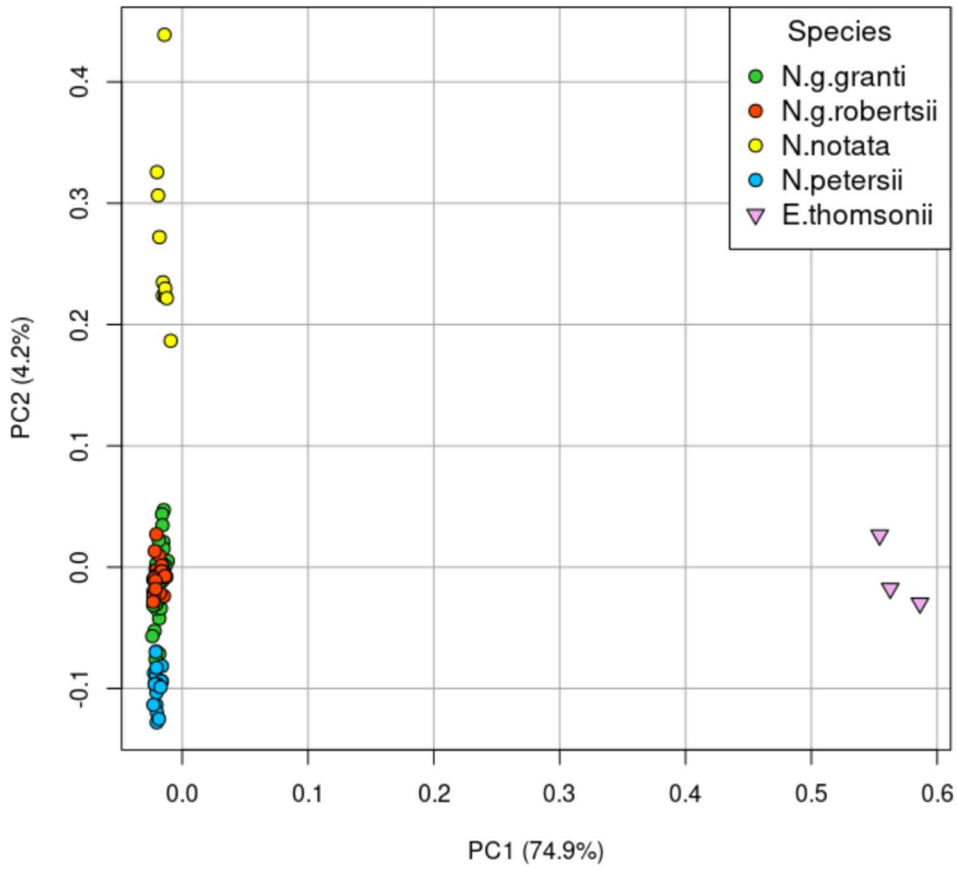


Figure 3. PCA of the genotype data divided by species, the outgroup is included.

In the second PCA we repeated the division by species but this time we removed the *thomsonii*. Without the outgroup, the two principal components captured respectively 77.2% and 4.1% of the variance. The plot in Figure 4 shows the presence of four main clusters, a first one composed by part of *g. granti* together with *g. robertsii*, a second one composed by *notata*, and two close clusters composed by part of *g. granti* and *petersii*. The grouping of *g. granti* with *g. robertsii* was expected since they are both members of the *granti* species, but it is interesting to observe that a subgroup of the *g. granti* are closer to the *petersii* than the rest of the individuals of the *granti* species. Since we know from the literature that the thinning of the Acacia-Cammiphora woodland allowed the granti and petersii to come into contact [4], the presence of these two close clusters could be explained as the results of an admixture event between the two groups.

