Pleistocene climate fluctuations drove demographic history of African golden wolves (*Canis lupaster*)
Pleistocene climate change impacted entire ecosystems throughout the world. In the northern hemisphere, the distribution of Arctic species expanded during glacial periods, while more temperate and mesic species contracted into climatic refugia, where isolation drove genetic divergence. Cycles of local cooling and warming in the Sahara region of northern Africa caused repeated contractions and expansions of savannah-like environments which connected mesic species isolated in refugia during interglacial times, possibly driving population expansions and contractions; divergence and geneflow in the associated fauna.

Here we use whole genome sequences of African golden wolves (*Canis lupaster*), a generalist mesopredator with a wide distribution in northern Africa to estimate their demographic history and past episodes of geneflow. We detect a correlation between divergence times and cycles of increased aridity-associated Pleistocene glacial cycles. A complex demographic history with responses to local climate change in different lineages was found, including a relict lineage north of the High Atlas Mountains of Morocco that has been isolated for more than 18,000 years, possibly a distinct ecotype.

Keywords: carnivore, genomics, MiSTI, PSMC, *Canis anthus*
INTRODUCTION

Pleistocene climatic fluctuations shaped phylogeographic and demographic histories of many species in the northern hemisphere (Bolfíková et al., 2017; Feliner, 2011; Gómez & Lunt, 2007; Hewitt, 2000; Hewitt, 1999; Tison et al., 2014), which was periodically glaciated (Clark et al., 2009). The patterns left from repeated range shifts into glacial refugia and subsequent expansion across the higher latitudes has been best characterized in Europe (Hewitt, 2000; Hewitt, 1999). Species adapted to temperate climates saw a reduction in range and numbers of individuals during glacial periods in European and North American ecosystems, and benefited from milder climates during interglacials (Bolfíková et al., 2017; Dufresnes et al., 2020; Gómez & Lunt, 2007; Sommer & Nadachowski, 2006; Stóck et al., 2012). In tropical regions a combination of changes in incoming solar radiation and glacial-interglacial cycles influenced the extension and activity of the monsoon systems, and resulted in shifts between humid and arid conditions (Drake, Blench, Armitage, Bristow, & White, 2011; Ehrmann, Schmiedl, Beuscher, & Krüger, 2017; Emeis, Såkamoto, Wehausen, & Brumsack, 2000; Heinrich, 1988; Hoffmann et al., 2016; Larrasoaña, Roberts, & Rohling, 2013; Lézine, Hély, Grenier, Braconnot, & Krinner, 2011; Rohling, Mayewski, & Challenor, 2003; Smith, 2012). A pattern of expansion of savannah associated north African species was observed during humid periods (Bertola et al., 2016; Cosson et al., 2005; Dinis et al., 2019; Husemann, Schmitt, Zachos, Ulrich, & Habel, 2014; Iyengar et al., 2007; Leite et al., 2015; Lerp, Wronski, Pfenninger, & Plath, 2011). Desert adapted species show the opposite pattern of expansion during dry periods and reductions in distribution and population size in humid periods (Moutinho et al., 2020; Tamar et al., 2018).

The northern coast of north Africa is currently dominated by a temperate Mediterranean climate, which brings westerly rains that hardly penetrate 200 km away from coastal areas (Larrasoaña et al., 2013). Further south, the climate of tropical north Africa is driven by two monsoonal systems - the west and east African monsoons - that result in higher rainfall in the equatorial zone that decreases progressively towards the north. Although westerly- and monsoon-driven precipitation extend into the continental interior during extreme events, the overall scarcity of rainfall dictates the presence of the Sahara Desert, a hyperarid zone between 28-30 °N and 18 °N where semi-arid or savannah fauna live isolated in refugia around the periphery and in oases (Dinis et al., 2019; Husemann et al., 2014; Nicolas, Granjon, Duplantier, Cruaud, & Dobigny, 2009; Rato et al., 2007)(Fig. 1). Cyclic variations in the Earth’s orbit have driven changes in the amount of solar radiation received in the tropics, which is the engine that powers the monsoon system and has resulted in periodical expansion of savannah landscapes throughout much of the Sahara Desert since its inception about 14 million years ago (Ma) (Castañeda et al., 2009; Drake et al., 2008; Drake et al., 2011; Drake, Breeze, & Parker, 2013; DeMenocal, 2004; Ehrmann et al., 2017; Geyh & Thiedig, 2008; Larrasoaña et al., 2013; Smith, 2012; Tjallingii et al., 2008; Weldeab, Lea, Schneider, & Andersen, 2007). Such expansions, known as Green Sahara Periods (GSP), typically lasted for 3 to 6 thousand years (kyr) and occurred during bundles of high-amplitude boreal summer insolation maxima driven by maxima in the

This article is protected by copyright. All rights reserved
eccentricity of the Earth’s orbit. During periods of eccentricity minima, when boreal summer insolation peaks were subdued, the monsoon system was weakened and influenced by glacial-boundary conditions via changes in north Atlantic sea-surface temperatures (DeMenocal, 2004; Ehrmann et al., 2017; Tjallingii et al., 2008; Weldeab et al., 2007). Intensification of glacial cycles in the northern hemisphere after 2.7 Ma amplified the effect of glacial-boundary conditions and led to an overall shift to drier conditions over the Sahara (DeMenocal, 2004). Glacial periods occurred with a 100-kyr cyclicity after ca. 1.2 Ma (McClymont, Sosdian, Rosell-Melé, & Rosenthal, 2013) and, at least in the case of the last glacial period, were punctuated by short, colder periods (Heinrich stadials) that that lasted up to 1500 years (Heinrich, 1988; Hemming, 2004) and enhanced the dry conditions over the desert (Ehrmann et al., 2017). If animals are tightly associated with their habitat, their populations should expand and contract with their preferred habitat, so in GSPs semi-arid species should expand in population size and gene flow should increase between regions isolated during drier, glacial times (Dinis et al., 2019; Karssene, Nowak, Chammem, Cocchiaro, & Nouira, 2019; Leite et al., 2015; Lerp et al., 2011; Rato et al., 2007). Short-lived epochs of enhanced rainfall also occurred during the last glacial period along the northernmost (Drake et al., 2013; Hoffmann et al., 2016; Smith, 2012) and southernmost (Castañeda et al., 2009; Weldeab et al., 2007) fringes of the Sahara, and might have also driven changes in population size and gene flow in north African species.

The African golden wolf (Canis lupaster) is a recently re-discovered (Koepfli et al., 2015) wild canid species of north and east Africa. Its current distribution encompasses approximately 11 million square kilometers from Morocco to Egypt in the north and from Senegal to Kenya in the south (Kebede, Sciences, Box, & Sodo, 2017; Kingdon, 2013) (Fig. 1). This large distribution makes them useful for answering questions about how past climate periods could have affected mesic mammal communities in north Africa. Until recently, African golden wolves were considered the same species as the Eurasian golden jackal (Canis aureus), and most ecological data have been collected in Eurasia, making African golden wolves one of the least studied canid species in the world (Admasu, Thirgood, Bekele, & Laurenson, 2004; Amroun, Bensidhoum, Delattre, & Gaubert, 2014; Amroun, Giraudoux, & Delattre, 2006; McShane & Grettenberger, 1984). Their wide distribution and generalist predatory style has allowed them to adapt to a wide variety of habitats including arid or semi-arid landscapes, grasslands, savannahs, forests and high elevation areas in Morocco and Ethiopia as well as anthropized zones (Amroun, Oubellil, & Gaubert, 2014; Amroun et al., 2006; Cuzin, 2003; Kebede et al., 2017; McShane & Grettenberger, 1984).

Despite their large home ranges and ecological plasticity (Admasu et al., 2004; Amroun et al., 2014, 2006; Fuller, Biknevicius, Kat, Van Valkenburgh, & Wayne, 1989; Karssene et al., 2018; McShane & Grettenberger, 1984; Moehlman, 1986), African golden wolves have not been found in hyperarid areas (Kingdon, 2013) although they have been reported in archaeological sites in current-day desert areas that

This article is protected by copyright. All rights reserved
were greener in the past (Sereno et al., 2008). This suggests that African golden wolves had a wider
distribution during GSP. In addition to climate, competition between African golden wolves and bigger
carnivores with overlapping distributions, such as hyenas (Kebede et al., 2017; Kingdon, 2013) and black-
backed jackals (Van Valkenburgh & Wayne, 1994), could have affected their distribution. Recent genetic
evidence has suggested that African golden wolves may have benefited from increases in human
populations since the Neolithic, either through the introduction of caprid livestock that could have served as
a food source, and/or through predator control against competing species (Eddine, Mostefai, et al., 2020).
African golden wolves have been shown to scavenge around anthropized zones (Admasu et al., 2004;
Amroun et al., 2014; Amroun et al., 2006; McShane & Grettenberger, 1984), but have disappeared from
very intensively exploited agricultural areas (Aulagnier, 1992; Cuzin, 2003). Finally, recent evidence from
whole genomes of African golden wolves (Chavez et al., 2019; Gopalakrishnan et al., 2018; Liu et al.,
2018) have shown a complex genetic history with two separate populations at the extremes of their
distribution and possibly events of introgression from other related canids in the past.