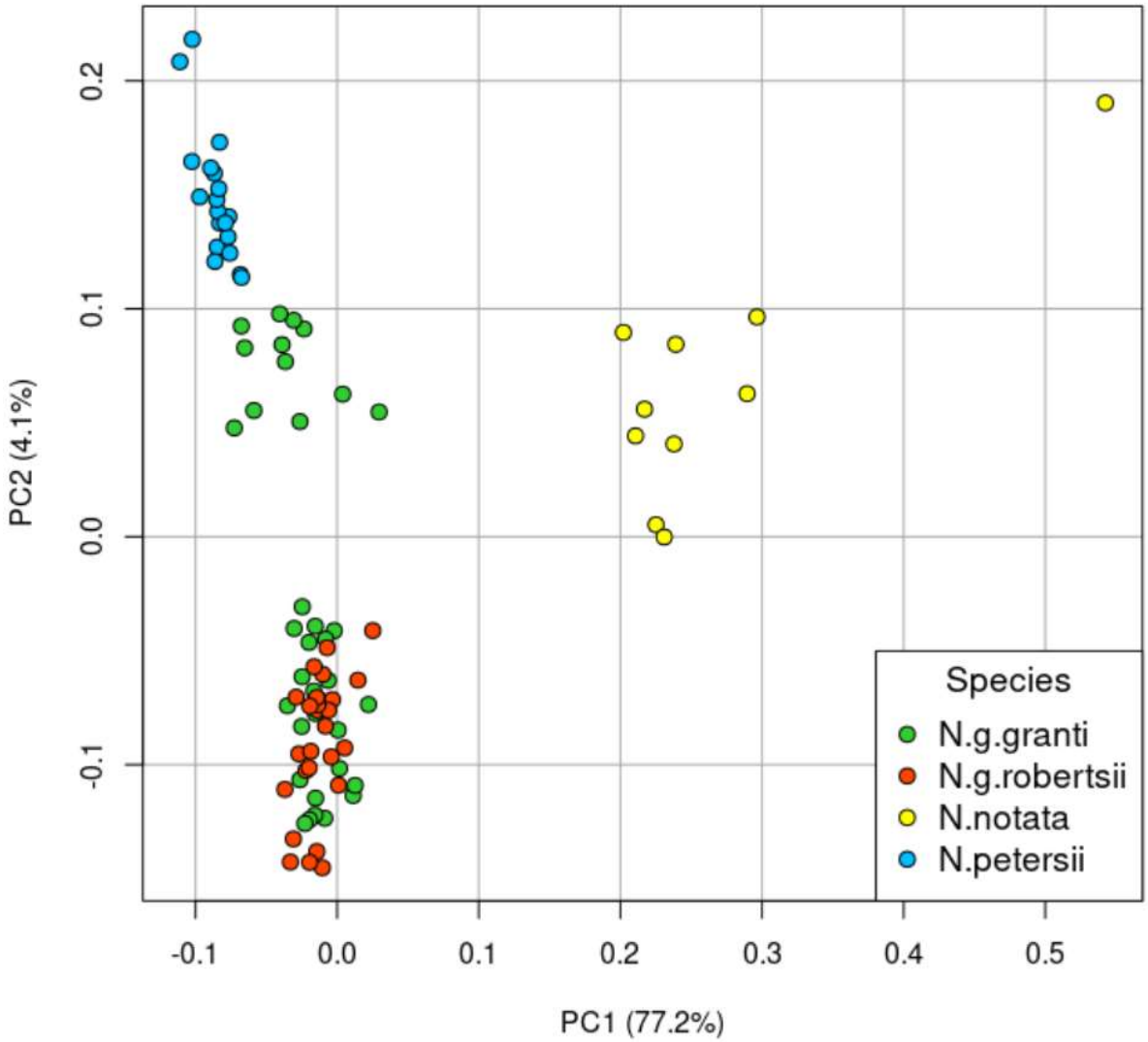


Figure 4. PCA of the genotype data divided by species, the outgroup is excluded.

In the last PCA we removed the outgroup and we divided the data by both species (data-points shapes) and localities (data-points color). In the PCA plot in Figure 5, it is possible to observe the localities of the four clusters, in particular, we can define the two populations of the two close clusters that we observed in the previous PCA (Figure 4). It is interesting to note that the subgroup of the *g. granti*, that formed an individual cluster close to the *petersii*, is composed of the samples collected from Mkomazi.

As expected, the cluster composed by the *petersii* contains individuals collected from Tsavo and Aruba Dam. Observing the map (Figure 1) we can see that the Mkomazi and the *petersii* populations live in neighboring ranges and the boundary between Mkomazi and Tsavo is delimited by a thin Acacia-Cammiphora wooded grassland (appendix Figure 2).

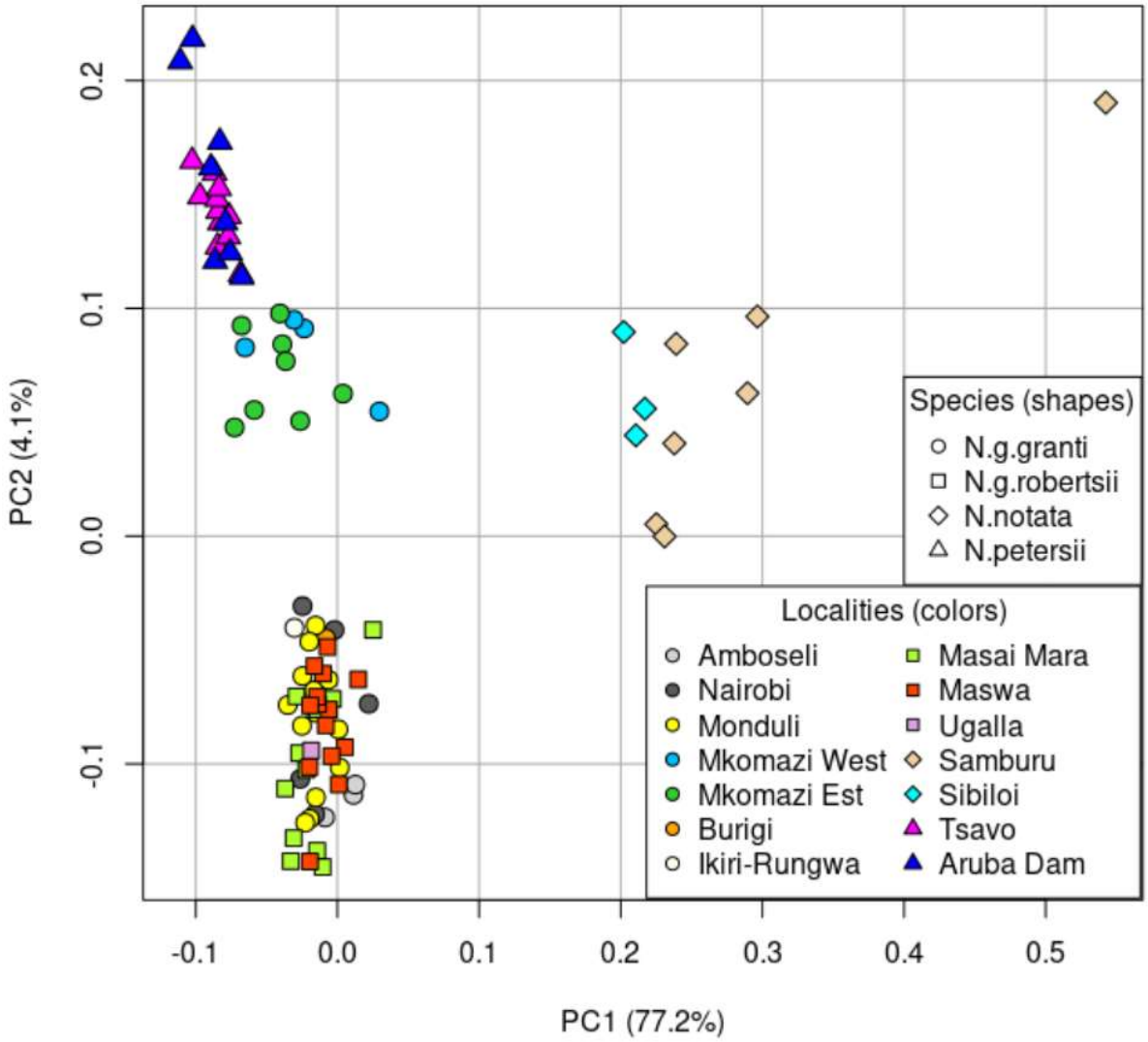


Figure 5. PCA of the genotype data divided by localities and species, the outgroup is excluded.