Here we use seven whole genome sequences from across the distribution of African golden wolves
and recently developed analytical methods to evaluate population structure, deep demographic history, and
past episodes of gene flow. We integrate the genomic results with past environmental variability in north
Africa (Castañeda et al., 2009; Drake et al., 2008; Drake et al., 2011; Drake et al., 2013; Geyh & Thiedig,
2008; Larrasoaña et al., 2013; Tjallingii et al., 2008; Weldeab et al., 2007) in order to evaluate the role of
monsoon variability and its teleconnection with glacial-interglacial climates in shaping the demography of
the African golden wolf. We also compare the genomic diversity, structure, and inbreeding coefficients
across the species with similar geographic distributions in two closely related canids (gray wolves, *Canis
lupus*; and coyotes, *C. latrans*) with very different social systems and ecology. We find a complex
demographic history with different, well defined populations that diverged thousands of generations ago
coinciding with Late Pleistocene climatic fluctuations, and an isolated mountain lineage with a high degree
of inbreeding whose genome-wide heterozygosity has been drastically reduced. Although the morphology
and size of the African golden wolves is more similar to North American coyotes, the observed genetic
population structure is more similar to what is found in gray wolves, possibly suggesting a social system
and ecology more similar to the latter.

METHODS

Materials

A sample from a previous study (Urios, Donat-Torres, Monroy-Vilchis, & Idrissi, 2015) from
which the mitochondrial genome has already been published (KT378605) was used to construct a shotgun
library as in Camacho-Sanchez et al., (2018) and sequenced. The individual was a roadkill found at
N32°33.364’ W5°50.848’, in Beni Mellal province, Morocco. The area is a hill slope at around 2000 m asl
elevation in the north of the High Atlas Mountains. We refer to this sample as “west Morocco” and a previously published genome (Gopalakrishnan et al., 2018) from another Moroccan individual as “east Morocco”. An additional 26 genomes were obtained from the literature: six African golden wolves, seven domestic dogs, six gray wolves, two Eurasian golden jackals one Ethiopian wolf, one African hunting dog, and three coyotes (Table S1).

Pre-processing pipeline

Adapters were trimmed with cutadapt (Martin, 2011) and quality of the reads was evaluated with FastQC (Andrews, 2010). Reads were mapped using bwa mem v1.3 (Li & Durbin, 2010) to the reference genome CanFam3.1 (Canis familiaris, domestic dog) (Lindblad-Toh et al., 2005), with the Y chromosome (Oetjens, Martin, Veeramah, & Kidd, 2018). Reads were also mapped to an assembled reference genome of an African hunting dog (Lycaon pictus; Campana et al., 2016) to avoid bias from mapping to the dog genome, which is in the ingroup for some analyses. We sorted and filtered low quality mapped reads with samtools v1.9 (Li, 2011) and removed duplicates with GATK v3.7 (McKenna et al., 2010). All reads had a sequencing quality higher than 20 and had a complementary read in the same chromosome (Table S1).

Variant calling and quality filters

We filtered reads mapping to sex chromosomes and mitochondrial DNA and retained only autosomes. We filtered low quality mapped reads and called genotype likelihoods using ANGSD. Sites with depth <5X or more than twice the mean coverage depth were filtered out (Freedman et al., 2014). Genotype likelihood frequency files (.glf.gz) were used to call SNPs using the -doplink option in ANGSD and an exclusion zone of 10 kilobases (kb) upstream and downstream the genic regions was defined using a refseq annotation file of CanFam3 from the UCSC Genome browser (Kent et al., 2002) and custom scripts (Supplementary Code File 02.b). Genotype likelihoods were computed and SNPs called for autosomes mapped against both reference genomes (CanFam3.1 and African hunting dog).

Genetic structure

A principal component analysis (PCA) was performed using genotype posterior probabilities of genotypes of the five Old World species with the ngsCovar package in ngsTools (Fumagalli, Vieira, Linderoth, & Nielsen, 2014) and Rscript v3.4.4 (The R Core Development Team, 2017). We estimated admixture proportions from genotype likelihoods using NGSadmix (Skotte & Albrechtsen, 2013), setting number of clusters (K) between 5 and 14. To ensure reproducibility of results, we ran the NGSadmix test five times and visually compared the plots.

We used .glf.gz files to call SNPs with ANGSD (-doplink option, settings: -pvalue 1e-5, -doMaf 1, -doPost 1). Genic regions, deviations from Hardy-Weinberg (HW) equilibrium, and SNPs in linkage disequilibrium were filtered from the dataset in PLINK v1.9 (Purcell et al., 2007) (settings: --hwe 0.001, --maf 0.05, --indep-pairwise 50 5 0.5). We performed a PCA using flashPCA (Abraham & Inouye, 2014) and
estimated admixture composition using ADMIXTURE (Alexander, Novembre, & Lange, 2009) with a number of clusters (K) between 5 and 14. Visual comparisons were made of plots with genotype likelihoods and with discovered SNPs in genomes mapped against CanFam3.1 and the African hunting dog genome.

Demographic history

We down sampled .bam files of the Kenyan African golden wolf (AGW) to 7X, 9X, 11.2X, and 15X from 24X using samtools view -bs (Li, 2011) to visually estimate the best false negative rate (FNR) due to low coverage following (Hawkins et al., 2018; Kim et al., 2014; Nadachowska-Brzyska, Burri, Smeds, & Ellegren, 2016) and the different coverages of the Kenyan genome were plotted with and without FNR correction (Fig. S1A,B).

To avoid misinterpretations of heterozygous sites as homozygous in low coverage samples (e.g., 7X-15X) (Nadachowska-Brzyska, Burri, Smeds, & Ellegren, 2016), we repeated the process with the actual coverage and downsampled .bam files of the Kenyan African golden wolf (7X, 9X, 11.12X, 15X and 24X) to visually estimate the best false negative rate (FNR) following (Hawkins et al., 2018; Kim et al., 2014; Nadachowska-Brzyska et al., 2016). After determining the FNR, consensus sequences with genic regions were called with bcftools mpileup (settings: -C 50, d 5, -D 100). PSMC was called using 64 atomic time intervals (settings: -p “1*6+58*1”) as in previous studies with AGW (Freedman et al., 2014), and initial theta per individual and coverage was corrected (Li & Durbin, 2011). A round of 50 iterations of bootstrapping per genome was applied to draw a multisample PSMC plot. A mutation rate of 4.5 *10^-9 (Koch et al., 2019) and generation time of 3 years (Chavez et al., 2019; Freedman et al., 2014; Gopalakrishnan et al., 2018; Koepfli et al., 2015; Liu et al., 2018) were used. To explore extreme mutation rates (2.7-7.1*10^-9 as estimated by Koch et al. 2019), we also generated PSMC plots for all AGWs except Egypt without bootstrapping with these values.

More recent demographic population history was studied with ngsPSMC (Shchur et al., 2019). We generated files for ngsPSMC in ANGSD (option: -dopsmc) filtering for a minimum depth of 5X per genome. We ran ngsPSMC for 50 iterations with the same atomic time intervals as in PSMC, using initial population sizes 10^5 years ago as observed in the PSMC plot. Calculated \( \Theta \pi \) was used as theta (\( \theta \)) (see Supplementary Code File 03) and we estimated genome wide rho from a recombination map for dogs (Auton et al., 2013). Mutation rate and generation time were the same as in PSMC. NgsPSMC is still under development, so bootstrapped plots could not be generated.

We considered both low- and high-latitude climate mechanisms influencing past environmental variability in the Sahara back to 1.5 million years ago (Castañeda et al., 2009; Drake et al., 2013; Ehrmann et al., 2017; Larrasoña et al., 2013; McClymont et al., 2013; Rohling et al., 2003; Smith, 2012; Tjallingii

This article is protected by copyright. All rights reserved
et al., 2008; Weldeab et al., 2007) that could have affected the demographic history of AGW. We compared the timing of these events with the PSMC and ngsPSMC maxima and minima and observed possible correlations between climatic events and increases or decreases of population. We also added the speciation event that led to African golden wolves (Chavez et al., 2019; Koepfli et al., 2015) with a confidence interval of 400 kyr.

213 **Summary statistics**

214 We inferred changes in Ne with thetas and neutrality tests based on likelihood-based estimation of site frequency spectrum (SFS) in ANGSD (Nielsen, Korneliussen, Albrechtsen, Li, & Wang, 2012). The reference genome of the African hunting dog (Campana et al., 2016) was used as ancestral to call unfolded SFS, assuming HW equilibrium, multisample GL estimation (-dosaf 1), and an upper depth filter of 2.5X, mean read depth per sample and per population as in Table S1. We calculated the genome wide heterozygosity per individual as in Gopalakrishnan et al. (2018). Different coverages of the Kenyan AGW (7X, 9X, 11.2X, 15X, 24X, as above) were used to repeat SFS calling and correct mean heterozygosity for all samples.

222 We used a joint SFS between pairs of populations (2DSFS) to estimate average genome wide Fst and 95% confidence intervals from a 50 kb sliding window scan. Finally, we computed a series of nucleotide diversity indexes (Tajima’s D (Tajima, 1989), Fu and Li’s F and D (Fu & Li, 1993), Fay’s H (Zeng, Fu, Shi, & Wu, 2006), Zeng’s E (Zeng et al., 2006)) and thetas (θW, θπ, θFL, θH, θL) (Durrett, 2008; Fay & Wu, 2000; Fu & Li, 1993; Tajima, 1989; Watterson, 1975; Zeng et al., 2006) using the -doThetas 1 option in ANGSD with population-based SFS as prior information (-pest), divided in 50-kb windows across the genome and excluding genic regions.

229 **Heterozygosity**

230 We defined four populations: afr_north (AGW from Algeria, Egypt, east Morocco, west Morocco, Senegal), afr_east (AGW from Ethiopia, Kenya), coyote (coyotes from California, Midwest and Mexico) and gwolf_me (gray wolves from S. Arabia, Iran and Syria). We calculated the inbreeding coefficient or F<sub>i</sub> per individual based on individual genotype likelihoods using ngsF (Vieira, Fumagalli, Albrechtsen, & Nielsen, 2013), computed across 20 iterations in ngsF. Genotype-based calculation of F<sub>i</sub> per individual was performed in PLINK with –het on a dataset of SNPs discovered using ANGSD (option -doPlink).

236 We used two approaches to calculate ROHs across the whole genome. The first method uses PLINK and the SNP dataset from the F<sub>i</sub> calculation. In each population we removed SNPs in close linkage disequilibrium in 200 basepair (bp) windows with a step size of 100 bp and a R<sup>2</sup> of 0.9 using option –indep-pairwise 200 100 0.90 in PLINK and generated ROHs as in Sams & Boyko (2019). The second method uses the software ROHan (Renaud, Hanghøj, Korneliussen, Willerslev, & Orlando, 2019) on mapped .bam
files. We ran ROHAn in windows of 500kb in .bam files of all AGW, coyotes and gray wolves of the Middle East. Expected theta in ROHs was set to $2 \times 10^{-5}$ and default options for Illumina error rate. Plots of local heterozygosity were computed across the whole genome and a summary of ROHs was calculated. Finally, we calculated inbreeding coefficients from ROHs ($F_{ROH}$) as in (McQuillan et al., 2008; Sams & Boyko, 2019):

$$F_{ROHj} = \frac{\sum_{k} length(ROH_k)}{L},$$

where ROH$_k$ is the $k$th ROH in individual $j$’s genome and $L$ is the total length of the genome.

### Divergence dating

We used MiSTI (Shchur, 2019) to estimate times of divergence between local lineages represented by our seven AGW individuals. We used a list of green Sahara periods (Ehrmann et al., 2017; Larrañaga et al., 2013) and of cold stadials (Heinrich, 1988; Hemming, 2004; Rohling et al., 2003) associated with increased aridity of the Sahara region (Ehrmann et al., 2017) to define humid (with potentially more connectivity among lineages) and dry (potential times for divergence) time segments. We used pairwise time scales generated using PSMC and 2DSFS files from previous sections. GNU Parallel (Tange, 2018) was used to model simultaneously different scenarios of divergence among lineages with different migration rates in dry periods and GSPs, permitting MiSTI to automatically optimize calculations of migration rate per time segment. We extracted a table of splitting times from MiSTI and plotted log likelihoods per proposed splitting time against time. Finally, a polynomial curve was fitted per group of data where $R^2 \geq 0.99$ to estimate the maximum point of the curve using the Newton-Raphson approach and a confidence interval of the upper 5%, 1% and 0.1% of log likelihood points. Since this is a new software, we evaluated the replicability of our results by testing divergence time estimations with different coverages of the same genomes and different combinations of pre-defined time segments. We also compared the results with an estimation of divergence time using the Cavalli-Sforza et al. (1969) equation (Supplementary Methods).

### RESULTS

Although coverage was different between samples, for each sample coverage across the genome was similar when reads were mapped against the two reference genomes, CanFam3.1 and the African hunting dog. Coverage for samples mapped against CanFam3.1 ranged from 5.16X to 34.54X and from 4.58X to 32.1X when mapped to the African hunting dog reference genome (Campana et al., 2016)(Tables S2 and S3). Mapping against CanFam3.1 resulted in a slightly higher mappability, higher coverage and less duplicates than against the African hunting dog genome (Table S4). The new sample, west Morocco, had a
total of 284,332,735 raw reads. Average mappability for this sample against CanFam3.1 and African hunting dog reference genomes was 93.99% and 91.55%, respectively. Genomewide coverage was 11.26X mapped against CanFam3.1 and 10.58X mapped against African hunting dog (Tables S2 and S3).

Population structure driven by a north-south boundary

Genotype probabilities were calculated across 23 genomes of five species (7 African golden wolves (AGW), 7 dogs, 6 gray wolves, 1 Ethiopian wolf, 2 Eurasian golden jackals) with the -doGeno 32 option in ANGSD. Nearly three million (2.98M) sites were called when mapping against the CanFam3.1 genome, and 2.54M sites when mapping against the African hunting dog. Genotype likelihood-based SNP calling with ANGSD produced 689,000 (698K) sites for those genomes mapped against CanFam3.1, and 625K sites for genomes mapped against the African hunting dog. All SNPs were more than 10kb away from any genic zone and were in HW equilibrium. We identified population structure, with all wolves northwest (Senegal, east Morocco, west Morocco, Algeria) and southeast of the Sahara (Kenyan and Ethiopian) in different clusters (Fig. S2). The Egyptian AGW appeared at an intermediate genetic distance between the northwestern cluster and a cluster that encompassed all gray wolves and dogs, which is consistent with previously published studies that suggested it has introgressed ancestry (Gopalakrishnan et al., 2018; Liu et al., 2018). While different PCAs showed Ethiopian wolves at different positions with respect to the African golden wolves depending on the reference genome and the source data (glf.gz or SNPs), the structure of African golden wolves, dogs, and gray wolves remained consistent regardless of the reference genome used to map reads against (CanFam3.1 or African hunting dog) or what method was used to generate the PCA plots (genotype likelihoods or SNPs) (Fig. S3).

After filtering for genic sites plus 10kb windows, NGSadmix retained 10.5M genotype likelihood sites for admixture analyses in genomes mapped to CanFam3.1 and 9.89M sites for genomes mapped against the African hunting dog. Using the full dataset of 23 genomes, NGSadmix failed to identify a most probable value of K based on the maximum likelihood estimation implemented in the program (Skotte & Albrechtsen, 2013), with high numbers of K receiving the highest likelihood. We extracted and plotted in R the best likelihood values for every K and attempted to find the best-fit K, but the best likelihood was found at very large K values, even beyond the number of individuals included in the analysis (Fig. S4). A relative maximum was found at K=21. This issue may indicate sub-clustering within the species studied (Pilot et al., 2019), and does not necessarily mean lack of population structure. Nonetheless, a series of general trends could be identified. Starting from K=6 onwards, the Ethiopian and Kenyan African golden wolves clustered together in a different population from the other AGWs. From K=5 to K=10, the Egyptian AGW appears to have mixed ancestry between African golden wolves and possibly Middle Eastern gray wolves or dogs from Africa (Fig. 2A). Notably, at K=9 the Ethiopian AGW seems to share some proportion of

This article is protected by copyright. All rights reserved
alleles from a postulated ancestral population with Ethiopian wolves, but this trend is not shown at any other K or level of clustering.

When considering African golden wolves alone (Fig. 2B), some structure is detected between the southeast population (Ethiopia, Kenya) and the northern population (Algeria, Egypt, Morocco, Senegal). This pattern is consistent with a general trend in big African carnivores and ungulates to cluster in northern and southeast populations divided by the Rift Valley (Bertola et al., 2016; Brown et al., 2007; Charruau et al., 2011; Flagstad, Syvertsen, Stenseth, & Jakobsen, 2001; Lorenzen, Heller, & Siegismund, 2012; Moodley & Bruford, 2007; Muwanika, Nyakaana, Siegismund, & Arctander, 2003; Smits et al., 2013), and is also consistent with previous studies on African golden wolves (Gopalakrishnan et al., 2018). No mixed composition was detected in the Egyptian individual from other different African golden wolves when no dogs or gray wolves were present at the NGSadmix study. SNP-based Admixture plots showed very similar results, and only small local proportions of admixture between dog and gray wolf were detected when comparing NGSadmix and Admixture plots generated for mapped genomes against CanFam3.1 and African hunting dog (Fig. S5). Since we did not detect major reference biases in African golden wolf genomes, all subsequent analyses were performed using autosomes mapped against CanFam3.1.