Nucleotide diversity

From Table 2 and the plots in Figures 6 and 7, we can observe that *thomsonii* and *petersii* had the lowest level of nucleotide diversity, respectively 0.000234 and 0.000235, suggesting smaller historic population size and a potential stronger drift that may have reduced their genetic variation. *Notata* had a value of 0.000385 and *granti* (*g. granti* and *g. robertsii*) presented the largest nucleotide diversity with a value of 0.000402. Within the *granti* species the *g. granti* had larger genetic diversity than *g. robertsii* with values of 0.000409 and 0.000339 respectively. This may be explained by the fact that *g. granti* has a larger distribution range. It is interesting to note that Mkomazi, with a value of 0.000362 had the largest genetic diversity within the granti species.

Since the population from Mkomazi could have been a hypothetical hybrid population resulted from the contact between *g. granti* and *petersii*, we also analyzed the genetic diversity of *granti* species and *g. granti* subspecies excluding the Mkomazi. The two groups resulted to have a value of 0.000364 (*granti* species) and 0.00035 (*g. granti* subspecies), indicating that by removing the Mkomazi we observed a reduction in the nucleotide diversity of the two groups by 14.54% and 9.36% respectively. Also, we can observe that by removing the Mkomazi population from *granti* species, *notata* become the species with the largest genetic variation.

Table 2. Information about average nucleotide diversity (π), sampling localities and number of samples for each species and population. In *granti* species and *g. granti* subspecies, π is also measured without the Mkomazi.

Species	Subspecies	Locality	Samples	π
granti	*granti*	Amboseli	3	0.0002987
		Nairobi	5	0.0003065
		Monduli	15	0.0003498
		Mkomazi	12	0.000362
		Burigi	1	
		Ikiri-Rungwa	1	
		Total	37	0.0004098
		(subspecies)	22	0.0003502
		Total no Mkomazi		
	robertsii	Masai Mara	11	0.0003347
		Maswa	14	0.0003164
		Ugalla	1	
		Total	26	0.0003396
		(subspecies)		
			63	0.0004027
		Total (g. species)	48	0.0003649
petersii		*Total no Mkomazi*	11	0.0002332
		Tsavo		
		Aruba Dam	8	0.0002265
		Total	19	0.0002352
notata		Samburu	7	0.0003718
		Sibiloi	3	0.000297
		Total	10	0.0003853
thomsonii		Ikiri-Rungwa	3	0.0002342

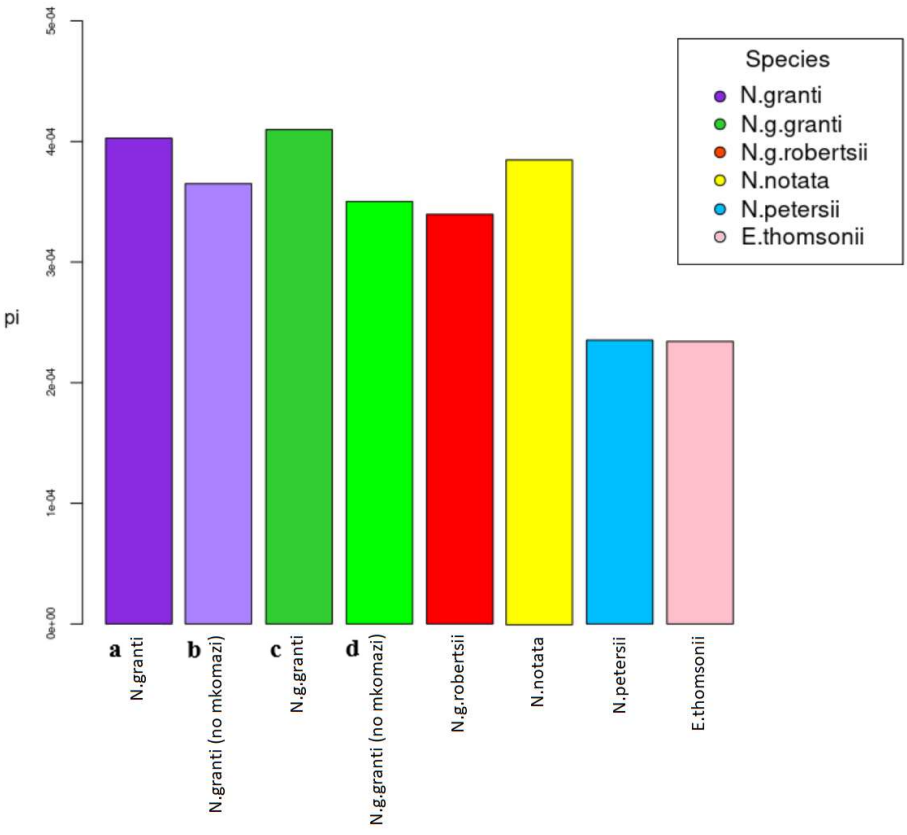


Figure 6. Nucleotide diversity (π) within species and subspecies. [a] *N. granti* (species) with Mkomazi included, [b] *N. granti* (species) with Mkomazi excluded, [c] *N. g. granti* (subspecies) with Mkomazi included, [d] *N. g. granti* (subspecies) with Mkomazi excluded.