Complex and heterogeneous demographic history

With PSMC, three possible groups of lineages with similar demographic histories could be seen: one for eastern AGW from Kenya and Ethiopia, another for the Senegalese and Algerian individuals, and another one for the two Moroccan individuals (Fig. 3). The Egyptian AGW showed a very steep increase in effective population size (Ne) towards the end of the graph that could be interpreted as introgression and could not be placed with the other samples at the plot.

While the Moroccan lineages could have benefited from a series of GSP identified at 124 kyr (Eemian), ca. 100 kyr and 80 kyr ago (Ehrmann et al., 2017; Larrasoña et al., 2013), the eastern group showed a relatively constant decrease in population since 300 kyr ago, reaching a minimum Ne of ca. 40,000 at around 45 kyr ago, after which animals from west and east of the Rift Valley follow different demographic trajectories (Fig. 3). Both Moroccan lineages seem to have followed a similar trajectory until about 28 kyr ago, after which the west Moroccan lineage has a steep reduction in population effective size, while the east Moroccan lineage remains constant. The Senegalese and Algerian individuals shared a demographic history until about 100 kyr ago, when they diverge, with a constant decrease in population for the Algerian individual and a very steep decrease for the Senegalese lineage after 24 kyr ago. Some proposed events of enhanced dry conditions in north Africa driven by Heinrich stadials could have caused decreases in these populations. Heinrich Event 6 (58.25-58.85 kyr ago; Rohling et al., 2003) appears to coincide with a reduction of population size in the Moroccan lineages, while all populations experienced a reduction after Heinrich Event 2 (ca. 23.6-25.9 kyr ago; Rohling et al., 2003) (Fig. 3). However, the
resolution for this time is not sufficient to confidently associate climate events with population contractions. Pleistocene climate changes in north Africa did not impact the African golden wolves equally or synchronously across their range.

After 70 kyr ago all lineages, including the Egyptian individual, show an increase in population size according to ngsPSMC (Fig. 4A), although at different times. The first lineages to experience this increase are found in the western part of the range, and later in the east when the Kenyan lineage reached an effective population size of more than 16,000 by 28-30 kyr ago (Fig. 4A). Strikingly, around 53 kyr ago we observe an increase in population size in the Algerian and Egyptian AGW that can be connected to a local wetter event in the northernmost fringe of the Sahara (Hoffmann et al., 2016). Western populations reached minimum effective population sizes around 30-40 kyr ago (Moroccan individuals), and 35-25 kyr (Algeria, Ethiopia), followed by local recoveries and relative maximums around 15-25 kyr ago. The Kenyan, Egyptian and Senegalese lineages (from the extremes of the distribution) show no recovery after their initial minima.

A general decrease in population sizes was observed from the last part of the last Glacial Maximum (LGM) around 18 kyr ago through the Younger Dryas (YD) with no recoveries during the wetter phase of the Holocene GSP ca. 10-5.5 kyr ago (Fig. 4B). Consistent with a differentiation in paleodistribution of these animals, there are general increases of effective population size that start at different times: while the easternmost (Ethiopia, Kenya) lineages reach a local minimum in population size around 3-3.5 kyr ago, the northwestern Moroccan lineages experience their minimum more recently at ca. 2.7-3 kyr ago and Senegal and Algeria reach theirs around 2-2.2 kyr ago. The western populations all reach their local maximum around 1000 years ago, to decrease again afterwards (Fig. 4B). A final steep increase in population size is seen only in the Egyptian and Kenyan lineages ca. 150 and 90 years ago, respectively. While the first one could be attributed to local hybridization with another species, the Kenyan increase could be the result of two converging lineages of the same species. Our analysis failed to identify any more changes in population size after 70 years ago (ca. 23 generations), and attempts to do so with IBD-based methods did not yield any results.

We observed similar behavior for all four populations in Watterson’s theta ($\Theta_w$) (Watterson, 1975), nucleotide diversity ($\Theta_r$) (Tajima, 1989), and Fu & Li’s theta ($\Theta_{FL}$) (Fu & Li, 1993) (Table 1, Fig. S6A). Although means showed very similar trends, Fu and Li’s ($\Theta_L$) (Fu & Li, 1993) and Fay and Wu’s ($\Theta_H$) (Fay & Wu, 2000) thetas were very different between the northwest and eastern African golden wolf populations.

Neutrality tests show very similar results in all four populations considered (Table 1, Fig. S6B). While Tajima’s D (Tajima, 1989) is negative and very close to zero in northwestern AGW, North American
This article is protected by copyright. All rights reserved
We found that African golden wolves had a mean genome wide heterozygosity of $7.83 \times 10^{-4}$ (standard deviation (SD): $1.82 \times 10^{-4}$); closer to North America coyotes ($7.95 \times 10^{-4}$, SD: $9.38 \times 10^{-5}$) than to gray wolves from the Middle East ($6.33 \times 10^{-4}$, SD: $1.89 \times 10^{-4}$) and higher than domestic dogs ($5.17 \times 10^{-4}$, SD: $9.71 \times 10^{-5}$) (Fig. S7). However, the west Moroccan individual had a low genome wide heterozygosity ($5.01 \times 10^{-4}$).

Although inbreeding coefficients varied among individuals, the Egyptian, Kenyan and west Moroccan African golden wolves were more inbred than the rest of the individuals of the species (Table 3), partially consistent with being at the extremes of the distribution. PLINK and ngsF’s estimation of inbreeding coefficient yielded some differences that could be explained by differences in the calculation methods of each software. Our efforts to determine the influence or bias across genomes of varying depths of coverage (7X, 9X, 11.2X and 15-24X) found that ngsF is less sensitive to changes in coverage than PLINK, and tends to overestimate the homozygosity of the Kenyan over the Egyptian individual (Table S6). In any case, PLINK also identified the Kenyan, Egyptian, and west Moroccan individuals as the most inbred.

This approach could not be replicated, however, when using ROH-based methods to estimate inbreeding. Except for samples with a high mean genome wide coverage (Kenyan AGW – 26X, Iranian gray wolf – 26X, Californian coyote – 23X), we detected almost no ROHs in AGWs (Table 3). ROH-based calculations mostly underestimated Fi ratios detected by both ngsF and PLINK, especially in lower coverage genomes. This failure could be due to the dependency of these methods to detect long stretches with homozygous sites when using low coverage samples, or on the wide variety in methods and thresholds to infer ROHs (Sams & Boyko, 2019).

Using a 50-kb windows-based approach across the west Moroccan AGW and the east Moroccan AGW genomes (Fig. S8), we observed several regions where heterozygosity approaches zero in the west Moroccan individual. This trend is especially remarkable when compared with the east Moroccan individual, who only has a single possible ROH in chromosome 11. Furthermore, the Senegal sample (7X coverage) did not show any increased homozygosity as compared to other samples of higher coverage.

The west Moroccan genome is the African golden wolf with the highest inbreeding coefficient observed in this dataset and the east Moroccan is the one with the lowest. This difference between geographically close lineages is remarkable when considering that other lineages more distant from each other share more alleles, have higher rates of heterozygosity, and lower pairwise Fst values (Table 2A).

**Divergence during glacial periods**

We observed highly consistent results in divergence time estimation regardless of genome coverage (see Supplementary File: “Replicability tests: results”, Table S7). MiSTI also gave coherent results regardless of the time segment definition used (Table S8), so we used time segments defining humid/dry conditions as the hydrological thresholds for glacial periods.

This article is protected by copyright. All rights reserved
periods of Sahara according to Larrasoaña et al., (2013). Every divergence time estimation was run covering all time steps until 150 kyr ago. We also found good consistency between our MiSTI and Cavalli-Sforza divergence estimates except for those divergence times involving either the Ethiopian or the west Moroccan genomes (Table S9), which had the lowest genome wide heterozygosity (see Table 3).

We found a recurrent pattern of isolation of local lineages at times corresponding to glacial maxima in the Northern Hemisphere. We failed to detect the splitting time of three lineages (Algeria, Senegal, east Morocco) (Fig. 5, Table S10), which is consistent with very low genome wide pairwise Fst values (Table 2). Since the last splitting time considered by MiSTI was ca. 3 kyr ago, the separation of east Morocco, Algeria and Senegal lineages (referred to as the EMAS cluster) may have happened very recently (less than 3 kya), or they may maintain geneflow. Splitting times of the EMAS cluster from Egyptian AGW happened around 9-11 kyr ago (Table 1), which overlaps or follows the Younger Dryas in north Africa ca. 10-13 kyr ago (Ehrmann et al., 2017).