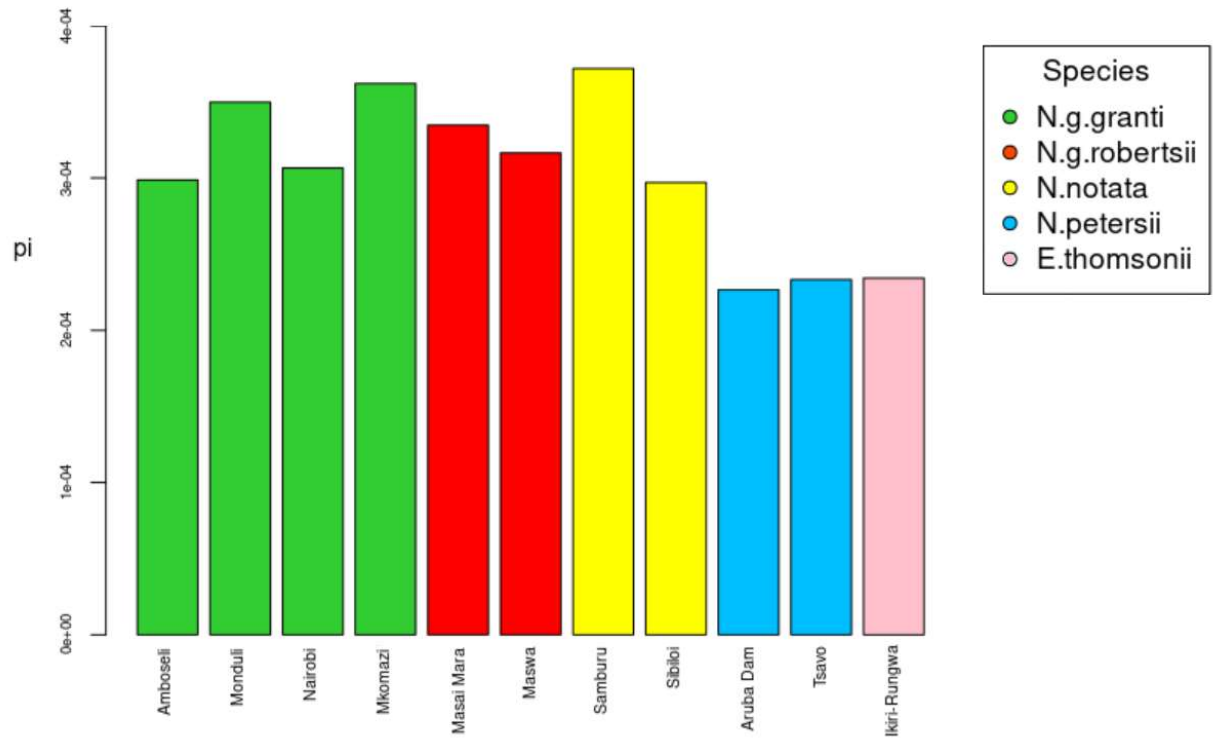


Figure 7. Nucleotide diversity (π) by localities, the colors are assigned depending on the species.

Fixation index

Table 3 shows the pairwise differences in the allele frequencies between species and subspecies. For this analysis, we excluded the Mkomazi samples from the *g. granti* subspecies and we considered them as an independent group. As expected, since they are members of the *granti* species, *g. granti* and *g. robertsii* had the lowest level of differentiation with an FST value of 0.0164. But the most interesting result was that the pair Mkomazi and *petersii*, with a fixation index of 0.0338, had a lower level of differentiation than the pair Mkomazi and *g. granti*, with an FST value of 0.037. This result revealed that the Mkomazi population is genetically more similar to *petersii* than to *g. granti* subspecies, which is the current taxonomical group used for their classification. This made us question the validity of the current classification of Mkomazi population as a member of the *g. granti* subspecies. The other pairs showed larger FST values and, as expected, the *thomsonii* pairs between Grants gazelle's species presented the largest level of differentiation.

FST					
	g. granti	***g. robertsii***	***Mkomazi***	***petersii***	***notata***
g. granti		0.01640628	0.03706388	0.05615573	0.06158969
g. robertsii	0.01640628		0.04400392	0.06362729	0.06594527
Mkomazi	0.03706388	0.04400392		0.03384605	0.04984582
petersii	0.05615573	0.06362729	0.03384605		0.09082416
notata	0.06158969	0.06594527	0.04984582	0.09082416	
thomsonii	0.29299145	0.30924673	0.29307661	0.34829738	0.26519969

Table 3. Pairwise FST values between species, subspecies and Mkomazi population, which is excluded from the *g. granti* subspecies and is considered as an independent group.

Admixture

To evaluate the different clustering of our data, we used ADMIXTURE [19] to perform different cross-validation test with K (number of ancestral populations) ranging from 3 to 8. Table 4 shows that K equals 7 resulted to have the lowest error.

K	CV error
3	0.16170
4	0.15732
5	0.15123
6	0.14432
7	0.14160
8	0.15132

Table 4. Admixture cross-validation error for different values of K.

In the plot (Figure 8), the colored bars indicate the ancestry proportion in each individual. With K equal 5, which is the number of clusters we observed in the PCA2 and 3 (four clusters plus outgroup), the Mkomazi had both *granti* and *petersii* ancestry, with a greater proportion of the latter. This supported our hypothesis that the *g. granti* population from Mkomazi is a hybrid population that resulted from the contact between *granti* and *petersii*. With K equal 4, the Mkomazi and *notata* populations had the same ancestry (*granti* and *petersii*). With K equal 6, the Mkomazi had a small proportion of *petersii* ancestry, and a larger one that was not found in the other populations. For all K larger than 3, the analysis also supported the validity of the division of *granti* species in *g. granti* and *g. robertsii* subspecies.

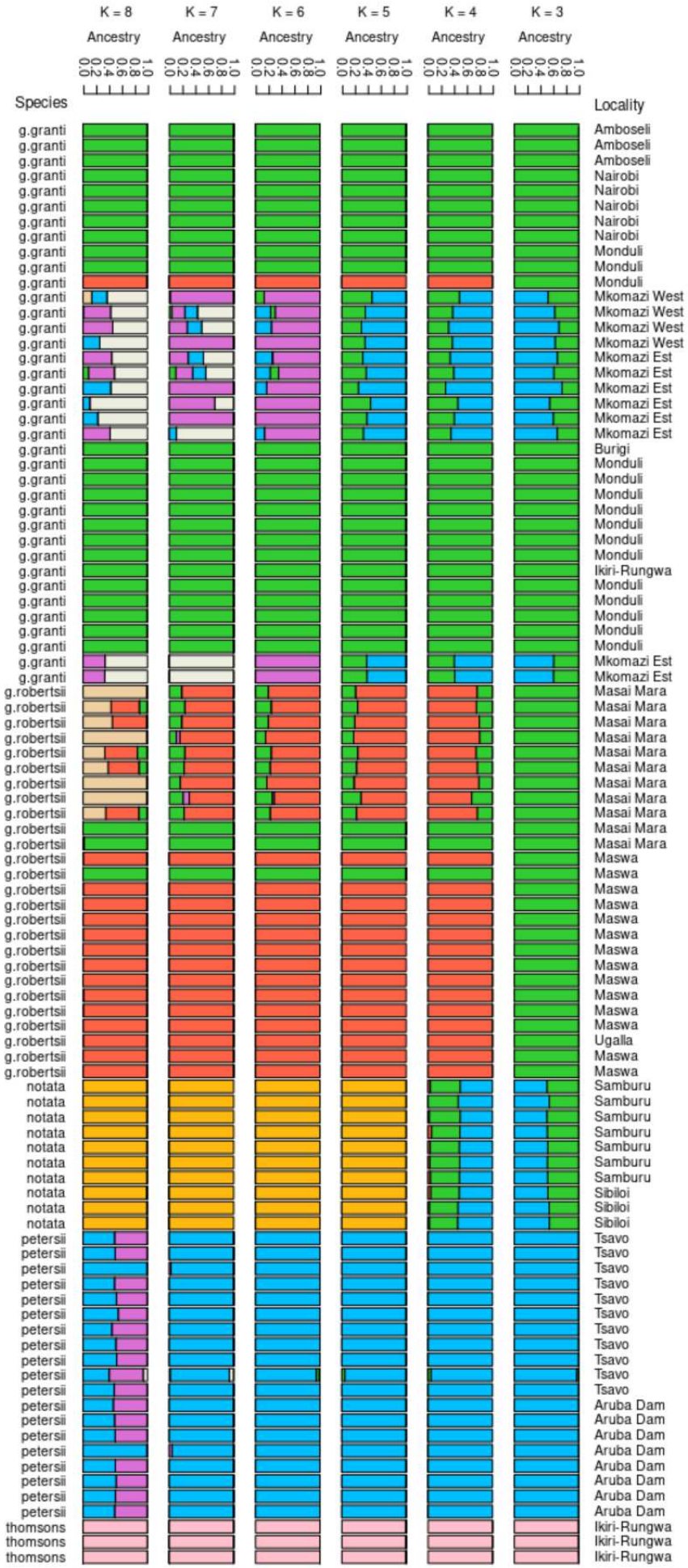


Figure 8. Admixture plot, generated by ADMIXTURE [19], with K-values ranging from 3 to 8.

Furthermore, it indicates the presence of a potential admixture event between the subspecies just mentioned, in particular between some individuals from the Masai Mara population and the *g. granti*. That may be a plausible event since they show no geographic isolation [4].

The admixture plot also shows that there are some hypothetical labeling errors in the dataset. First, there is a sample from Monduli that has been classified as *g. granti*, but we can see that it only had *g. robertsii* ancestry. Second, there are two individuals from Masai Mara and one from Maswa that have been classified as *g. robertsii*, but they only had *g. granti* ancestry.

TreeMix

We used the TreeMix [3], a genetic drift based program, to reconstruct the genetic relationship between populations. We generated two maximum likelihood trees, in the first one we subset the data by species and in the second by locations. As it is possible to see in Figures 9 and 10, in both trees we included the outgroup and, since we knew from the admixture analysis that we had two potential admixture events, we also allowed two migrations events. The colors of the migration arrows are defined according to their weight and they point in the direction toward the recipient group. The lengths of the horizontal branches are proportional to the amount of genetic drift that has occurred on the branch. The scale bar shows ten times the average standard error of the values in the sample covariance matrix of the allele frequencies [22], while the drift parameter is a relative measure of time.

In the first tree (Figure 9), *E. thomsonii* (the outgroup) and Grants gazelles' branches started from the root node. The first node in Grants gazelles' branch divided the petersii from the rest of the *Nanger granti*, the second one divided *notata* and *granti* species, and the last one divided *g. granti* and *g. robertsii* subspecies.