Most lineage divergences happened between ca. 16 kyr and 30 kyr ago (Table S10), overlapping with cold Heinrich stadials (Ehrmann et al., 2017; Heinrich, 1988; Rohling et al., 2003). The west Moroccan lineage diverged from the EMAS cluster ca. 16-21 kyr ago (Table S10), overlapping with Heinrich stadial H2 (Ehrmann et al., 2017; Heinrich, 1988; Rohling et al., 2003), as well as the divergence time of Ethiopian and Kenyan lineages, which probably happened around 17.5 kyr ago. The Kenyan lineage shows a more widespread divergence with all other lineages in the north, with a likely splitting time of 22.3 kyr with the EMAS cluster, 24.1 kyr with the Egyptian lineage, and ca. 29.4 kyr with the west Moroccan lineage, which could have followed cooling periods known as Heinrich stadials H2 and H3.

Divergence times between the Ethiopian lineage and EMAS cluster and the Egyptian lineage probably happened between 26.9-30.3 kyr ago, coinciding with Heinrich stadial H3 (Fig. S9B, Table S10). We have not found any signatures of divergence after the Holocene GSP ca. 10-5.5 kyr ago (Drake et al., 2013; Larrasoaña et al., 2013; Smith, 2012).

After calculating migration rates during time segments, we observe a general trend with increased migration rates at around 80 kyr, 100 kyr and 120 kyr for some lineages, therefore increasing gene flow (Table S11). This is consistent with previous paleoclimatological data that point to three GSPs around those periods (Drake et al., 2013; Ehrmann et al., 2017; Larrasoaña et al., 2013; Smith, 2012). Furthermore, we have detected certain periods of increased migration that could coincide with minor wet phases reported at ca 52.5-50.5 kyr and 37.5-33 kyr ago in the northern Sahara (Hoffmann et al., 2016), such as west Morocco → Egypt (35.2-51 kyr) and Senegal → east Morocco (29.9-36.3 kyr). However, time segments as defined by PSMC are wide and more resolution is needed to properly infer when these migration rates increased in the past.
DISCUSSION

Climate change-driven differentiation of African golden wolf lineages

Our results indicate that the divergence times of all lineages of African golden wolves occurred during the latest Pleistocene, between 50 and 10.5 kyr ago, with most divergence times clustered between 16 and 30 kyr ago, broadly coinciding with the Late Glacial Maximum (LGM) at ca. 33-16 kyr ago (Clark et al., 2009) (Fig. 5). Strikingly, all divergence events are associated to either Heinrich stadials H1 to H4 or to the Younger Dryas, clearly linking divergence times with periods of enhanced dry conditions in the Sahara. We hypothesize that these 1500-year-long drier periods reinforced the isolation of mesic and water-dependent species in refugia (Brito et al., 2014), where isolation during hundreds of generations caused genetic divergence. Although some relative warming and cooling periods have been proposed since the middle Holocene that could be linked to relative changes in population size, the worldwide impact of these local events is heavily debated (Neukom, Steiger, Gómez-Navarro, Wang, & Werner, 2019) and we could not find a direct correlation.

It has been proposed that the onset of wetter conditions during later GSP permitted mesic species to expand and reconnect, a pattern seen in a wide number of species in north Africa (Ben Faleh et al., 2012; Cosson et al., 2005; Lerp et al., 2011; Nicolas et al., 2009; Rato et al., 2007). The last GSP (ca. 10-5.5 kyr ago) reconnected animals isolated in several refugia (Kuper, & Kröpelin, 2006; Linstädter & Kröpelin, 2004; Yeakel et al., 2014). However, we have not detected strong signatures of gene flow between African golden wolf lineages during this period (except for the EMAS cluster), suggesting that this period was not wet enough to erase patterns of genetic differentiation. This is consistent with the drier nature of the Holocene GSP in comparison with previous GSP (Ehrmann et al., 2017).

An exception to this is the east Moroccan AGW. While the west Moroccan population seems to have been isolated for the last 18,000 years and have undergone a high degree of inbreeding, the east Moroccan individual presents one of the highest levels of genome wide heterozygosity and is genetically close to lineages from Senegal and Algeria, forming the EMAS cluster. It is characterized by a wide distribution (3,200 km wide), a relatively recent divergence (less than 3,000 years ago), and the exclusion of the west Moroccan lineage, all of which are intriguing features. African golden wolves are known to have large home ranges (up to 22 km² in Kenya, Fuller et al., 1989, and 64.8 km² in Ethiopia, Admasu et al., 2004) and large dispersal capabilities (in Tunisia, an individual was detected to have walked 230 km in 98 days, Karssene et al., 2018). A recent study (Eddine et al., 2020) detected almost no genetic structuring in a wide variety of samples from Algeria and Tunisia that encompassed roughly 1,200 km. The east and west Moroccan individuals are less than 1,000 km apart from each other. Why then, would the west
Moroccan individual present such high levels of inbreeding and be isolated from other African golden wolves for nearly 18,000 years?

Atlas: refugium during glacial times

Previous studies (Cosson et al., 2005; Husemann et al., 2014; Leite et al., 2015; Rato et al., 2007) have identified the Atlas Mountains of Maghreb as a glacial refugium during drier times for a variety of species. The highly heterogeneous landscape is formed by several mountain ranges (Anti-Atlas, High Atlas, Middle Atlas, Rif) with some of the highest peaks in north Africa (Jbel Toubkal, 4,165m), deep humid valleys, and relict high mountain cedar forests (Abel-Schaad et al., 2018) with isolated and endangered populations of Macaca sylvanus (Ciani et al., 2005) and rivers that act as barriers to arid-adapted species (Rato et al., 2007). These features are key for the establishment of “refugia within refugia”, with high endemism, barriers that subdivide habitats and the potential for genetic divergence and speciation (Dufresnes et al., 2020; Gómez & Lunt, 2007).

The isolation of the west Moroccan lineage could have arisen in a variety of ways. The ‘rear edge hypothesis’ of Hampe & Petit (2005) proposes a scenario where previously isolated demes in interglacial refugia converge in an admixture zone, leaving behind groups that have undergone local adaptations in response to abiotic stresses. This could explain why the east Moroccan individual appears to have merged with the Algerian and Senegalese lineages in recent times, even though the demographic histories of the west and east Moroccan individuals appear overlapped until 60,000 years ago (Fig. 3). In this scenario, the west Moroccan lineage could be one of the groups from the rear edge that failed to reconnect with others. Hampe and Petit (2005) explain how relict populations are not necessarily the source of postglacial expansions and this could be the case of the west Moroccan lineage.

Another possibility is that the west Moroccan lineage belongs to a population that has not contributed to the EMAS cluster for thousands of years. In a previous study on red fox (Vulpes vulpes), Leite et al. (2015) detected two main lineages in northwestern Africa: animals from Atlantic Sahara to Tunisia (Maghreb 1) and a more isolated lineage restricted to the valley of Oukaimeden, north of the High Atlas Mountains (Maghreb 2). The west Moroccan individual was found north of the High Atlas, which is the tallest mountain range of the Atlas Mountains (over 4000 m high) and could have restricted movements of African golden wolves. However, African golden wolves have been observed living at medium to high elevations (2,000-3,000m) in Algeria (Amroun et al., 2014), Ethiopia (Admasu et al., 2004; Gaubert et al., 2012; Rueness et al., 2011; Simeneh, 2010), Tanzania (Temu, Nahonyo, & Moehlman, 2018), and Morocco (Cuzin, 2003; Urios et al., 2015; Waters, Harrad, Amhaouch, Taiqui, & Senn, 2015). There is a possibility, therefore, that the west Moroccan lineage represents a different, adapted ecotype of the same species of African golden wolf, parallel to what has been seen with North American gray wolves, where ecology

This article is protected by copyright. All rights reserved
drives differentiation in local populations (Leonard, 2014). Further studies and more sampling across the High Atlas Mountains will reveal whether African golden wolves follow a localized pattern of specialized demes associated with different ecosystems.

**Recent admixture in the EMAS cluster**

In a previous study based on 13 autosomal microsatellites and mitochondrial control region sequences, Eddine et al. (2020) failed to identify structure between Algerian and Tunisian African golden wolves, and estimated an increase in population size between 3,840 and 6,720 years ago. However, across this time period we found an almost constant decrease of population size with ngsPSMC (Fig. 4B). Higher values of $\Theta_H$ indicate a greater abundance of high-frequency variants than in northwest AGW, which could suggest a high number of shared variants among the northwestern individuals and possible recent gene flow between different lineages of wolves. In contrast, eastern AGW show a lower abundance of high frequency variants, which is possibly indicative of a process of local divergence of the two lineages in the region (Ethiopia, Kenya).

The high genome wide heterozygosity observed in the EMAS cluster could come from a recent admixture of distant lineages (as we failed to detect a divergence time using MiSTI), increasing genetic variability. Previous work in red fox also detected a widespread lineage that extended from Atlantic Sahara to Tunisia (Leite et al., 2015). Allele-based estimations of $\text{Ne}$ have been found to be affected by recent admixture processes in human populations (Lohmueller, Bustamante, & Clark, 2010) and recent studies have focused on independently ascertaining demographic histories of genomic portions of diverse origin in admixed individuals (Browning & Browning, 2015; Browning et al., 2018; Skov et al., 2020). Future IBD-based studies of whole genomes of individuals from the EMAS cluster could help disentangle the hypotheses of a recent Neolithic expansion versus a recent admixture that enriched the diversity of local demes.