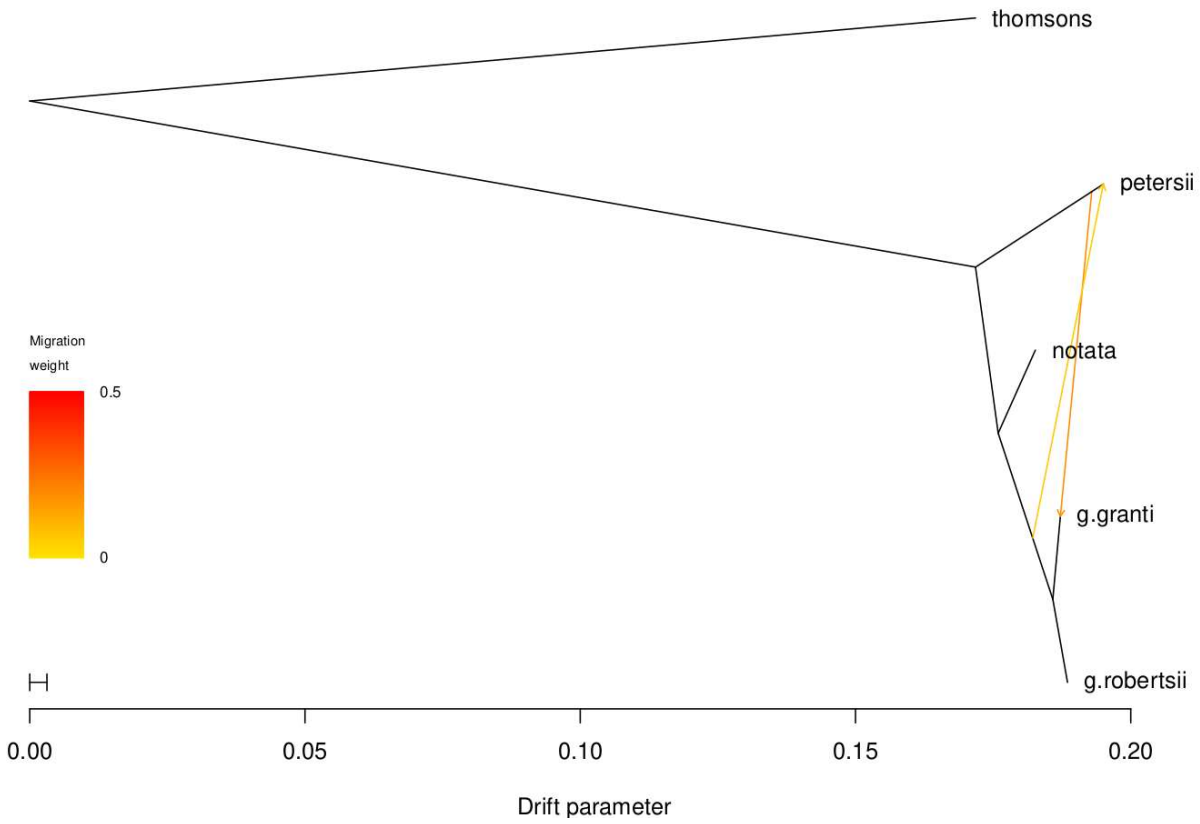


Figure 9. Maximum likelihood drift tree generated by TreeMix [3]. Two migration events are allowed, and data is clustered by species.

The migration arrows indicate that there was gene flow from *petersii* to *g. granti* and from *granti* lineage to *petersii*, the latter with lower weight than the former. This supported the presence of the first admixture event (Mkomazi and *petersii*) we found in the admixture analysis, but not the second one (Masai Mara and *g. granti*). Also, we can observe that *petersii* has been subjected to stronger genetic drift than the rest of the Grants gazelles, this was also supported by the lowest genetic variability we found within this species in the nucleotide diversity analysis (Table 2).

As in the first one, in the second tree (Figure 10) we placed the outgroup (this time Ikiri-Rungwa population) at the root. The first node of the Grants gazelle populations branch divided the *granti* populations (except Mkomazi) from *petersii* and *notata* populations. It is interesting to note that the *g. granti* population from Mkomazi was placed in the same branch as *petersii*. This time the migration arrows indicate that there is gene flow from the Amboseli *g. granti* population to Mkomazi. These migration events are of doubtful interpretation, but the placement of Mkomazi close to Tsavo and Aruba Dam populations supported the hypothesis that Mkomazi is strongly related to *petersii*. Finally, we can see that the populations of *g. robertsii* and *g. granti* has been placed on two different branches, supporting the validity of the division of the *granti* species in the two subspecies.

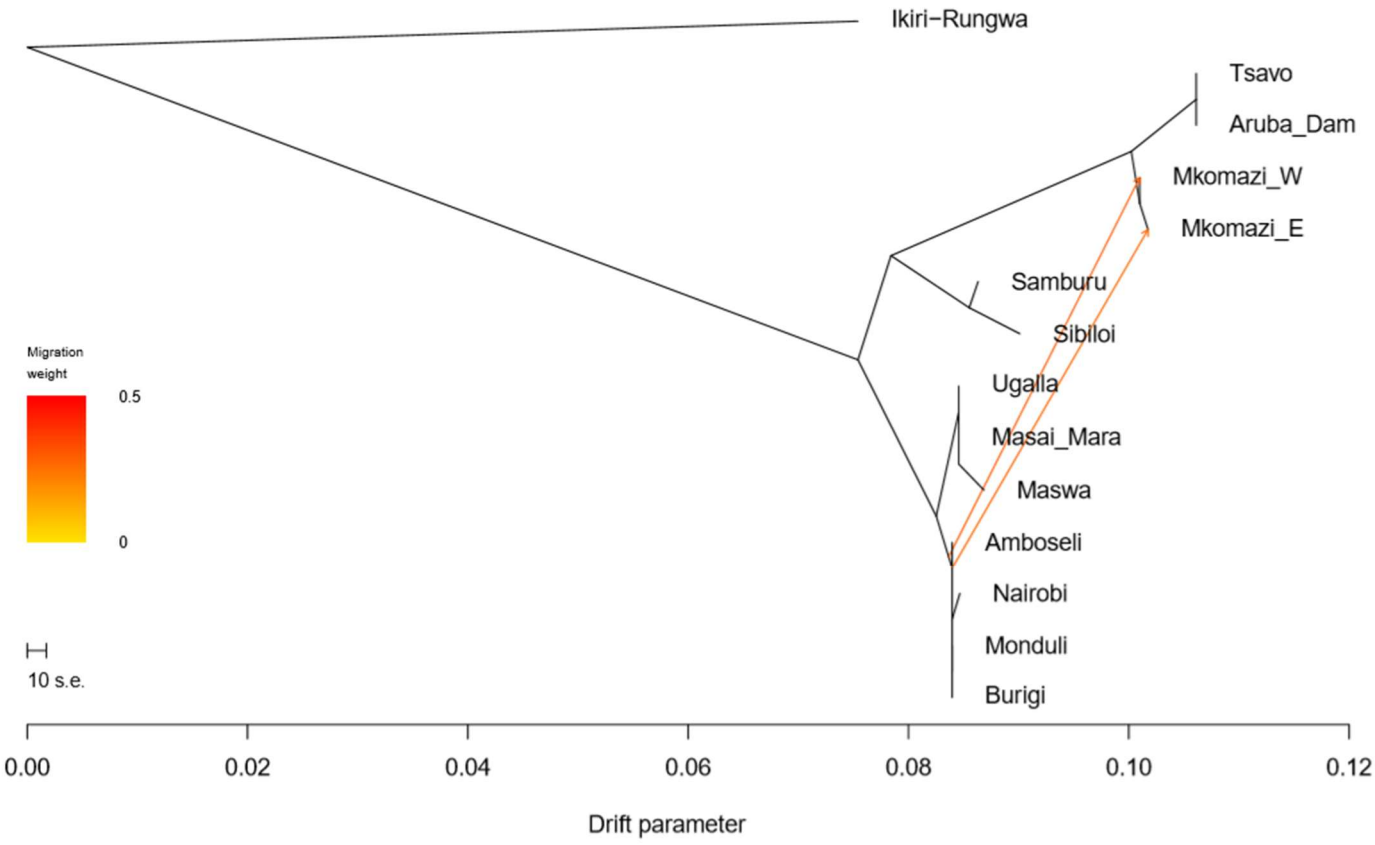


Figure 10. Maximum likelihood drift tree generated by TreeMix [3]. Two migration events are allowed, and data is clustered by locations.

4. Discussion

Based on our results, we can confirm the current taxonomical classification that considers Grant's gazelle complex as a superspecies comprising the three distinct and allopatric species, *granti*, *petersii* and *notata* [1]. This is supported by previous studies [1] on mtDNA which showed that, despite they live in geographic proximity with no apparent geographical or ecological barriers, the three lineages reflected distinct evolutionary trajectories. It has been suggested by Siegismund et al. (2013) [4] that the genetic differentiation of the species may have evolved during the late Pleistocene era, where the populations may have been separated due to repeated contraction and expansion of habitat. The lack of interbreeding between the three species, that has led to their differentiation, could have been caused by chromosomal differences which contributed to reproductive barriers [23].

Also, from our results and particularly from genetic admixture and TreeMix analysis, we confirm the subdivision of the *granti* species as *g. granti* and *g. robertsii* subspecies. This has been proposed by Lorenzen et al. (2007) [1], but their genetic data were inconclusive and they did not consider the subspecies well authenticated.

Moreover, our results raised the necessity to question whether the current taxonomical classification of the Mkomazi population into the *g. granti* subspecies is correct. It is reasonable to think that the Mkomazi population is the result of an admixture event between *g. granti* and *petersii*. This is also supported in the literature, Grubb (1994) [24] described an intermediary coat coloring of specimens collected from the borderlands between Kenya and Tanzania, and he proposed that hybridization had occurred between *granti* and *petersii*. During recent decades, they may have come into close contact thanks to the thinning of the Acacia-Cammiphora woodland [4], which, in the nineteenth century, was a dense vegetation barrier to gene flow between these two distinct populations [1].

We showed evidence that Mkomazi population is genetically more similar to *petersii* than *granti* species, and for this reason, we propose that it should be placed within *petersii* taxonomic lineage. This arises important consequences in terms of conservation management of these populations. Grant's gazelle populations have decreased considerably over the last decades and this had a different impact on the different species [4]. *Granti* species live in well protected areas, its population size probably exceeds 75000 [4] and it has a large genetic diversity (Table 2). *Notata* species has a lower proportion of its populations in protected areas but it has an estimated population size of 50000 animals, and it also has a large genetic diversity (Table 2). *Petersii* has a population size that is probably less than 15000, Tsavo National Park and Tana River National Reserve are the only protected area where it is distributed [4], and it has the lowest genetic variability within Grant's gazelle complex (Table 2). For all these reasons it might be more threatened than the other two species [4], and so we propose that the conservation status of the three species, and in particular that of the Mkomazi population, is re-evaluated considering the latter as a member of the *petersii* species.

5. Acknowledgments

Martina Cardinali, Jean-Baptiste Michel P Van Den Broucke, Alexander Henrik Welford, Charlie Stender Cordes are thanked for their collaboration in the analysis of the data. Genís Garcia-Erill and Hans Siegismund are thanked for providing guidance and technical assistance. Casper-Emil Pedersen, Peter Frandsen and Genís Garcia-Erill for providing the basic structure and guidelines for the laboration of the code. The study was supported by Ida Moltke, Hans Siegismund, Rasmus Heller, Anders Albrechtsen and the other members of the Population Genetics team 2020 of the University of Copenhagen.

6. Reference

[1] Eline D. Lorenzen, Peter Arctander, Hans R. Siegismund (2007). Three reciprocally monophyletic mtDNA lineages elucidate the taxonomic status of Grant's gazelles. Conserv Genet 9:593–601.

[2] Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. (2013). Special features of RAD Sequencing data: implications for genotyping. Molecular Ecology 22: 3151–3164.

[3] Keinan, A., Mullikin, J. C., Patterson, N., and Reich, D. (2007). Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat Genet, 39(10):1251–5.

[4] Hans R. Siegismund, Eline D. Lorenzen e Peter Arctander (2013): Nanger (granti) Grant's Gazelle Species Group. In: Jonathan Kingdon, David Happold, Michael Hoffmann, Thomas Butynski, Meredith Happold e Jan Kalina: Mammals of Africa Volume VI, pp. 373-379.

[5] Kingdon J (1982). East African mammals. An atlas of evolution in Africa, vol III, Part D (Bovids). The University of Chicago Press, pp 414–421.

[6] Grubb P (2005). Nanger granti. Mammal species of the world a taxonomic and geographic reference. Johns Hopkins University Press, p 684.

[7] Leuthold W. (1981). Contact between formerly allopatric subspecies of Grant's gazelle (Gazella granti Brooke, 1872) owing to vegetation changes in Tsavo National Park, Kenya. Z Säugetierkd 46:48–55.

[8] IUCN SSC Antelope Specialist Group 2016. Nanger granti. The IUCN Red List of Threatened Species 2016: e.T8971A50186774.

[9] Lacy RC (1997). Importance of genetic variation to the viability of mammalian populations. J. Mammal. 78:320–35.

[10] Sadalmelik I. (2007). Own work released as public domain in Wikimedia Commons. Topographic map of Tanzania and Kenya. Created with GMT from public domain GLOBE data.

[11] Jonathan P. Weeks (2010). plink: An R Package for Linking Mixed-Format Tests Using IRT-Based Methods. Journal of Statistical Software, 35(12), 1-33.

[12] Graham Coop, (2020). Population and quantitative genetics. Department of Evolution and Ecology & Center for Population Biology, University of California, Davis.

[13] Jensen-Seaman MI, Chemnick L, Ryder O, Li WH (2004). Nucleotide diversity in gorillas. Genetics. 166: 1375-83.