**Dynamic demographic histories**

With PSMC we have found a complex demographic history, where lineages do not follow the same trends of expansion or contraction through time (Fig. 3). This could be due to a variety of causes. First, current lineage locations might not represent those of the past. Consistent with the detected pattern in the west Moroccan sample, other lineages might have remained isolated for thousands of generations in a refugium and not have contributed for several GSP to the postglacial admixed populations, therefore increasing genome wide inbreeding and decreasing $\text{Ne}$.

Second, GSP did not lead to homogeneous savannah landscapes throughout the desert, a circumstance that has been shown for the two GSP for which more paleoenvironmental evidence is available (Eemian, ca. 122-128. kyr ago; and Holocene, 5.5-10 kyr ago). Complex topography modulated
the S-N and W-E gradient of decreasing monsoonal rainfall, thereby creating areas of more or less aridity with different environments (Larrasoaña et al., 2013). With such a heterogeneous landscape, mild GSP would not have had much impact on local populations closer to arid zones while other populations would have benefitted from more humid environments. Different refugia within the Sahara could have hosted distinct populations of African golden wolves with diverse reactions to climate change. Further, competitive pressure from black-backed jackals (Van Valkenburgh & Wayne, 1994) and/or hyaenas (Kebede et al., 2017; Kingdon, 2013) may not have been equal throughout the entire Sahara region if their past distributions were not entirely overlapping with the distribution of African golden wolves.

A third possibility is that we are detecting different ecotypes of the same species. In gray wolves, the onset of postglacial conditions benefited some populations, while others, possibly adapted to colder conditions and/or bigger prey, declined or went extinct (Ersmark et al., 2016; Leonard et al., 2007). Although current African golden wolves are generalist predators and maintain a varied diet composed of small mammals, birds, plants, insects and waste (Amroun et al., 2014; Amroun et al., 2006; McShane & Grettenberger, 1984), predation of gazelles has been detected in Niger, Kenya and Tanzania (Fuller et al., 1989; McShane & Grettenberger, 1984; Moehlman, 1986; Temu, Nahonyo, & Moehlman, 2016; Temu et al., 2018). It is unknown if the decline of gazelles after ca. 25 kyr ago (Lerp et al., 2011) impacted local populations of AGW, or if some of them were more specialized predators of ungulates. Another factor is habitat preference. While African golden wolves have been detected in all sorts of environments in Algeria (Amroun et al., 2014; Amroun et al., 2006), Tunisia (Karssene, Chammem, Khorchani, Nouira, & Li, 2017; Karssene et al., 2018), Morocco (Cuzin, 2003) and Niger (McShane & Grettenberger, 1984), some prefer farmlands and covered woodlands over open environments in Ethiopia (Admasu et al., 2004; Simeneh, 2010). In places where AGW are sympatric with competitors such as black-backed jackals, they are found mostly in open grasslands and rarely in woodlands (Fuller et al., 1989; Moehlman, 1986). Human pressure is another important factor to take into consideration. African golden wolves are found close to highly anthropized zones in Algeria and Ethiopia (Admasu et al., 2004; Amroun et al., 2014; Amroun et al., 2006; Simeneh, 2010), while they tend to avoid humans in Morocco, where higher human pressures exist (Brito et al., 2014; Cuzin, 2003). While it has been proposed that increasing anthropization could have driven a recent population expansion in northeast Africa either through the arrival of caprid livestock or through predator control (Eddine et al., 2020), almost negligible amounts of livestock remains were found in fecal diet studies in Niger, Tanzania and Ethiopia (Admasu et al., 2004; Fuller et al., 1989; McShane & Grettenberger, 1984; Simeneh, 2010). It remains plausible to consider that different abiotic and biotic factors could have driven different populations of African golden wolf to specialize into ecotypes. It is unknown whether current African golden wolves arose from bottlenecked specialized ecotypes from the past, but it is a possibility that cannot be excluded.

This article is protected by copyright. All rights reserved
In this study we sampled the extremes of the distribution that were most probably not directly connected (with the exception of the EMAS cluster) after the Younger Dryas; however, “jackal-like” forms have been discovered in several archaeological sites with dates closer to Holocene GSP conditions (di Lernia, 1998; Guagnin, 2015; Sereno et al., 2008) that could indicate higher connectivity among locations that we have not detected with our dataset. A higher number of samples from more diverse locations -and possibly, historic or ancient DNA – could ascertain whether extinct ecotypes of African golden wolf roamed the once green Sahara landscape.

**Speciation and the mid-Pleistocene transition**

The mid-Pleistocene transition was characterized by the shift of 41 kyr long glacial-interglacial cycles to much longer and more intense 100 kyr long glacial-interglacial cycles (McClymont et al., 2013) and a general trend of increased aridity in northern Africa (DeMenocal, 2004; Trauth, Larrasoaña, & Mudelsee, 2009). Based on genetic evidence, two studies suggested a date of 1.32 million years ago (Ma) for the speciation event that gave rise to AGW, with a confidence interval of either 1.0-1.65 Ma (Koepfli et al., 2015) or 1.1-1.5 Ma (Chavez et al., 2019). This may be consistent with the oldest “jackal-like” fossils referred to as golden jackals (Canis aureus) dating back to the Middle Pleistocene in Morocco (Geraads, 2011). Overall, these data suggest that a shift to enhanced aridity in the Sahara at 1.44 +/- 0.2 Ma (Trauth et al., 2009) drove the speciation of African golden wolves (Fig. 6) (Chavez et al., 2019; Koepfli et al., 2015) and reinforces the view that the mid-Pleistocene transition was a prime driver of speciation events in north Africa (DeMenocal, 2004). The increase of aridity ca. 1.2-1.4 Ma also coincides with the expansion and formation of new haplogroups of scimitar-horned onyx (Iyengar et al., 2007), the appearance of several clades of rodents (Praomys rostratus; Nicolas et al., 2008; genus Acomys; Nicolas et al., 2009; desert-adapted Gerbillus tarabuli, Ndiaye et al., 2011) and the divergence of red foxes and Rueppell foxes, the latter being a species more adapted to arid conditions (Leite et al., 2015). In east Africa, the mid-Pleistocene transition is suggested to be connected to the appearance of modern carnivores, especially those of the genus Canis (Werdelin & Lewis, 2005)

**ACKNOWLEDGEMENTS**

We are grateful to I. Salado, S. Ravagni, A. Cornellas and M. Camacho-Sánchez for their assistance in lab work and in the analyses. Logistical support was provided by Laboratorio de Ecología Molecular (LEM-EBD) and by the infrastructures offered by Doñana’s Singular Scientific-Technical Infrastructure (ICTS-EBD). Informatic technical support was provided by the High-Performance Computing (HPC) Research Center in Princeton University. CS was supported by a PhD fellowship from Programa Internacional de Becas “La Caixa-Severo Ochoa” of the Spanish Ministerio de Economía y

This article is protected by copyright. All rights reserved
Competitividad and La Caixa bank (BES-2015-074331). This project was funded by the Frontera grant P18-FR-5099 from the Junta de Andalucía. EBD-CSIC received support from the Spanish Ministry of Economy and Competitiveness under the 'Centro de Excelencia Severo Ochoa 2013-2017' program, SEV-2012-0262.

AUTHOR CONTRIBUTIONS

J.A.L. developed the initial concept which was further developed with C.S. and B.V.H. V.U. provided the sample of the west Moroccan African golden wolf. C.S. conducted data analyses. B.V.H. and J.A.L. assisted with interpretation of genomic results. J.C.L. assisted with interpretation of paleoclimatological results. All authors discussed the results and provided edits and approval of the manuscript.

DATA AVAILABILITY STATEMENT

All data generated in this study have been made available in Github: https://github.com/cdomsar/DivgLupaster. Newly sequenced African golden wolf has been deposited into EMBL-EBI server under accession number ERP123054.


This article is protected by copyright. All rights reserved
This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


https://github.com/vlshchur/MiSTI


Simeneh, G. (2010). *Habitat use and Diet of Golden Jackal (Canis aureus) and Human-Carnivore Conflict in Guassa Community Conservation Area, Menz.* (MSc thesis) Addis Ababa University, Department of Biology.