[14] Mattias Jakobsson, Michael D. Edge, Noah A. Rosenberg (2013). The relationship between Fst and the frequency of the most frequent allele. Genetics. 193: 512-528

[15] Holsinger, Kent & Weir, Bruce (2009). Genetics in geographically structured populations: defining, estimating and interpreting Fst. Genetics.

[16] Nielsen, Slatkin. (2013). An introduction to population genetics: theory and applications. Sinauer Associates, pp 61.

[17] Weir BS, Cockerham CC. (1984). Estimating F-Statistics for the Analysis of Population-Structure. Evolution 38: 1358-1370.

[18] Chen G, Yuan A, Shriner D, Tekola-Ayele F, Zhou J, et al. (2015) An Improved Fst Estimator. Plus one 10(8): e0135368.

[19] Alexander, D.H., Lange, K. (2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12**,** 246).

[20] Mejia AF, Nebel MB, Eloyan A, Caffo B, Lindquist MA (2017). PCA leverage: outlier detection for high-dimensional functional magnetic resonance imaging data.

[21] Kindt, Roeland & Breugel, Paulo & Lillesø, Jens-Peter Barnekow & Bingham, M & Demissew, Sebsebe & Dudley, C & Friis, Ib & Gachathi, F & Kalema, James & Mbago, Frank & others,. (2011). Potential natural vegetation of Eastern Africa (Ethiopia, Kenya, Malawi, Rwanda, Tanzania, Uganda and Zambia): Volume 3.

[22] Pickrell JK, Pritchard JK (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8(11):e1002967.

[23] Peter Arctander, Pieter W. Kat, Rashid A. Aman & Hans R. Siegismund (1995). Extreme genetic differences among populations of Gazella granti, Grant's gazelle, in Kenya.

[24] Grubb P (1994) Genetic analyses of African bovids. Gnusletter 13(1–2):4–5

7. Appendices

7.1. Additional figures

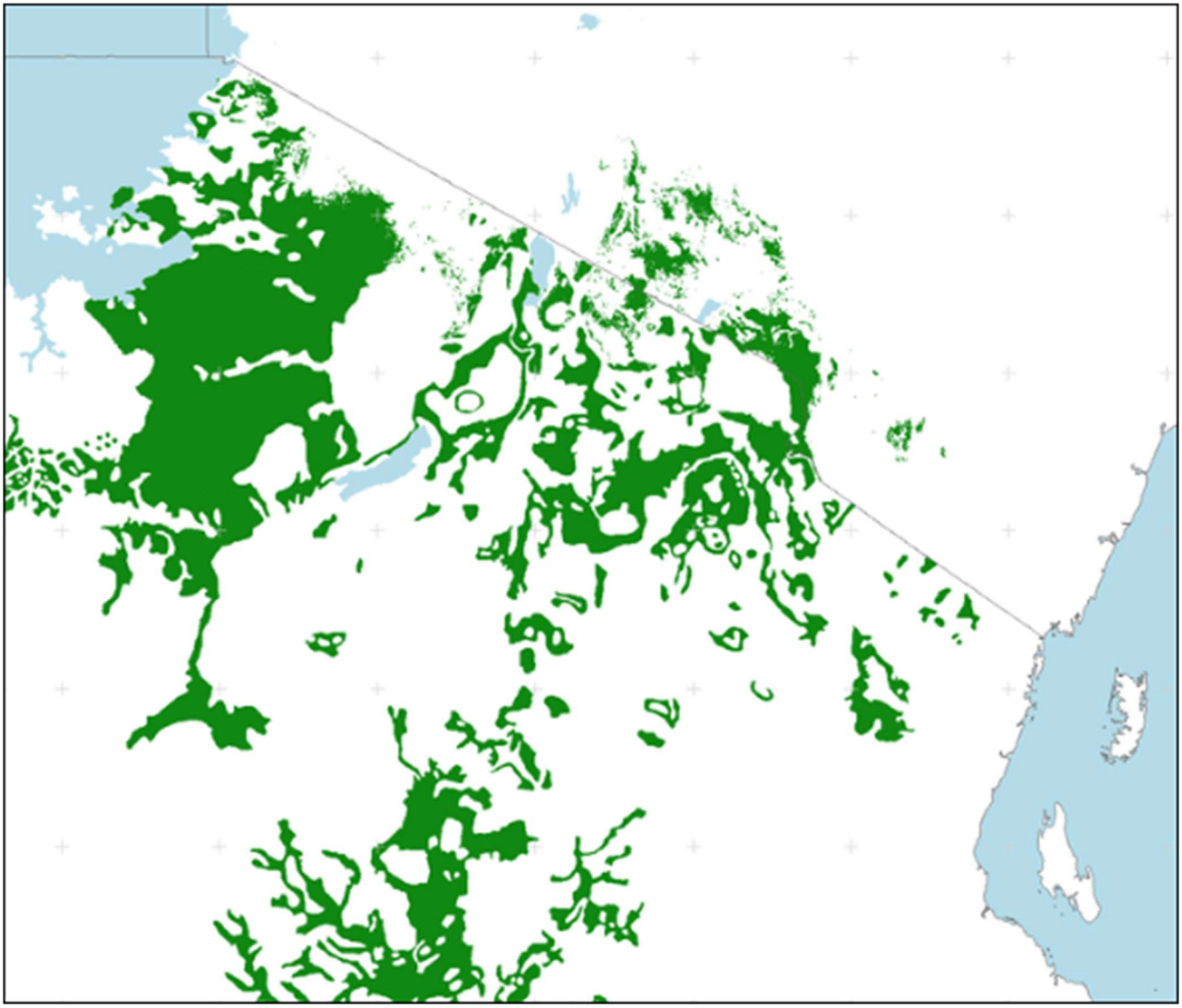


Figure 2. Distribution of Acacia-Cammiphora deciduous wooded grassland in Tanzania and Kenya (2011) [20].

7.2. Scripts

1. Principal Component Analysis
2. Fixation index
3. Genetic admixture
4. Nucleotide diversity
5. TreeMix