This article is protected by copyright. All rights reserved
This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved
Figure 1. Map of Africa showing isohyets, elevations, main water sources, African golden wolf distribution (in green) according to IUCN and location of the seven samples included in this study (in red).
Figure 2. NGSadmix plots of Old World canids showing admixture proportions, including the 23 individuals used at this study mapped against African hunting dog (A) and only African golden wolves (B). Eastern (Kenya, Ethiopia) African golden wolves cluster in a different group from those from the north (Egypt, Algeria, east Morocco, west Morocco, Senegal). The Egyptian individual also seems to have ancestry from gray wolves or domestic dogs. EW: Ethiopian wolf. EGJ: Eurasian golden jackal. Plots are based in 9.885M sites (see text for details).
Figure 3. PSMC plot of six African golden wolf individuals: Algeria, west Morocco, east Morocco, Senegal, Ethiopia, Kenya. Consensus sequences were extracted from .bam files and $\Theta_0$ values were corrected according to the False Negative Rate (FNR) calculated for each actual coverage by down sampling the Kenyan genome (24X). Individual PSMC plots were bootstrapped 50 times each. Proposed Green Sahara Periods (GSPs) were included from Larrasoána et al., (2013) and Ehrmann et al., (2017). Heinrich events (H) of local cooling were also included according to Rohling et al. (2003) and Ehrmann et al. (2017). The black line defines the event of speciation (ca. 1.3 Myr ago) according to Koepfli et al. (2015) and Chavez et al. (2019) and a confidence interval of 1.10-1.5 Myr (Chavez et al., 2019). Numbers after each individual mean FNR calculated by visually adjusting using the psmc_plot.py program from the PSMC package.
Figure 4. ngsPSMC plots of the seven African golden wolf individuals of this study. (A) ngsPSMC plot of all individuals using maximum Ne = 17000. (B) ngsPSMC plot using maximum Ne = 2000. Cooling events (H) are the same as in Figure 3. Local events of wetter conditions are marked up (north Sahara: 65-61 ka, 52.5-50.5 ka and 37.5-33 ka (Hoffmann et al., 2016)) and down (Sahel: 55-60 ka; (Tjallingi et al., 2008; Weldeab et al., 2007)) of Figure 4A. HO = Holocene Optimum as in Larrasoaña et al., (2013). YD = Younger Dryas.
Figure 5. Times of divergence between pairs of lineages of African golden wolves. δ¹⁸O data from the North Greenland Ice Core Project (NGRIP members, 2004) are shown, along with main glacial and interglacials periods, for the last 125,000 years. Intervals for divergence of pairs of lineages were considering the top 5% (light gray) and 1% (dark gray) of values for polynomial equations of adjusted graphs to the data points after $R^2=0.99$. For further explanation, see Table S8. H bars: Heinrich events. YD: Younger Dryas.
Figure 6. Dust flux record of Ocean Drilling Program (ODP) Leg 160, site 967 (Larrasoaña et al., 2003) of eastern Mediterranean Sea from last three million years (Ma). We have incorporated an estimation of the onset
of the mid-Pleistocene climate Transition (MPT) (McClymont et al., 2013), the shift from 41-kyr- to 100-kyr-
short aridity cycles and a represents a breakpoint in north African aridity (Trauth et al., 2009). Estimation of the
speciation event of African golden wolves circa 1.32 million years ago (Ma) is also included with confidence
intervals (blue: 1.0-1.65, Koepfli et al. (2015); red: 1.1-1.5, Chavez et al. (2019)).
Supplementary figure captions

Supplementary Figure 1. PSMC plots of African golden wolf (AGW) genomes under different conditions. A, B represent PSMC plots of the Kenyan AGW with the normal (24X) and downsamled coverages (15X, 11.2X, 9X, 7X) without (A) and with (B) False Negative Rate correction for low heterozygosity due to low coverages. C, D represent PSMC plots of six AGW (Algeria, Ethiopia, Kenya, East Morocco, West Morocco, Senegal) with lower (C) and upper (D) bounds of the mutation rate estimation by Koch et al., (2019) (2.7-7.1*10^-9).

Supplementary Figure 2. Genotype likelihood-based Principal Component Analysis (PCA) generated by ngsCovar from the ngsTools package. PCA was called using 2.54 million sites in 16 genomes of wild Old World canids (African golden wolves, gray wolves, Ethiopian wolves, Eurasian golden jackals) and 7 genomes of domestic dogs.

Supplementary Figure 3. SNP-based Principal Component Analysis (PCA) of 23 canid individuals. SNPs were called based on genotype likelihood using ANGSD with the -doPlink option, curated and filtered for Hardy-Weinberg equilibrium and linkage disequilibrium using PLINK v1.9. PCA was generated by flashPCA using 625,000 sites in 16 genomes of wild Old World canids (African golden wolves, gray wolves, Ethiopian wolves, Eurasian golden jackals) and 7 genomes of domestic dogs.

Supplementary Figure 4. Best-fit calculation of K (likelihood) vs values of K as calculated by NGSadmix using 23 genomes of Old World canids mapped against African hunting dog (admixture plot in Figure 2).

Supplementary Figure 5. SNP-based Admixture plots of Old World canids mapped against African hunting dog showing admixture proportions, including the 23 individuals used at this study. SNPs were called as in Supplementary Figure 2. Eastern (Kenya, Ethiopia) African golden wolves cluster in a different group from those from the north (Egypt, Algeria, East Morocco, West Morocco, Senegal). EW: Ethiopian wolf. EGJ: Eurasian golden jackal. This plot is based in 625,000 unlinked sites.

Supplementary Figure 6. Thetas per site and neutrality tests of four populations: east African Golden Wolves (AGW) (Ethiopia, Kenya), northwest AGW (Algeria, east Morocco, west Morocco, Senegal), Coyote (California, Midwest, Mexico), Gray wolves of the Middle East (ME) (S. Arabia, Iran, Syria). We considered 50-kb non-overlapping windows across the whole genome and filtered out those windows with a number of
sites outside the 99.7% of the distribution (mean +/- 3 standard deviations). Theta statistics were calculated dividing by the total number of sites. Neutrality tests were averaged per window.

**Supplementary Figure 7.** Genome wide heterozygosity calculated per individual and population. Heterozygosity was calculated using the fraction of singletons from the unfolded Site-Frequency Spectrum (SFS). Genome wide heterozygositides were corrected using the Kenyan African golden wolf genome (24X) and down sampling it to each coverage, calculating proportion of lost heterozygosity for each coverage. Corrections are marked in darker colors. AGW: African golden wolves. GW: gray wolves. EGJ: Eurasian golden jackals.

**Supplementary Figure 8.** Heterozygosity plots of east and west Moroccan individuals per chromosome. Plots were generated with ROHan using 500kB windows, minimum coverage of 5X and maximum coverage of 2.5 times the mean coverage per genome. --rohmu option were set as 2e-5. All other settings were left as default.

**Supplementary Figure 9.** Log-likelihood of divergence between members of the north African golden wolf cluster (A) and north vs. east African golden wolf cluster (B) vs time. Likelihood of divergence times was calculated paralllizing MiSTI with GNU Parallel using the default optimization round. Time segments were defined using likely aridization / regreening Sahara periods defined in the literature. A polynomial curve equation was adjusted to the fifth degree and plotted when R² > 0.99.
Table 1. Mean values and standard deviations of genome wide thetas per site and neutrality tests of four populations: northwest African Golden Wolves (AGW) (Algeria, east Morocco, west Morocco, Senegal), east AGW (Ethiopia, Kenya), gray wolves of the Middle East (ME) (S. Arabia, Iran, Syria), and coyote (California, Midwest, Mexico). We considered 50-kb non-overlapping windows across the whole genome and filtered out those windows with a number of sites outside the 99.7% of the distribution (mean +/- 3 standard deviations). Theta statistics were calculated dividing by the total number of sites. Neutrality tests were averaged per window. $\Theta_W$: Watterson’s theta (Watterson, 1975). $\Theta_\pi$: Nucleotide diversity (Tajima, 1989). $\Theta_{FL}$: Fu and Li’s theta (Fu & Li, 1993). $\Theta_H$: Fay and Wu’s theta (Fay & Wu, 2000). $\Theta_L$: Fu and Li’s L (Fu & Li, 1993). Neutrality tests: Tajima’s D (Tajima, 1989); Fu’s F and D (Fu & Li, 1993); Fay’s H (Fay et al., 2006); Zeng’s E (Zeng et al., 2006).

<table>
<thead>
<tr>
<th>Populations</th>
<th>$\Theta_W$</th>
<th>$\Theta_\pi$</th>
<th>$\Theta_{FL}$</th>
<th>$\Theta_H$</th>
<th>$\Theta_L$</th>
<th>Tajima’s D</th>
<th>Fu’s F</th>
<th>Fu’s D</th>
<th>Fay’s H</th>
<th>Zeng’s E</th>
</tr>
</thead>
<tbody>
<tr>
<td>northwest AGW</td>
<td>8.52 ± 3.96*10^{-4}</td>
<td>8.55 ± 4.15*10^{-4}</td>
<td>4.62 ± 2.35*10^{-4}</td>
<td>1.7 ± 0.77*10^{-3}</td>
<td>1.28 ± 0.59*10^{-3}</td>
<td>-0.34 ± 3.98*10^{-1}</td>
<td>7.92 ± 4.32*10^{-1}</td>
<td>9.06 ± 3.63*10^{-1}</td>
<td>-1.54 ± 0.41</td>
<td>9.13 ± 2.08*10^{-1}</td>
</tr>
<tr>
<td>east AGW</td>
<td>7.77 ± 4.42*10^{-4}</td>
<td>7.93 ± 4.53*10^{-4}</td>
<td>4.61 ± 2.99*10^{-4}</td>
<td>1.17 ± 0.64*10^{-3}</td>
<td>9.55 ± 5.43*10^{-4}</td>
<td>2.09 ± 6.12*10^{-1}</td>
<td>1.04 ± 0.65</td>
<td>1.04 ± 0.55</td>
<td>-1.22 ± 0.58</td>
<td>7.79 ± 2.69*10^{-1}</td>
</tr>
<tr>
<td>ME gray wolves</td>
<td>8.47 ± 4.06*10^{-4}</td>
<td>8.49 ± 4.19*10^{-4}</td>
<td>4.87 ± 2.53*10^{-4}</td>
<td>1.49 ± 0.71*10^{-3}</td>
<td>1.17 ± 0.56*10^{-3}</td>
<td>-0.19 ± 4.02*10^{-1}</td>
<td>8.03 ± 4.45*10^{-1}</td>
<td>8.87 ± 3.77*10^{-1}</td>
<td>-1.46 ± 0.44</td>
<td>8.81 ± 2.2*10^{-1}</td>
</tr>
<tr>
<td>coyote</td>
<td>7.61 ± 3.88*10^{-4}</td>
<td>7.71 ± 4.02*10^{-4}</td>
<td>4.24 ± 2.48*10^{-4}</td>
<td>1.33 ± 0.68*10^{-3}</td>
<td>1.05 ± 0.54*10^{-3}</td>
<td>-0.19 ± 4.02*10^{-1}</td>
<td>8.03 ± 4.45*10^{-1}</td>
<td>8.87 ± 3.77*10^{-1}</td>
<td>-1.46 ± 0.44</td>
<td>8.81 ± 2.2*10^{-1}</td>
</tr>
</tbody>
</table>
Table 2. Pairwise genome wide $F_{ST}$ values of African golden wolves (A), gray wolves (B) and coyotes (C). $F_{ST}$ values were calculated with the ngsTools package. Distances in km between coordinates were calculated in Marble (Linux).

African golden wolves

<table>
<thead>
<tr>
<th></th>
<th>Algeria</th>
<th>Egypt</th>
<th>east Morocco</th>
<th>west Morocco</th>
<th>Senegal</th>
<th>Ethiopia</th>
<th>Kenya</th>
</tr>
</thead>
<tbody>
<tr>
<td>distances</td>
<td>2590</td>
<td>920</td>
<td>1300</td>
<td>3250</td>
<td>4610</td>
<td>5170</td>
<td></td>
</tr>
<tr>
<td>km</td>
<td>2.01x10^{-1}</td>
<td>6.09x10^{-2}</td>
<td>2.05x10^{-1}</td>
<td>9.5x10^{-2}</td>
<td>2.38x10^{-1}</td>
<td>4.05x10^{-1}</td>
<td>Algeria</td>
</tr>
<tr>
<td>Egypt</td>
<td>1.67x10^{-1}</td>
<td>3.04x10^{-1}</td>
<td>2.09x10^{-1}</td>
<td>2.92x10^{-1}</td>
<td>4.4x10^{-1}</td>
<td></td>
<td>Egypt</td>
</tr>
<tr>
<td>east Morocco</td>
<td>1.7x10^{-1}</td>
<td>8.11x10^{-2}</td>
<td>2.15x10^{-1}</td>
<td>3.62x10^{-1}</td>
<td></td>
<td></td>
<td>east Morocco</td>
</tr>
<tr>
<td>west Morocco</td>
<td>2.19x10^{-1}</td>
<td>3.46x10^{-1}</td>
<td>5.04x10^{-1}</td>
<td></td>
<td></td>
<td></td>
<td>west Morocco</td>
</tr>
<tr>
<td>Senegal</td>
<td>2.35x10^{-1}</td>
<td>4.04x10^{-1}</td>
<td></td>
<td>2.89x10^{-1}</td>
<td></td>
<td></td>
<td>Senegal</td>
</tr>
<tr>
<td>Ethiopia</td>
<td></td>
<td></td>
<td>6110</td>
<td>2.89x10^{-1}</td>
<td></td>
<td></td>
<td>Ethiopia</td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
<td>850</td>
<td></td>
<td></td>
<td></td>
<td>Kenya</td>
</tr>
</tbody>
</table>

Gray wolves

<table>
<thead>
<tr>
<th></th>
<th>S. Arabia</th>
<th>Iran</th>
<th>Syria</th>
</tr>
</thead>
<tbody>
<tr>
<td>distances</td>
<td>1270</td>
<td>1390</td>
<td></td>
</tr>
<tr>
<td>km</td>
<td>1.73x10^{-1}</td>
<td>1.66x10^{-1}</td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>2.54x10^{-1}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syria</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coyotes

<table>
<thead>
<tr>
<th></th>
<th>California</th>
<th>Mexico</th>
<th>Midwest</th>
</tr>
</thead>
<tbody>
<tr>
<td>distances</td>
<td>1060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>km</td>
<td>1.62x10^{-1}</td>
<td>1.34x10^{-1}</td>
<td>California</td>
</tr>
<tr>
<td>Mexico</td>
<td>1.15x10^{-1}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved
<table>
<thead>
<tr>
<th></th>
<th>2770</th>
<th>2590</th>
<th>Midwest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Values of inbreeding coefficients (Fi) using genomewide genotype likelihood and SNP-based approaches (ngsF, PLINK and ROHan, respectively) and using the ROH-based method of McQuillan et al. (2008) with ROHs calculated with genotype likelihood and SNP-based approaches (ROHan and PLINK, respectively). Values higher than 0.1 are marked in gray.

<table>
<thead>
<tr>
<th></th>
<th>F_i (ngsF)</th>
<th>F_i (PLINK)</th>
<th>F_ROH (PLINK)</th>
<th>F_ROH (ROHan)</th>
<th>F_ROH (ROHan)</th>
<th>F_ROH (ROHan)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500kb</td>
<td>min 500kb</td>
<td>med 500kb</td>
<td>max 500kb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algeria</td>
<td>1.17x10^{-2}</td>
<td>4.66x10^{-2}</td>
<td>2.34x10^{-7}</td>
<td>0</td>
<td>0</td>
<td>2.27x10^{-4}</td>
</tr>
<tr>
<td>Egypt</td>
<td>1.35x10^{-1}</td>
<td>6.49x10^{-2}</td>
<td>8.46x10^{-6}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>East Morocco</td>
<td>5.88x10^{-4}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.27x10^{-4}</td>
</tr>
<tr>
<td>West Morocco</td>
<td>2.3x10^{-1}</td>
<td>2.93x10^{-1}</td>
<td>3.34x10^{-6}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Senegal</td>
<td>2.65x10^{-2}</td>
<td>8.49x10^{-2}</td>
<td>7.17x10^{-7}</td>
<td>0</td>
<td>0</td>
<td>6.81x10^{-4}</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>2.20x10^{-5}</td>
<td>0</td>
<td>9.70x10^{-4}</td>
<td>0</td>
<td>0</td>
<td>2.27x10^{-4}</td>
</tr>
<tr>
<td>Kenya</td>
<td>6.04x10^{-2}</td>
<td>1.47x10^{-1}</td>
<td>9.53x10^{-4}</td>
<td>1.16x10^{-2}</td>
<td>2.68x10^{-2}</td>
<td>2.68x10^{-2}</td>
</tr>
<tr>
<td>GW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Arabia</td>
<td>8.47x10^{-2}</td>
<td>1.81x10^{-1}</td>
<td>1.47x10^{-5}</td>
<td>0</td>
<td>0</td>
<td>4.54x10^{-4}</td>
</tr>
<tr>
<td>Iran</td>
<td>1.90x10^{-5}</td>
<td>0</td>
<td>1.05x10^{-5}</td>
<td>4.76x10^{-3}</td>
<td>1.18x10^{-2}</td>
<td>1.66x10^{-2}</td>
</tr>
<tr>
<td>Syria</td>
<td>7.21x10^{-2}</td>
<td>1.57x10^{-1}</td>
<td>1.77x10^{-5}</td>
<td>0</td>
<td>0</td>
<td>4.54x10^{-4}</td>
</tr>
<tr>
<td>Coyotes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>9.60x10^{-5}</td>
<td>0</td>
<td>4.27x10^{-5}</td>
<td>1.86x10^{-2}</td>
<td>4.13x10^{-2}</td>
<td>4.70x10^{-2}</td>
</tr>
<tr>
<td>Mexico</td>
<td>9.77x10^{-3}</td>
<td>5.70x10^{-2}</td>
<td>2.10x10^{-5}</td>
<td>0</td>
<td>0</td>
<td>8.17x10^{-3}</td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved