



# High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox

Rémi Allio, Marie-Ka Tilak, Celine Scornavacca, Nico L Avenant, Andrew C Kitchener, Erwan Corre, Benoit Nabholz, Frédéric Delsuc

## ► To cite this version:

Rémi Allio, Marie-Ka Tilak, Celine Scornavacca, Nico L Avenant, Andrew C Kitchener, et al.. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. *eLife*, 2021, 10, 10.7554/eLife.63167 . hal-03147106

**HAL Id: hal-03147106**

**<https://hal.sorbonne-universite.fr/hal-03147106>**

Submitted on 19 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox

Rémi Allio<sup>1\*</sup>, Marie-Ka Tilak<sup>1</sup>, Céline Scornavacca<sup>1</sup>, Nico L. Avenant<sup>2</sup>, Andrew C. Kitchener<sup>3</sup>, Erwan Corre<sup>4</sup>, Benoit Nabholz<sup>1,5</sup>, and Frédéric Delsuc<sup>1\*</sup>

<sup>1</sup>Institut des Sciences de l'Evolution de Montpellier (ISEM), CNRS, IRD, EPHE, Université de Montpellier, France [remi.allio@umontpellier.fr](mailto:remi.allio@umontpellier.fr) [marie-ka.tilak@umontpellier.fr](mailto:marie-ka.tilak@umontpellier.fr) [celine.scornavacca@umontpellier.fr](mailto:celine.scornavacca@umontpellier.fr) [benoit.nabholz@umontpellier.fr](mailto:benoit.nabholz@umontpellier.fr) [frederic.delsuc@umontpellier.fr](mailto:frederic.delsuc@umontpellier.fr)

<sup>2</sup>National Museum and Centre for Environmental Management, University of the Free State, Bloemfontein, South Africa [navenant@nasmus.co.za](mailto:navenant@nasmus.co.za)

<sup>3</sup>Department of Natural Sciences, National Museums Scotland, Edinburgh, UK [a.kitchener@nms.ac.uk](mailto:a.kitchener@nms.ac.uk)

<sup>4</sup>CNRS, Sorbonne Université, FR2424, ABiMS, Station Biologique de Roscoff, 29680 Roscoff, France [corre@sb-roscoff.fr](mailto:corre@sb-roscoff.fr)

<sup>5</sup>Institut Universitaire de France (IUF)

\*Correspondence: [remi.allio@umontpellier.fr](mailto:remi.allio@umontpellier.fr), [frederic.delsuc@umontpellier.fr](mailto:frederic.delsuc@umontpellier.fr).

## ORCID

Allio, Rémi 0000-0003-3885-5410

Tilak, Marie-Ka 0000-0001-8995-3462

Scornavacca, Céline

Avenant, Nico L. 0000-0002-5390-9010

Kitchener, Andrew C. 0000-0003-2594-0827

Corre, Erwan 0000-0001-6354-2278

Nabholz, Benoit 0000-0003-0447-1451

Delsuc, Frédéric 0000-0002-6501-6287



## Abstract

In a context of ongoing biodiversity erosion, obtaining genomic resources from wildlife is essential for conservation. The thousands of yearly mammalian roadkill provide a useful source material for genomic surveys. To illustrate the potential of this underexploited resource, we used roadkill samples to study the genomic diversity of the bat-eared fox (*Otocyon megalotis*) and the aardwolf (*Proteles cristatus*), both having subspecies with similar disjunct distributions in Eastern and Southern Africa. First, we obtained reference genomes with high contiguity and gene completeness by combining Nanopore long reads and Illumina short reads. Then, we showed that the two subspecies of aardwolf might warrant species status (*P. cristatus* and *P. septentrionalis*) by comparing their genome-wide genetic differentiation to pairs of well-defined species across Carnivora with a new Genetic Differentiation index (GDi) based on only a few resequenced individuals. Finally, we obtained a genome-scale Carnivora phylogeny including the new aardwolf species.

## Keywords

Roadkill, Genomics, Population genomics, Phylogenomics, Species delineation, Carnivora, Systematics, Genetic differentiation, Mitogenomes, Africa.

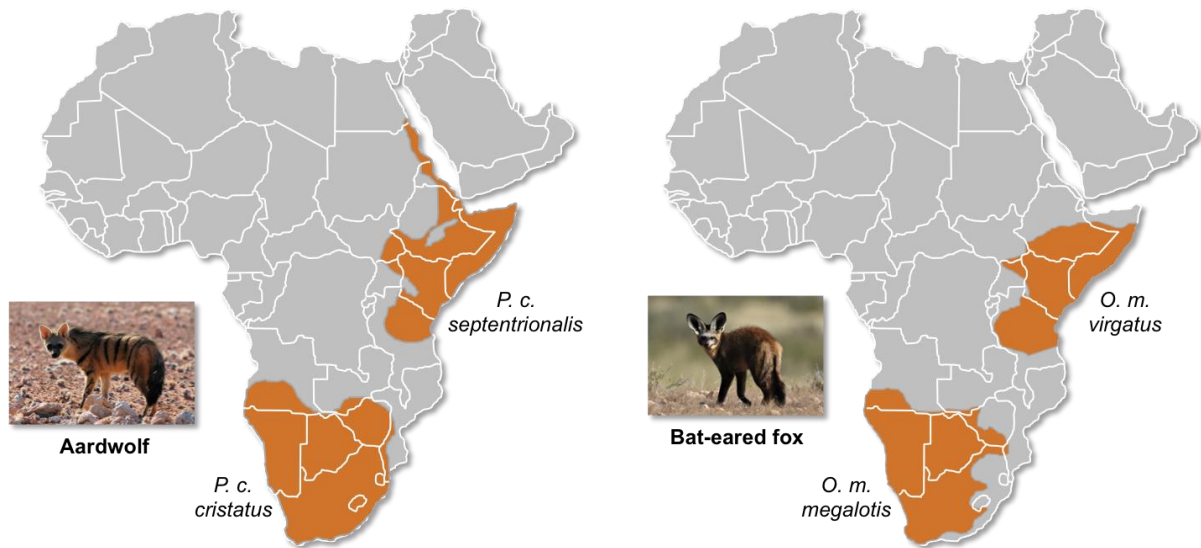
## Introduction

In the context of worldwide erosion of biodiversity, obtaining large-scale genomic resources from wildlife is essential for biodiversity assessment and species conservation. An underexploited, but potentially useful, source of material for genomics is the many thousands of annual wildlife fatalities due to collisions with cars. In particular, mammalian roadkill is unfortunately so frequent that several citizen-science surveys have been implemented on this subject in recent decades (Périquet et al., 2018; Shilling et al., 2015). For example, in South Africa alone, over 12,000 wildlife road mortality incidents were recorded by The Endangered Wildlife Trust's Wildlife and Roads Project from 1949 to 2017 (Endangered Wildlife Trust 2017). Initially developed to measure the impact of roads on wildlife, these web-based systems highlight the numbers of car-wildlife collisions. The possibility of retrieving DNA from roadkill tissue samples (Etherington et al., 2020; Maigret, 2019) could provide new opportunities in genomics by giving access not only to a large number of specimens of commonly encountered species, but also to more elusive and endangered species that might be difficult to sample otherwise.

Recent advances in the development of high-throughput sequencing technologies have made the sequencing of hundreds or thousands of genetic loci cost-efficient and have offered the possibility of using ethanol-preserved tissues, old DNA extracts, and museum specimens (Blaimer et al., 2016; Guschanski et al., 2013). This method, combined with third-generation long-read sequencing technologies, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) sequencing, have increased the sizes of the sequenced molecules from several kilobases to several megabases. The relatively high level of sequencing errors (10-15%) associated with these technologies can be compensated by sequencing at a high depth-of-coverage to avoid sequencing errors in *de novo* genome assembly and thus obtain reference genomes with high base accuracy, contiguity, and

completeness (Koren et al., 2017; Shafin et al., 2020; Vaser et al., 2017). Originally designed to allow direct sequencing of DNA molecules with simplified library preparation procedures, ONT instruments, such as the MinION (Jain et al., 2016), have been co-opted as a portable sequencing method in the field that proved useful in a diversity of environmental conditions (Blanco et al., 2019; Parker et al., 2017; Pomerantz et al., 2018; Srivathsan et al., 2018). This approach is particularly suitable for sequencing roadkill specimens, for which it is notoriously difficult to obtain a large amount of high-quality DNA because of post-mortem DNA degradation processes in high ambient environmental temperatures. Furthermore, it is possible to correct errors in ONT long reads by combining them with Illumina short reads, either to polish *de novo* long-read-based genome assemblies (Batra et al., 2019; Jain et al., 2018; Nicholls et al., 2019; Walker et al., 2014) or to construct hybrid assemblies (Di Genova et al., 2018; Gan et al., 2019; Tan et al., 2018; Zimin et al., 2013). In hybrid assembly approaches the accuracy of short reads with high depth-of-coverage (50-100x) allows the use of long reads at lower depths of coverage (10-30x) essentially for scaffolding (Armstrong et al., 2020; Kwan et al., 2019). A promising hybrid assembly approach, combining short- and long-read sequencing data has been implemented in MaSuRCA software (Zimin et al., 2017, 2013). This approach consists of transforming large numbers of short reads into a much smaller number of longer highly accurate “super reads”, allowing the use of a mixture of read lengths. Furthermore, this method is designed to tolerate a significant level of sequencing error. Initially developed to address short reads from Sanger sequencing and longer reads from 454 Life Sciences instruments, this method has already shown promising results for combining Illumina and ONT/PacBio sequencing data in several taxonomic groups, such as plants (Scott et al., 2020; Wang et al., 2019; Zimin et al., 2017), birds (Gan et al., 2019), and fishes (Jiang et al., 2019; Kadobianskyi et al., 2019; Tan et al., 2018), but not yet in mammals.

Here, we studied two of the most frequently encountered mammalian roadkill species in South Africa (Périquet et al., 2018): the bat-eared fox (*Otocyon megalotis*, Canidae) and the aardwolf (*Proteles cristatus*, Hyaenidae). These two species are among several African vertebrate taxa disjunct distributions between Southern and Eastern Africa that are separated by more than a thousand kilometres (e.g. ostrich, Miller et al., 2011; ungulates, Lorenzen et al., 2012). Diverse biogeographical scenarios, involving the survival and divergence of populations in isolated savanna refugia during the climatic oscillations of the Pleistocene, have been proposed to explain these disjunct distributions in ungulates (Lorenzen et al., 2012). Among the Carnivora subspecies have been defined based on this peculiar allopatric distribution not only for the black-backed jackal (*Lupulella mesomelas*; Walton and Joly 2003) but also for both the bat-eared fox (Clark, 2005) and the aardwolf (Koehler and Richardson, 1990) (**Fig. 1**). The bat-eared fox is divided into the Southern bat-eared fox (*O. megalotis megalotis*) and the Eastern bat-eared fox (*O. megalotis virgatus*) (Clark, 2005), and the aardwolf is divided into the Southern aardwolf (*P. cristatus cristatus*) and the Eastern aardwolf (*P. cristatus septentrionalis*) (Koehler and Richardson, 1990). However, despite known differences in behaviour between the subspecies of both species groups (Wilson et al., 2009), no genetic or genomic assessment of population differentiation has been conducted to date. In other taxa similar allopatric distributions have led to genetic differences between populations and several studies reported substantial intraspecific genetic structuration between Eastern and Southern populations (Atickem et al., 2018; Barnett et al., 2006; Dehghani et al., 2008; Lorenzen et al., 2012; Miller et al., 2011; Rohland et al., 2005). Here, with a novel approach based on a few individuals, we investigate whether significant genetic structuration and population differentiation have occurred between subspecies of bat-eared fox and aardwolf using whole genome data.



**Figure 1.** Disjunct distributions of the aardwolf (*Proteles cristatus*) and the bat-eared fox (*Otocyon megalotis*) in Eastern and Southern Africa. Within each species, two subspecies have been recognized based on their distributions and morphological differences (Clark, 2005; Koehler and Richardson, 1990). Picture credits: Southern aardwolf (*P. cristatus cristatus*) copyright Dominik Käufferle; Southern bat-eared fox (*O. megalotis megalotis*) copyright Derek Keats.

To evaluate the taxonomic status of the proposed subspecies within both *O. megalotis* and *P. cristatus*, we first sequenced and assembled two reference genomes from roadkill samples by combining ONT long reads and Illumina short reads using the MaSuRCA hybrid assembler. The quality of our genome assemblies was assessed by comparison to available mammalian genome assemblies. Then, to estimate the genetic diversity of these species and to perform comparative genome-scale species delineation analyses, two additional individuals from the disjunct South African and Tanzanian populations of both species were resequenced at high depth-of-coverage using Illumina short reads. Using these additional individuals, we estimated the genetic diversity and differentiation of each subspecies pair via an  $F_{ST}$ -like measure, which we called the genetic differentiation index, and compared the results with the genetic differentiation among pairs of well-established carnivoran sister species. Based on measures of genetic differentiation, we found that the two subspecies of *P. cristatus* warrant separate species status, whereas the subspecies of *O. megalotis* do not show such differentiation. Our results show that high-quality reference mammalian genomes could be obtained through a combination of short- and long-read sequencing methods providing opportunities for large-scale population genomic studies of mammalian wildlife using (re)sequencing of samples collected from roadkill.

## Results

### *Mitochondrial diversity within Carnivora*

The first dataset, composed of complete carnivoran mitogenomes available in GenBank combined with the newly generated sequences of the two subspecies of *P. cristatus*, the two subspecies of *O. megalotis*, *Parahyaena brunnea*, *Speothos venaticus* and *Vulpes vulpes*, plus the sequences extracted from Ultra Conserved Elements (UCE) libraries for *Bdeogale nigripes*, *Fossa fossana*, and *Viverra zibetha* (see *Methods* for more details), comprised

142 species or subspecies representing all families of Carnivora, including five *O. megalotis* and 10 *P. cristatus* individuals. Maximum likelihood (ML) analyses reconstructed a robust mitogenomic phylogeny, with 91.4% of the nodes (128 out of 140) recovered with bootstrap support higher than 95% (**Fig. 2a**). The patristic distances, extracted from the phylogenetic tree inferred with complete mitogenomes between the allopatric subspecies of aardwolf and bat-eared fox, were 0.045 and 0.020 substitutions per site, respectively (**Supplementary File 1**). These genetic distances are comparable to those observed between different well-defined species of Carnivora, such as the red fox (*Vulpes vulpes*) and the fennec (*V. zerda*) (0.029) or the steppe polecat (*Mustela eversmanii*) and the Siberian weasel (*M. sibirica*) (0.034) (see **Supplementary File 1**).

To further assess the genetic distances between the two pairs of subspecies and compare them to both polymorphism and divergence values observed across Carnivora, two supplemental datasets, including at least two individuals per species, were assembled by retrieving all COX1 and CYTB sequences, which are the two widely sequenced mitochondrial markers for carnivorans, available on GenBank. These datasets include 3,657 COX1 sequences for 150 species and 6,159 CYTB sequences for 203 species of Carnivora. After adding the corresponding sequences from the newly assembled mitogenomes, ML phylogenetic inference was conducted on each dataset. The patristic distances between all tips of the resulting phylogenetic trees were measured and classified into two categories: (i) intraspecific variation (polymorphism) for distances inferred among individuals of the same species and (ii) interspecific divergence for distances inferred among individuals of different species. Despite an overlap between polymorphism and divergence in both mitochondrial genes, this analysis revealed a threshold between polymorphism and divergence of approximately 0.02 substitutions per site for Carnivora (**Fig. 2b**). With a nucleotide distance of 0.054 for both COX1 and CYTB, the genetic distance observed between the two

subspecies of aardwolf (*Proteles* ssp.) was higher than the majority of the intraspecific distances observed across Carnivora. However, with nucleotide distances of 0.020 for COX1 and 0.032 for CYTB, the genetic distances observed between the two subspecies of bat-eared fox (*Otocyon* ssp.) were clearly in the ambiguous zone and did not provide a clear indication of the specific taxonomic status of these populations.

Finally, to test whether the two pairs of allopatric subspecies diverged synchronously or in two different time periods, Bayesian molecular dating inferences were performed on the 142-taxon ML mitogenomic tree. The resulting divergence times were slightly different depending on the clock model used (strict clock [CL], autocorrelated [LN or TK02] and uncorrelated [UGAM or UCLM]) despite the convergence of the MCMC chains for all models. Cross-validation analyses resulted in the selection of the LN and UGAM models as the models with the best fit based on a higher cross-likelihood score than that of CL (LN and UGAM versus CL mean scores = 35.8). Unfortunately, these two statistically indistinguishable models provided different divergence times for the two pairs of subspecies, with LN favouring a synchronous divergence (approximately 1 Mya [95% credibility interval (CI) : 0.672 - 1.043]; **Supplementary File 2**), while UGAM favoured an asynchronous divergence (~0.6 [CI: 0.83 - 0.39] Mya for *O. megalotis* ssp. and ~1.3 [CI: 1.88 - 0.93] Mya for *P. cristatus* ssp.; **Supplementary File 2**). However, the three chains performed with the UGAM model recovered highly similar ages for the two nodes of interest with low CI 95% values, whereas the three chains performed with the LN model recovered less similar ages between chains and high CI 95% values (**Supplementary File 2**).





## Assembling reference genomes from roadkill

Considering the DNA quality and purity required to perform single-molecule sequencing with ONT, a specific protocol to extract DNA from roadkill was developed (Tilak et al., 2020). This protocol was designed to specifically select the longest DNA fragments present in the extract, which also contained short degraded fragments resulting from post-mortem DNA degradation processes. This protocol increased the median size of the sequenced raw DNA fragments three-fold in the case of aardwolf (Tilak et al., 2020). In total, after high-accuracy basecalling, adapter trimming, and quality filtering, 27.3 Gb of raw Nanopore long reads were sequenced using 16 MinION flow cells for the Southern aardwolf (*P. c. cristatus*) and 33.0 Gb using 13 flow cells for the Southern bat-eared fox (*O. m. megalotis*) (**Table 1**). Owing to quality differences among the extracted tissues for both species, the N50 of the DNA fragment size for *P. cristatus* (9,175 bp) was about two times higher than the N50 of the DNA fragment size obtained for *O. megalotis* (4,393 bp). The quality of the reads basecalled with the *high accuracy* option of Guppy was significantly higher than the quality of those translated with the *fast* option, which led to better assemblies (see **Appendix 1 – Figure 1**). Complementary Illumina sequencing returned 522.8 and 584.4 million quality-filtered reads per species corresponding to 129.5 Gb (expected coverage = 51.8x) and 154.8 Gb (expected coverage = 61.6x) for *P. c. cristatus* and *O. m. megalotis*, respectively. Regarding the resequenced individuals of each species, on average 153.5 Gb were obtained with Illumina resequencing (**Table 1**).

231 **Table 1.** Summary of sequencing and assembly statistics of the genomes generated in this study.

Individuals			Illumina				Oxford Nanopore Sequencing					Assembly statistics							
Species	Subspecies	Voucher	Raw reads (M)	Cleaned reads	Nbr of gigabases	Estimated coverage	Nbr of flowcells	Nbr of bases (Gb)	N50	Average size	Estimated coverage	Genome size (Gb)	Nbr of scaff.	N50 (kb)	Busco score	OMM genes	Missing data (%)		
<i>Proteles cristatus</i>	cristatus	TS307	716.7	522.8	129.50	51.8	16	27.3	9,175	5,555	10.9	2.39	5,669	1,309	92.8	12,062	22.43		
<i>Proteles cristatus</i>	cristatus	TS491	663.8	526.1	140.73	56.3	NA											NA	
<i>Proteles cristatus</i>	septentrionalis	NMSZ201854	750.9	516.2	132.44	53.0	NA											12,050	22.96
<i>Otocyon megalotis</i>	megalotis	TS305	710.2	584.4	154.81	61.6	13	33	4,393	3,092	13.2	2.75	11,081	728	92.9	11,981	22.02		
<i>Otocyon megalotis</i>	megalotis	TS306	861.2	820	240.71	96.3	NA											NA	
<i>Otocyon megalotis</i>	virgatus	FMNH158128	661.7	554.1	100.30	40.1	NA											NA	

The two reference genomes were assembled using MinION long reads and Illumina short reads in combination with MaSuRCA v3.2.9 (Zimin et al., 2013). Hybrid assemblies for both species were obtained with a high degree of contiguity with only 5,669 scaffolds and an N50 of 1.3 Mb for the aardwolf (*P. cristatus*) and 11,081 scaffolds and an N50 of 728 kb for the bat-eared fox (*O. megalotis*) (**Table 1**). Our two new genomes compared favourably with the available carnivoran genome assemblies in terms of (i) contiguity showing slightly less than the median N50 and a lower number of scaffolds than the majority of the other assemblies (**Appendix 1 – Figure 2, Supplementary File 3**) and (ii) completeness showing high BUSCO scores (see **Appendix 1 – Figure 3** and **Supplementary File 4** for BUSCO score comparisons among carnivoran genomes). Comparison of two hybrid assemblies with Illumina-only assemblies obtained with SOAPdenovo illustrated the positive effect of introducing Nanopore long reads even at moderate coverage by reducing the number of scaffolds from 409,724 to 5,669 (aardwolf) and from 433,209 to 11,081 (bat-eared fox), while increasing the N50 from 17.3 kb to 1.3 Mb (aardwolf) and from 22.3 kb to 728 kb (bat-eared fox).

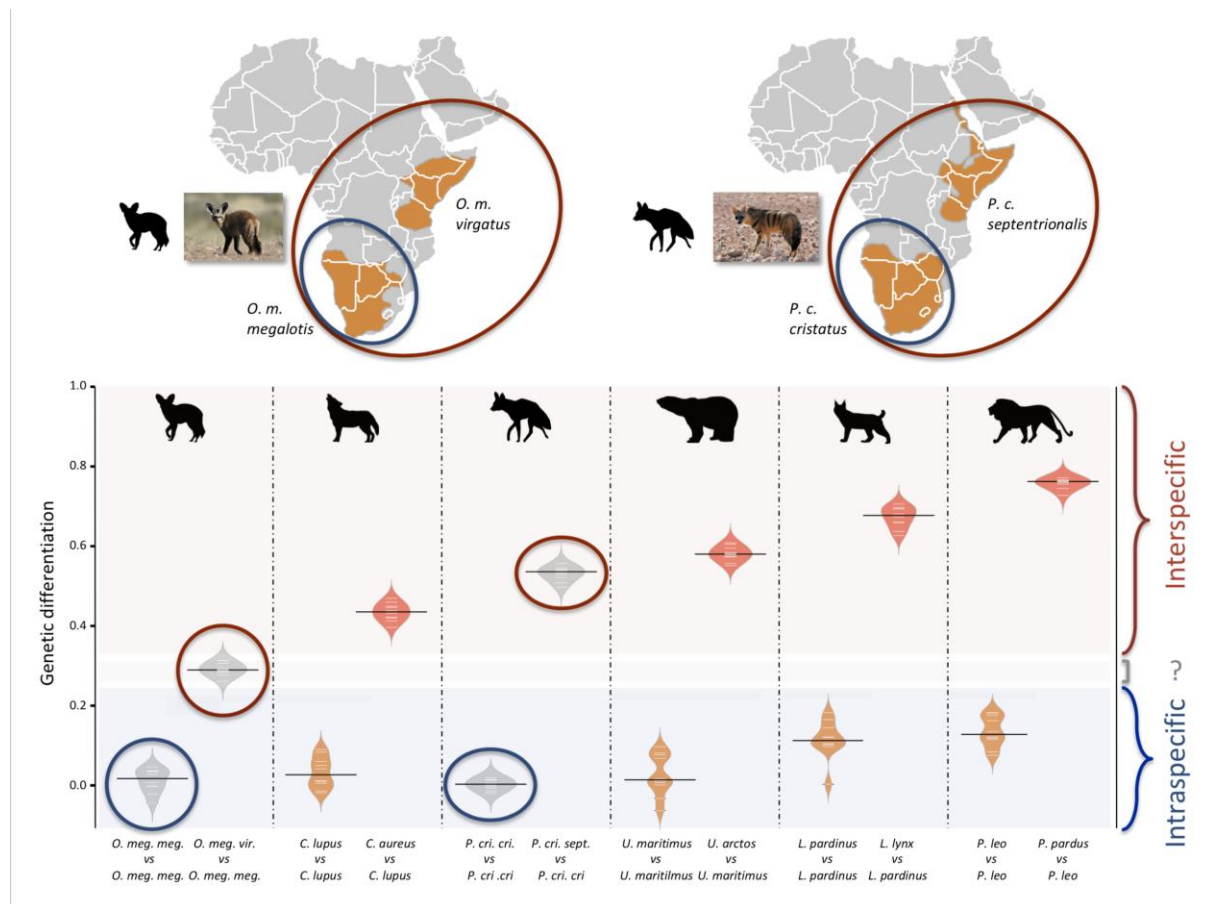
### ***Genome-wide analyses of population structure and differentiation***

To evaluate the population structure between the subspecies of *P. cristatus* and *O. megalotis*, the number of shared heterozygous sites, unique heterozygous sites, and homozygous sites between individuals was computed to estimate an  $F_{ST}$ -like statistic (hereafter called the *genetic differentiation index* or GDI). Since we were in possession of two individuals for the Southern subspecies and only one for the Eastern subspecies of both species, the genetic differentiation between the two individuals within the Southern subspecies and between the Southern and Eastern subspecies was computed. To account for the variation across the genome, 10 replicates of 100 regions with a length of 100 kb were randomly chosen to

estimate genetic differentiation. Interestingly, in both species the mean heterozygosity was higher in the Southern subspecies than in the Eastern subspecies. For the aardwolf the mean heterozygosity was 0.189 per kb (sd = 0.010) in the Southern population and 0.121 per kb (sd = 0.008) in the Eastern population. For the bat-eared fox the mean heterozygosity was 0.209 per kb (sd = 0.013) in the Southern population and 0.127 per kb (sd = 0.003) in the Eastern population. This heterozygosity level is low compared to that of other large mammals (Díez-del-Molino et al., 2018) and is comparable to that of the Iberian lynx, the cheetah or the brown hyaena, which have notoriously low genetic diversity (Abascal et al., 2016; Casas-Marce et al., 2013; Westbury et al., 2018).

Since we had very limited power to fit the evolution of the genetic differentiation statistics with a hypothetical demographic scenario because of our limited sample size ( $n = 3$ ), we chose a comparative approach and applied the same analyses to four well-defined species pairs of carnivorans, for which similar individual sampling was available. The genetic differentiation estimates between the two individuals belonging to the same subspecies (Southern populations in both cases) were on average equal to 0.005 and 0.014 for *P. c. cristatus* and *O. m. megalotis*, respectively. This indicated that the polymorphism observed in the two individuals within the Southern subspecies of each species was comparable (genetic differentiation index close to 0) and thus that these two subpopulations are likely panmictic (**Fig. 3 - Figure supplement 1**). In contrast, the genetic differentiation estimates for the two pairs of individuals belonging to the different subspecies were respectively equal to on average 0.533 and 0.294 for *P. cristatus* ssp. and *O. megalotis* ssp., indicating that the two disjunct populations are genetically structured. To contextualize these results, the same genetic differentiation measures were estimated using three individuals for four other well-defined species pairs (**Fig. 3 - Figure supplement 1**). First, the comparison of the polymorphism of two individuals of the same species led to intraspecific GDIs ranging from

0.029 on average for polar bear (*Ursus maritimus*) to 0.137 for lion (*Panthera leo*). As expected, comparing the polymorphisms of two individuals between closely related species led to a higher interspecific GDI ranging from 0.437 on average for the wolf/golden jackal (*Canis lupus/Canis aureus*) pair to 0.760 for the lion/leopard (*P. leo/Panthera pardus*) pair (**Fig. 3**). The genetic differentiation indices between the grey wolf (*C. lupus*) and the golden jackal (*C. aureus*) averaged 0.44, indicating that the two subspecies of aardwolf (GDI = 0.533) are genetically more differentiated than these two well-defined species, and only slightly less differentiated than the brown bear and the polar bear. Conversely, the genetic differentiation obtained between the bat-eared fox subspecies (GDI = 0.294) was lower than the genetic differentiation estimates obtained for any of the four reference species pairs evaluated here (**Fig. 3 - Figure supplement 1**). We verified that differences in depth-of-coverage among individuals did not bias our genetic differentiation estimates by subsampling reads at 15x (**Fig. 3 - Figure supplement 1**). We also checked that randomly sampling only three individuals was enough to accurately estimate genetic differentiation in the case of the brown vs. polar bear comparison (**Fig. 3 - Figure supplement 2**).



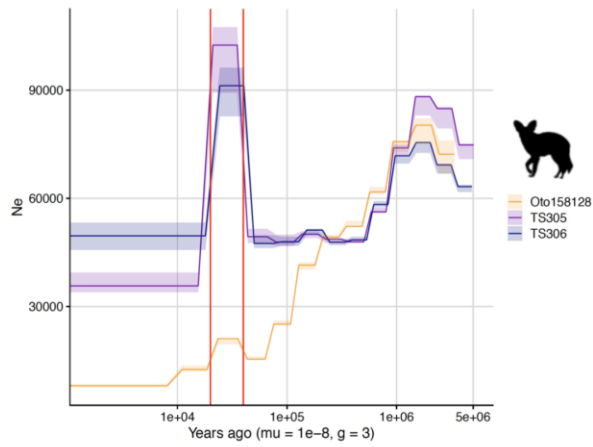
**Figure 3.** Genetic differentiation indices obtained from a comparison of intraspecific (orange) and interspecific (red) polymorphisms in four pairs of well-defined Carnivora species and for the subspecies of aardwolf (*Proteles cristatus*) and bat-eared fox (*Otocyon megalotis*) (grey). Silhouettes from <http://phylopic.org/>.

### *Effective population size reconstructions*

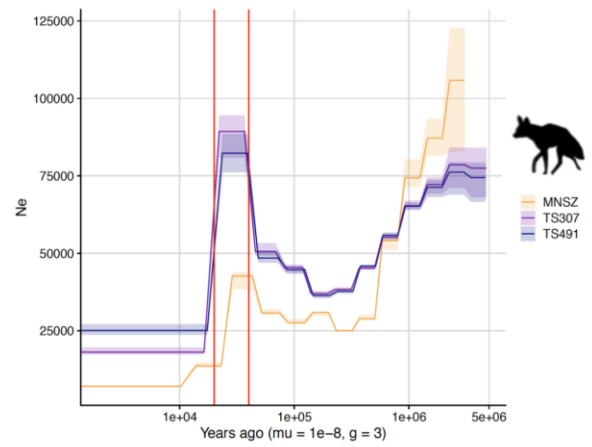
We used the pairwise sequential Markovian coalescent (PSMC) model to estimate the ancestral effective population size ( $N_e$ ) trajectory over time for each sequenced individual. For both the aardwolf and the bat-eared fox the individual from Eastern African populations showed a continuous decrease in  $N_e$  over time, leading to the recent  $N_e$  being lower than that in Southern African populations (**Fig. 4**). This is in agreement with the lower heterozygosity observed in the Eastern individuals of both species. For the bat-eared fox the trajectories of the three sampled individuals were synchronised approximately 200 kya ago (**Fig. 4a**), which could correspond to the time of divergence between the Southern and Eastern populations. In contrast,  $N_e$  trajectories for the aardwolf populations did not synchronise over the whole period (~2 Myrs). Interestingly, the Southern populations of both species showed a marked increase in population size between ~10-30 kya before sharply decreasing in more recent times (**Fig. 4**).



a) Bat-eared fox (*O. megalotis* ssp.)



b) Aardwolf (*P. cristatus* ssp.)



**Figure 4.** PSMC estimates of changes in effective population size over time for the Eastern (orange) and Southern (blue and purple) populations of a) bat-eared fox and b) aardwolf.  $\mu$  = mutation rate of  $10^{-8}$  mutations per site per generation and  $g$  = generation time of 2 years. Vertical red lines indicate 20 kyrs and 40 kyrs. Silhouettes from <http://phylopic.org/>.

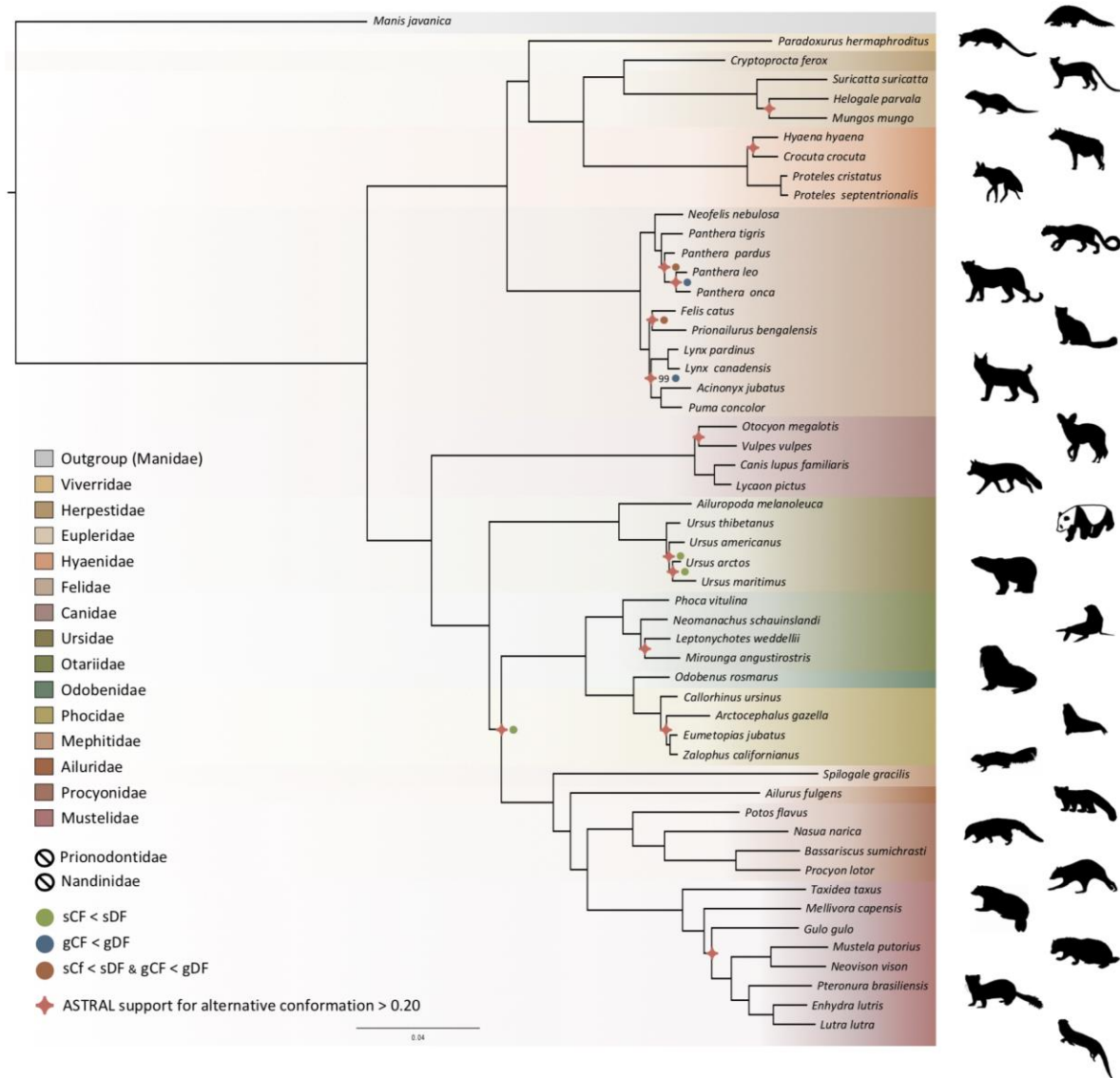
## Phylogenomics of the Carnivora

Phylogenetic relationships within the Carnivora were inferred from a phylogenomic dataset comprising 52 carnivoran species (including the likely new *Proteles septentrionalis* species), representing all but two families of the Carnivora (Nandiniidae and Prionodontidae). The non-annotated genome assemblies of these different species were annotated with a median of 18,131 functional protein-coding genes recovered for each species. Then, single-copy orthologous gene identification resulted in a median of 12,062 out of the 14,509 single-copy orthologues extracted from the OrthoMaM database for each species, ranging from a minimum of 6,305 genes for the California sea lion (*Zalophus californianus*) and a maximum of 13,808 for the dog (*Canis familiaris*) (**Supplementary File 5**). Our new hybrid assemblies allowed the recovery of 12,062 genes for the Southern aardwolf (*P. c. cristatus*), 12,050 for the Eastern aardwolf (*P. c. septentrionalis*), and 11,981 for the Southern bat-eared fox (*O. m. megalotis*) (**Table 1**). These gene sets were used to create a supermatrix consisting of 14,307 genes representing a total of 24,041,987 nucleotide sites with 6,495,611 distinct patterns (27.0%) and 22.8% gaps or undetermined nucleotides.

Phylogenomic inference was first performed on the whole supermatrix using ML. The resulting phylogenetic tree was highly supported, with all but one node being supported by maximum bootstrap (UFBS) values (**Fig. 5**). To further dissect the phylogenetic signal underlying this ML concatenated topology, we measured gene concordance (gCF) and site concordance (sCF) factors to complement traditional bootstrap node-support values. For each node, the proportion of genes (gCF) or sites (sCF) that supported the node inferred with the whole supermatrix was compared to the proportion of the genes (gDF) or sites (sDF) that supported an alternative resolution of the node (**Fig. 5**). Finally, a coalescent-based approximate species tree inference was performed using ASTRAL-III based on individual gene trees. Overall, the three different analyses provided well-supported and almost identical

350 results (**Fig. 5**). The order Carnivora was divided into two distinct suborders: a cat-related  
351 clade (Feliformia) and a dog-related clade (Caniformia). Within the Feliformia the first split  
352 separated the Felidae (felids) from the Viverroidea, a clade composed of the four families  
353 Viverridae (civets and genets), Eupleridae (fossa), Herpestidae (mongooses), and Hyaenidae  
354 (hyaenas). In hyaenids the two species of termite-eating aardwolves (*P. cristatus* and *P.*  
355 *septrionalis*) were the sister-group of a clade composed of the carnivorous spotted  
356 (*Crocuta crocuta*) and striped (*Hyaena hyaena*) hyaenas. Congruent phylogenetic  
357 relationships among Feliformia families and within hyaenids were also retrieved with the  
358 mitogenomic data set (**Fig. 2a**). The short internal nodes of the Felidae were the principal  
359 source of incongruence among the three different analyses with concordance factor analyses  
360 pointing to three nodes for which many sites and genes support alternative topologies (**Fig.**  
361 **5**), including one node for which the coalescent-based approximate species tree inference  
362 supported an alternative topology to the one obtained with ML on the concatenated  
363 supermatrix. In the Viverroidea the Viverridae split early from the Herpestoidea, regrouping  
364 the Hyaenidae, Herpestidae, and Eupleridae, within which the Herpestidae and Eupleridae  
365 formed a sister clade to the Hyaenidae. Within the Caniformia the Canidae (canids) was  
366 recovered as a sister group to the Arctoidea. Within the Canidae, in accordance with the  
367 mitogenomic phylogeny, the Vulpini tribe, represented by *O. megalotis* and *V. vulpes*, was  
368 recovered as the sister clade of the Canini tribe, represented here by *Lycaon pictus* and *C.*  
369 *familiaris*. The Arctoidea was recovered as a major clade composed of eight families grouped  
370 into three subclades: Ursoidea (Ursidae), Pinnipedia (Otariidae, Odobedinae, and Phocidae),  
371 and Musteloidea, composed of Ailuridae (red pandas), Mephitidae (skunks), Procyonidae  
372 (raccoons), and Mustelidae (badgers, martens, weasels, and otters). Within the Arctoidea the  
373 ML phylogenetic inference on the concatenation provided support for grouping the  
374 Pinnipedia and the Musteloidea to the exclusion of the Ursidae (bears) with maximum

bootstrap support (**Fig. 5**), as in the mitogenomic tree (**Fig. 2a**). However, the concordance factor analyses revealed that many sites and many genes actually supported alternative topological conformations for this node characterised by a very short branch length (sCF = 34.1, SDF1 = 29.2, sDF2 = 36.7, gCF = 46.9, gDF1 = 18.6, gDF2 = 18.2, gDFP = 16.3) (**Fig. 5**). In the Pinnipedia the clade Odobenidae (walruses) plus Otariidae (eared seals) was recovered to the exclusion of the Phocidae (true seals), which was also in agreement with the mitogenomic scenario (**Fig. 2a**). Finally, within the Musteloidea the Mephitidae represented the first offshoot, followed by the Ailuridae, and a clade grouping the Procyonidae and the Mustelidae. Phylogenetic relationships within Musteloidea were incongruent with the mitogenomic tree, which alternatively supported the grouping of the Ailuridae and the Mephitidae (**Fig. 2a**).



**Figure 5.** Phylogenomic tree reconstructed from the nucleotide supermatrix composed of 14,307 single-copy orthologous genes for 52 species of Carnivora plus one outgroup (*Manis javanica*). The family names in the legend are ordered as in the phylogeny. Silhouettes from <http://phylopic.org/>.

## Discussion

### *High-quality mammalian genomes from roadkill using MaSuRCA hybrid assembly*

With an increasing number of species being threatened worldwide, obtaining genomic resources from mammalian wildlife can be difficult. We decided to test the potential of using roadkill samples, an abundant and valuable resource in ecological studies (Schwartz et al., 2020) but a currently underexploited source material for genomics (Etherington et al., 2020; Maigret, 2019). Roadkill are indeed relatively easy to survey and the potential coordination with ongoing monitoring and citizen science projects (*e.g.* Périquet et al., 2018; Waetjen and Shilling, 2017) could potentially give access to large numbers of tissue samples for frequently encountered species. Even though roadkill may represent a biased sample of species populations (Brown and Bomberger Brown, 2013; Loughry and McDonough, 1996), they can also be relevant to generate reference genomes for elusive species that could hardly be sampled otherwise. Despite limited knowledge and difficulties associated with *de novo* assembly of non-model species (Etherington et al., 2020), we designed a protocol to produce DNA extracts of suitable quality for Nanopore long-read sequencing from roadkill (Tilak et al., 2020). Additionally, we tested the impact of the accuracy of the MinION base-calling step on the quality of the resulting MaSuRCA hybrid assemblies. In line with previous studies (Wenger et al., 2019; Wick et al., 2019) we found that using the *high accuracy* option rather than the *fast* option of Guppy 3.1.5 leads to more contiguous assemblies by increasing the N50 value. By relying on this protocol, we were able to generate two hybrid assemblies by combining Illumina reads at relatively high coverage (80x) and MinION long reads at relatively moderate coverage (12x), which provided genomes with high contiguity and completeness. These represent the first two mammalian genomes obtained with such a hybrid Illumina/Nanopore approach using the MaSuRCA assembler for non-model carnivoran species: the aardwolf (*P. cristatus*) and the bat-eared fox (*O. megalotis*). Despite the use of

roadkill samples our assemblies compare favourably, in terms of both contiguity and completeness, with the best carnivoran genomes obtained so far from classical genome sequencing approaches that do not rely on complementary optical mapping or chromatin conformation approaches. Overall, our carnivoran hybrid assemblies are fairly comparable to those obtained using the classic Illumina-based genome sequencing protocol involving the sequencing of both paired-end and mate-paired libraries (Li et al., 2010). The benefit of adding Nanopore long reads is demonstrated by the fact that our hybrid assemblies are of better quality than all the draft genome assemblies generated using the DISCOVAR *de novo* protocol based on a PCR-free single Illumina 250 bp paired-end library (Weisenfeld et al., 2014; DISCOVAR) used in the 200 Mammals Project of the Broad Genome Institute (Zoonomia consortium, 2020). These results confirm the capacity of the MaSuRCA hybrid assembler to produce quality assemblies for large and complex genomes by leveraging the power of long Nanopore reads (Wang et al., 2020). Moreover, these two hybrid assemblies could form the basis for future chromosome-length assemblies by adding complementary HiC data (van Berkum et al., 2010) as proposed in initiatives such as the Vertebrate Genome Project (Koepfli et al., 2015) and the DNA Zoo (Dudchenko et al., 2017). Our results demonstrate the feasibility of producing high-quality mammalian genome assemblies at moderate cost (\$5,000-10,000 USD for each of our Carnivora genomes) using roadkill and should encourage genome sequencing of non-model mammalian species in ecology and evolution laboratories.

#### ***Genomic evidence for two distinct species of aardwolves***

The mitogenomic distances inferred between the subspecies of *O. megalotis* and *P. cristatus* were comparable to those observed for other well-defined species within the Carnivora. Furthermore, by comparing the genetic diversity between several well-defined species

(divergence) and several individuals of the same species (polymorphism) based on the COX1 and CYTB genes across Carnivora, we were able to pinpoint a threshold of approximately 0.02 substitutions per base separating divergence from polymorphism, which is in accordance with a recent study of naturally occurring hybrids in Carnivora (Allen et al., 2020). This method, also known as the barcoding-gap method (Meyer and Paulay, 2005), allowed us to show that the two subspecies of *P. cristatus* present a genetic divergence greater than the threshold, whereas the divergence is slightly lower for the two subspecies of *O. megalotis*. These results seem to indicate that the subspecies *P. c. septentrionalis* should be elevated to species level (*P. septentrionalis*). Conversely, for *O. megalotis*, this first genetic indicator seems to confirm the distinction at the subspecies level. However, mitochondrial markers have some well-identified limitations (Galtier et al., 2009), and it is difficult to properly determine a threshold between polymorphism and divergence across the Carnivora. The measure of mtDNA sequence distances can thus be seen only as a first useful indicator for species delineation. The examination of variation at multiple genomic loci in a phylogenetic context, combined with morphological, behavioural and ecological data, is required to establish accurate species boundaries.

The newly generated reference genomes allowed us to perform genome-wide evaluation of the genetic differentiation between subspecies using short-read resequencing data of a few additional individuals of both species. Traditionally, the reduction in polymorphism in two subdivided populations (*p within*) compared to the population at large (*p between*) is measured with several individuals per population ( $F_{ST}$ ; Hudson et al., 1992). However, given that the two alleles of one individual are the results of the combination of two *a priori* non-related individuals of the population (*i.e.*, the parents), with a large number of SNPs, the measurement of heterozygosity can be extended to estimation of the (sub)population polymorphism. Furthermore, in a panmictic population with recombination



along the genome, different chromosomal regions can be considered to be independent and can be used as replicates for heterozygosity estimation. In this way, genome-wide analyses of heterozygosity provide a way to assess the level of polymorphism in a population and a way to compare genetic differentiation between two populations. If we hypothesize that the two compared populations are panmictic, picking one individual or another of the population has no effect (*i.e.*, there is no individual with excess homozygous alleles due to mating preference across the population), and the population structure can be assessed by comparing the heterozygosity of the individuals of each population compared to the heterozygosity observed for two individuals of the same population (see *Methods*). Such an index of genetic differentiation, by measuring the level of population structure, could provide support to establish accurate species boundaries. In fact, delineating species has been and still is a complex task in evolutionary biology (Galtier, 2019; Ravinet et al., 2016; Roux et al., 2016). Given that accurately defining the species taxonomic level is essential for a number of research fields, such as macroevolution (Faurby et al., 2016) or conservation (Frankham et al., 2012), defining thresholds to discriminate between populations or subspecies in different species is an important challenge in biology. However, due to the disagreement on the definition of species, the different routes of speciation observed *in natura* and the different amounts of data available among taxa, adapting a standardised procedure for species delineation seems complicated (Galtier, 2019).

As proposed by Galtier (2019), we decided to test the taxonomic level of the *P. cristatus* and *O. megalotis* subspecies by comparing the genetic differentiation observed between Eastern and Southern populations within these species to the genetic differentiation measured for well-defined Carnivora species. Indeed, estimation of the genetic differentiation either within well-defined species (polymorphism) or between two closely related species (divergence) allowed us to define a threshold between genetic polymorphism and genetic

divergence across the Carnivora (**Fig. 5**). Given these estimates, and in accordance with mitochondrial data, the two subspecies of *P. cristatus* (1) present more genetic differentiation between each other than the two well-defined species of golden jackal (*Canis aureus*) and wolf (*C. lupus*), and (2) present more genetic differentiation than the more polymorphic species of the dataset, the lion (*P. leo*). Despite known cases of natural hybridisation reported between *C. aureus* and *C. lupus* (Galov et al., 2015; Gopalakrishnan et al., 2018), the taxonomic rank of these two species is well accepted. In that sense, given the species used as a reference, both subspecies of *P. cristatus* seem to deserve to be elevated to species level. The situation is less clear regarding the subspecies of *O. megalotis*. Indeed, while the genetic differentiation observed between the two subspecies is significantly higher than the polymorphic distances observed for all the well-defined species of the dataset, there is no species in our dataset that exhibits equivalent or lower genetic divergence than a closely related species. This illustrates the limits of delineating closely related species due to the continuous nature of the divergence process (De Queiroz, 2007). The subspecies of *O. megalotis* fall into the “grey zone” of the speciation continuum (De Queiroz, 2007; Roux et al., 2016) and are likely undergoing speciation due to their vicariant distributions. To be congruent with the genetic divergence observed across closely related species of the Carnivora (according to our dataset), we thus propose that (1) the taxonomic level of the *P. cristatus* subspecies be reconsidered by elevating the two subspecies *P. c. cristatus* and *P. c. septentrionalis* to species level, and (2) the taxonomic level for the two subspecies of *O. megalotis* be maintained.

Although there is a distinct genetic difference between Eastern and Southern aardwolves, the evidence for a clear morphological difference is less obvious (**Fig. 6, Appendix 2 – Figure 1-3, Supplementary File 6-7**). The earliest available name for the East African aardwolf subspecies is *P. c. septentrionalis* (Rothschild, 1902). This subspecies was

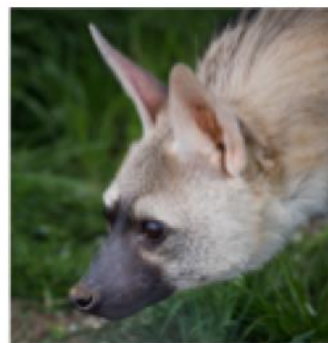
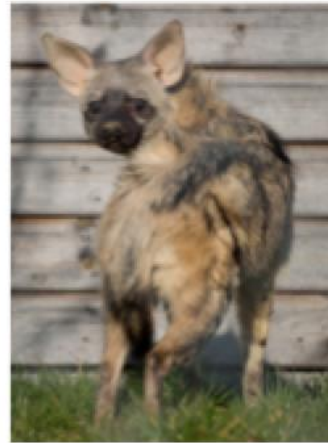
519 first distinguished based on pelage characteristics of a specimen from Somaliland, which has  
520 a creamy white pelage without any grey tinge, but washed slightly with buff in the neck and  
521 side of the rump (Rothschild, 1902). Also, the striping pattern is less well defined and breaks  
522 up into spots on the neck. In contrast, the Southern aardwolf subspecies *P. c. cristatus* was  
523 described as ashy grey, front and sides of neck greyish white, black stripes broad and well  
524 defined (Rothschild, 1902). Drake-Brockman (1910) also described Somali aardwolves as  
525 pale buff with a dark greyish-buff head, but Cabrera (1910) was the first to ascribe diagnostic  
526 characters to distinguish between the Eastern and Southern populations. He described a new  
527 subspecies *P. c. pallidior* from Suakin (Sudan) as a very pale yellowish cream, almost white  
528 ventrally and on the forehead. This contrasts with the grizzled grey of the forehead of *P. c.*  
529 *cristatus* (**Fig. 6**). Cabrera (1910) also described how the fur of *P. c. pallidior* is unicoloured  
530 and lacks the brown base of *P. c. cristatus*. This latter character appears to be consistent in an  
531 Ethiopian specimen compared with three skins of Namibian and South African origin in the  
532 collections of National Museums Scotland, although it would appear to be a difference in the  
533 coloration of the underfur. However, a further specimen from Zimbabwe also has pale  
534 underfur. In reviewing georeferenced photographs of aardwolves from throughout the range,  
535 the striping pattern appeared to be variable, but overall East African specimens tended to be  
536 paler, with more contrasting stripes with a pale forehead compared with the longer, greyer or  
537 ochre-grey fur in Southern African specimens, which have less distinctive stripes (A.C.K.  
538 pers. obs.). However, fur length and hence stripe distinctiveness may just be a phenotypic  
539 response to lower temperatures at higher latitudes compared with equatorial East African  
540 specimens. Cabrera (1910) also proposed differences in a skull measurement between Eastern  
541 and Southern African aardwolves. Three specimens from Eastern Africa had a wider inter-  
542 orbital breadth than two from Southern Africa. However, his measurements also showed that  
543 Eastern African aardwolves have larger postorbital breadths, brain case widths, and maxillary

544 widths at the canines. Adding in measurements of skulls from the literature (Allen et al.,  
545 1909; Heller, 1913; Hollister, 1918; Roberts, 1951, 1932) confirmed that postorbital breadth  
546 is significantly greater in *P. c. septentrionalis* than *P. c. cristatus* but revealed no significant  
547 differences between other skull measurements including condylobasal length of skull  
548 (**Appendix 2 – Figure 2-3, Supplementary File 7**). However, as noted above from skins,  
549 sample sizes are very limited and thus these morphological differences remain tentative  
550 subject to examination of a much larger sample with more powerful geometric  
551 morphometrics methods. These preliminary observations should nevertheless prompt a deeper  
552 investigation of morphological and behavioural differences that have been reported between  
553 the two proposed subspecies of aardwolf to formally validate our newly proposed taxonomic  
554 arrangement. Our results might also have conservation implications, as the status of the two  
555 distinct aardwolf species will have to be re-evaluated separately in the International Union for  
556 Conservation of Nature (IUCN) Red List of Threatened Species (IUCN, 2020).

Southern aardwolf  
(*Proteles cristatus*)



Eastern aardwolf  
(*Proteles septentrionalis*)



557

558 **Figure 6.** Phenotypic comparisons, highlighting the differences in fur coloration and stripe pattern, between  
559 captive individuals of Eastern (*P. septentrionalis*) and Southern (*P. cristatus*) aardwolves held at Hamerton Zoo  
560 Park (UK). All pictures copyright and used with permission from Robb Cadd.

561

562 *Population size variation and environmental change*

563 The Pairwise Sequentially Markovian Coalescent (PSMC) analyses revealed that the  
564 Southern and Eastern African populations have different effective population-size estimates  
565 over time, confirming that they have been genetically isolated for several thousand years,  
566 which is more so for the aardwolf than for the bat-eared fox. This supports the hypothesis of  
567 two separate events leading to the same disjunct distributions for the two taxa, in accordance  
568 with mitochondrial dating. Nevertheless, the population trends are rather similar and are  
569 characterized by continuous declines between 1 Mya and 100-200 kya that are followed by an  
570 increase that is much more pronounced in the Southern populations of both species between  
571 30-10 kya. The similar trajectories exhibited by both species suggest that they were under the  
572 influence of similar environmental factors, such as climate and vegetation variations.

573 Aardwolves and bat-eared foxes live in open environments including short-grass  
574 plains, shrubland, and open and tree savannas, and both are highly dependent on herbivorous  
575 termites for their diet. Therefore, the fluctuation of their populations could reflect the  
576 evolution of these semi-arid ecosystems determining prey abundance during the last million  
577 years. However, the global long-term Plio-Pleistocene African climate is still debated. For  
578 Eastern Africa, some studies have suggested an evolution towards increased aridity  
579 (deMenocal, 2004, 1995), whereas others have proposed the opposite (Grant et al., 2017;  
580 Maslin et al., 2014; Trauth et al., 2009). Therefore, our data support the latter hypothesis, as a  
581 global long-term tendency towards a wetter climate in East Africa could have been less  
582 favourable for species living in open environments.

583 Southern populations exhibit a similar decreasing trend between 1 Mya and 100 kya.  
584 Once again, the relevant records appear contradictory. This could be the result of regional  
585 variation across South Africa, with aridification in the Southwestern part and wetter

conditions in the Southeast (Caley et al., 2018; Johnson et al., 2016). Finally, the 30-10 kya period appears to have been more humid (Chase et al., 2019; Chevalier and Chase, 2015; Lim et al., 2016). This seems inconsistent with the large population increase detected in Southern populations of both species; however, the large regions of the Namib Desert that are currently unsuitable could have been more favourable in wetter conditions.

The global decrease in population size detected in the Southern and Eastern populations could also reflect the fragmentation of a continuous ancestral range. The global trend towards a wetter climate may have favoured the development of the tropical rainforest in central Africa, creating a belt of unsuitable habitat. This is in line with previous studies describing diverse biogeographical scenarios involving the survival and divergence of ungulate populations in isolated savanna refuges during Pleistocene climatic oscillations (Lorenzen et al., 2012). In this respect, it could be interesting to study population trends in other species living in semi-arid environments and having a similar range as disconnected populations. Interestingly, several bird species also have similar distributions including the Orange River francolin (*Scleroptila gutturalis*), the greater kestrel (*Falco rupicoloides*), the double-banded courser (*Smutsornis africanus*), the red-fronted tinkerbird (*Pogoniulus pusillus*), the Cape crow (*Corvus capensis*) and the black-faced waxbill (*Estrilda erythronotos*), supporting the role of the environment in the appearance of these disjunct distributions. Finally, these new demographic results, showing recent population size declines in both regions in both species, might be taken into account when assessing the conservation status of the two distinct aardwolf species and bat-eared fox subspecies.

### ***Genome-scale phylogeny of Carnivora***

In this study, we provide a new phylogeny of Carnivora including the newly recognized species of aardwolf (*P. septentrionalis*). The resulting phylogeny is fully resolved with all

nodes supported with UFBS values greater than 95% and is congruent with previous studies (Doronina et al., 2015; Eizirik et al., 2010) (**Fig. 5**). Across the Carnivora the monophyly of all superfamilies are strongly supported (Flynn et al., 2010) and are divided into two distinct suborders: a cat-related clade (Feliformia) and a dog-related clade (Caniformia). On the one hand, within the Feliformia, the different families and their relative relationships are well supported and are in accordance with previous studies (Eizirik et al., 2010). There is one interesting point regarding the Felidae. While almost all the nodes of the phylogeny were recovered as strongly supported from the three phylogenetic inference analyses (ML inferences, concordance factor analyses and coalescent-based inferences), one third of the nodes (3 out of 9) within the Felidae show controversial node supports. This result is not surprising and is consistent with previous studies arguing for ancient hybridisation among the Felidae (Li et al., 2019, 2016). Another interesting point regarding the Feliformia and particularly the Hyaenidae is the relationship of the two aardwolves. The two species, *P. cristata* and *P. septentrionalis* form a sister clade to the clade composed of the striped hyaena (*H. hyaena*) and the spotted hyaena (*C. crocuta*), in accordance with previous studies (Koepfli et al., 2006; Westbury et al., 2018) and the two subfamilies Protelinae and Hyaeninae that have been proposed for these two clades, respectively. However, although the phylogenetic inferences based on the supermatrix of 14,307 single-copy orthologues led to a robust resolution of this node according to the bootstrap supports, both concordance factors and coalescent-based analyses revealed conflicting signals with support for alternative topologies. In this sense, the description and acceptance of the Hyaeninae and Protelinae subfamilies still require further analyses, including genomic data for the brown hyaena (*Parahyena brunnea*) (Westbury et al., 2018).

On the other hand, within the Caniformia, the first split separates the Canidae from the Arctoidea. Within the Canidae the bat-eared fox (*O. megalotis*) is grouped with the red



636 fox (*Vulpes vulpes*) and the other representative of the Vulpini, but with a very short branch,  
637 and concordance analyses indicate conflicting signals on this node. Regarding the Arctoidea,  
638 historically the relationships between the three superfamilies of arctoids have been  
639 contradictory and debated. The least supported scenario from the literature is that in which  
640 the clade Ursoidea/Musteloidea is a sister group of the Pinnipedia (Flynn and Nedbal, 1998).  
641 Based on different types of phylogenetic characters, previous studies found support for both  
642 the clade Ursoidea/Pinnipedia (Agnarsson et al., 2010; Meredith et al., 2011; Rybczynski et  
643 al., 2009) and the clade Pinnipedia/Musteloidea (Arnason et al., 2007; Eizirik et al., 2010;  
644 Flynn et al., 2005; Sato et al., 2009, 2006; Schröder et al., 2009). However, investigations of  
645 the insertion patterns of retroposed elements revealed the occurrence of incomplete lineage  
646 sorting (ILS) at this node (Doronina et al., 2015). With a phylogeny inferred from 14,307  
647 single-copy orthologous genes, our study, based on both gene trees and supermatrix  
648 approaches, gives support to the variant Pinnipedia/Musteloidea excluding the Ursoidea as  
649 the best supported conformation for the Arctoidea tree (Doronina et al., 2015; Eizirik et al.,  
650 2010; Sato et al., 2006). Interestingly, in agreement with Doronina et al. (2015), our  
651 concordance factor analysis supports the idea that the different conformations of the  
652 Arctoidea tree are probably due to incomplete lineage sorting by finding almost the same  
653 number of sites supporting each of the three conformations (34.11%, 29.61% and 36.73%).  
654 However, although trifurcation of this node is supported by these proportions of sites, a  
655 majority of genes taken independently (gene concordance factors: 6,624 out of 14,307 genes)  
656 and the coalescent-based species tree approach (quartet posterior probabilities  $q_1 = 0.53$ ,  $q_2 =$   
657  $0.24$ ,  $q_3 = 0.24$ ) support the clade Pinnipedia/Musteloidea, excluding the Ursoidea.  
658 Considering these results, the difficulty of resolving this trifurcation among the Carnivora  
659 (Delisle and Strobeck, 2005) has likely been contradictory due to the ILS observed among  
660 these three subfamilies (Doronina et al., 2015), which led to different phylogenetic scenarios

depending on the methods (Peng et al., 2007) or markers (Yu and Zhang, 2006) used. Another controversial point, likely due to ILS (Doronina et al., 2015) within the Carnivora, is the question regarding which of the Ailuridae and Mephitidae is the most basal family of the Musteloidea (Doronina et al., 2015; Eizirik et al., 2010; Flynn et al., 2005; Sato et al., 2009). Interestingly, our phylogenetic reconstruction based on mitogenomic data recovered the clade Ailuridae/Mephitidae as a sister clade to all other Musteloidea families. The phylogenomic inferences based on the genome-scale supermatrix recovered the Mephitidae as the most basal family of the Musteloidea. This result is supported by both coalescent-based inferences and concordance factors. In that sense, despite incomplete lineage sorting (Doronina et al., 2015), at the genomic level, it seems that the Mephitidae is the sister-group to all other Musteloidea families.

Overall, the phylogenomic inference based on 14,307 single-copy orthologous genes provides a new vision of the evolution of Carnivora. The addition of information from both concordance factor analyses (Minh et al., 2020) and coalescent-based inference (Zhang et al., 2018) supports previous analyses showing controversial nodes in the Carnivora phylogeny. Indeed, this additional information seems essential in phylogenomic analyses based on thousands of markers, which can lead to highly resolved and well-supported phylogenies despite support for alternative topological conformations for controversial nodes (Allio et al., 2020b; Jeffroy et al., 2006; Kumar et al., 2012).

## Conclusions

The protocol developed here to extract the best part of the DNA from roadkill samples provides a good way to obtain genomic data from wildlife. Combining Illumina sequencing data and Oxford Nanopore long-read sequencing data using the MaSuRCA hybrid assembler allowed us to generate high-quality reference genomes for the Southern aardwolf (*P. c.*

*cristatus*) and the Southern bat-eared fox (*O. m. megalotis*). This cost-effective strategy provides opportunities for large-scale population genomic studies of mammalian wildlife using resequencing of samples collected from roadkill and opportunistic field collection. Indeed, by defining a genetic differentiation index based on only three individuals, we illustrated the potential of the approach for comparative genome-scale species delineation in both species for which subspecies have been defined based on disjunct distributions and morphological differences. Our results, based on both mitochondrial and nuclear genome analyses, indicate that the two subspecies of aardwolf warrant elevation to species level (*P. cristatus* and *P. septentrionalis*), but the *O. megalotis* subspecies do not warrant this status. Hence, by generating reference genomes with high contiguity and completeness, this study shows a practical application for genomics of roadkill samples.

## Methods

### Biological samples

We conducted fieldwork in the Free State province of South Africa in October 2016 and October 2018. While driving along the roads, we opportunistically collected tissue samples from four roadkill specimens, from which we sampled ear tissue preserved in 95% ethanol: two Southern bat-eared foxes (*O. megalotis megalotis* NMB 12639, GPS: 29°1'52"S, 25°9'38"E and NMB 12640, GPS: 29°2'33"S, 25°10'26"E), and two Southern aardwolves (*P. cristatus cristatus* NMB 12641, GPS: 29°48'45"S, 26°15'0"E and NMB 12667, GPS: 29°8'42"S, 25°39'4"E). As aardwolf specimen NMB 12641 was still very fresh, we also sampled muscle and salivary gland and preserved them in RNAlater™ stabilization solution (Thermo Fisher Scientific). These roadkill specimens were sampled under standing permit number S03016 issued by the Department of National Affairs in Pretoria (South Africa) granted to the National Museum, Bloemfontein. These samples have been sent to France

under export permits (JM 3007/2017 and JM 5042/2018) issued by the Free State Department of Economic, Small Business Development, Tourism and Environmental Affairs (DESTEA) in Bloemfontein (Free State, South Africa). All tissue samples collected in this study have been deposited in the mammalian tissue collection of the National Museum, Bloemfontein (Free State, South Africa). Additional tissue samples for an Eastern aardwolf (*P. c. septentrionalis*) male neonate (NMS.Z.2018.54) stillborn from Tanzanian parents in 2015 at Hamerton Zoo Park (UK) have been provided by the National Museums Scotland (Edinburgh, UK), and for an Eastern bat-eared fox (*O. m. virgatus*) from Tanzania (FMNH 158128) by the Field Museum of Natural History (Chicago, USA). As these two species are classified as Least Concern by the IUCN, and thus do not require CITES permits for international transport, the samples were transferred to France under import permits issued by the Direction régionale de l'environnement, de l'aménagement et du logement (DREAL) Occitanie in Toulouse (France).

## **Mitochondrial barcoding and phylogenetics**

### *Mitogenomic dataset construction*

In order to assemble a mitogenomic data set for assessing mitochondrial diversity among *P. cristatus* and *O. megalotis* subspecies, we generated seven new Carnivora mitogenomes using Illumina shotgun sequencing (**Supplementary File 8**). Briefly, we extracted total genomic DNA total using the DNeasy Blood and Tissue Kit (Qiagen) for *P. c. cristatus* (NMB 12641), *P. c. septentrionalis* (NMS Z.2018.54), *O. m. megalotis* (NMB 12639), *O. m. virgatus* (FMNH 158128), *Speothos venaticus* (ISEM T1624), *Vulpes vulpes* (ISEM T3611), and *Parahyaena brunnea* (ISEM FD126), prepared Illumina libraries following the protocol of Tilak et al. (2015), and sent libraries to the Montpellier GenomiX platform for single-end 100 bp sequencing on a Illumina HiSeq 2500 instrument to obtain about 5 to 10 million reads per

sample. We then assembled and annotated mitogenomes from these single-read shotgun sequencing data with MitoFinder v1.0.2 (Allio et al., 2020a) using default parameters. We also used MitoFinder to extract three additional mitogenomes from paired-end Illumina capture libraries of ultra-conserved elements (UCEs) and available from the Short Read Archive (SRA) of NCBI for *Viverra zangalunga*, *Bdeogale nigripes*, and *Fossa fossana*. Additional read mappings were done with Geneious (Kearse et al., 2012) to close gaps when the mitochondrial genome was fragmented. Finally, we downloaded all RefSeq carnivoran mitogenomes available in Genbank (135 species as of July 1<sup>st</sup>, 2019) and the mitogenome of the Malayan pangolin (*Manis javanica*) to use as an outgroup.

#### *Mitogenomic phylogenetics and dating*

Mitochondrial protein-coding genes were individually aligned using MACSE v2 (Ranwez et al., 2018) with default parameters, and ribosomal RNA genes using MAFFT (Katoh and Standley, 2013) algorithm FFT-NS-2 with option *--adjustdirection*. A nucleotide supermatrix was created by concatenating protein-coding and ribosomal RNA genes for the 142 taxa (140 species and two subspecies). Phylogenetic inferences were performed with Maximum likelihood (ML) as implemented in IQ-TREE 1.6.8 (Nguyen et al., 2015) with the GTR+G4+F model. Using the resulting topology, divergence time estimation was performed using Phylobayes v4.1c (Lartillot et al., 2013) with strict clock (CL), autocorrelated (LN or TK02), and uncorrelated (UGAM or UCLM) models combined with 18 fossil calibrations (**Supplementary File 9**). Three independent Markov chains Monte Carlo (MCMC) analyses starting from a random tree were run until 10,000 generated cycles with trees and associated model parameters sampled every cycle. A burn-in of 25% was applied before constructing the majority-rule Bayesian consensus tree with the *readdiv subprogram*. Finally, to determine the best-fitting clock model, cross-validation analyses were performed with Phylobayes by splitting the dataset randomly into two parts. Then, parameters of one model were estimated

on the first part of the dataset (here representing 90%) and the parameter values were used to compute the likelihood of the second part of the dataset (10%). This procedure was repeated ten times for each model. Finally, the likelihood of each repeated test was computed and summed for each model with the *readcv* and *sumcv* subprograms, respectively. The molecular clock model with the highest cross-likelihood scores was considered as the best fitting.

#### *Mitochondrial diversity and barcoding gap analyses*

To check if a threshold between intraspecific variation and interspecific divergence could be determined across the Carnivora (Meyer and Paulay, 2005), two mitochondrial barcoding datasets were assembled from all COX1 and CYTB sequences available for Carnivora plus the corresponding sequences for each of the two subspecies of *O. megalotis* and *P. cristatus*, respectively. After aligning each barcoding dataset with MACSE v2, ML phylogenetic inferences were performed with IQ-TREE 1.6.6 using the optimal substitution model as determined by ModelFinder (Kalyaanamoorthy et al., 2017). Then, pairwise patristic distances between all individuals were calculated from the resulting ML phylogram. Finally, based on the actual taxonomic assignment, patristic distances were considered as intraspecific variation between two individuals belonging to the same species and as interspecific divergence between individuals of different species.

### **Short reads and long reads hybrid assembly of reference genomes**

#### *Sampling*

To construct reference assemblies with high contiguity for the two focal species we selected the best-preserved roadkill samples: NMB 12639 for *O. megalotis* and NMB 12641 for *P. cristatus* (**Table 1, Supplementary File 8**). Total genomic DNA extractions were performed separately for Illumina short-read sequencing and MinION long-read sequencing.

#### *Illumina short-read sequencing*

Total genomic DNA extractions were performed from ear tissue samples from two individuals using the DNeasy Blood and Tissue Kit (Qiagen) following manufacturer's instructions. A total amount of 1.0µg DNA per sample was sent as input material for Illumina library preparation and sequencing to Novogene Europe (Cambridge, UK). Sequencing libraries were generated using NEBNext® DNA Library Prep Kit following manufacturer's recommendations and indices were added to each sample. Genomic DNA was randomly fragmented to a size of 350 bp by shearing, then DNA fragments were end-polished, A-tailed, and ligated with the NEBNext adapter for Illumina sequencing, and further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified (AMPure XP system) and the resulting libraries were analysed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR. Since the genome sizes for these two species was estimated to be about 2.5 Gb, Illumina paired-end 250 bp sequencing was run on HiSeqX10 and NovaSeq instruments to obtain about 200 Gb per sample corresponding to a genome depth-of-coverage of about 80x.

#### *MinION long-read sequencing*

Considering the DNA quality required to perform sequencing with Oxford Nanopore Technologies (ONT), a specific protocol to extract DNA from roadkill was designed (Tilak et al., 2020). First, genomic DNA was extracted by using the classical phenol-chloroform method. Then, we evaluated the cleanliness of the extractions by using (1) a binocular magnifying glass to check the absence of suspended particles (*e.g.* hairpieces), and (2) both Nanodrop and Qubit/Nanodrop ratio. To select the longest DNA fragments, we applied a specific ratio of 0.4x of AMPure beads applied (Tilak et al., 2020). Extracted-DNA size was then homogenized using covaris G-tubes to optimize sequencing yield. Finally, long-read ONT sequencing was performed through MinION flowcells (FLO-MIN-106) using libraries

prepared with the ONT Ligation Sequencing kit SQK-LSK109. For both species, we run MinION sequencing until about 30 Gb per sample were obtained to reach a genome depth-of-coverage of about 12x.

#### *Hybrid assembly of short and long reads*

Short reads were cleaned using Trimmomatic 0.33 (Bolger et al., 2014) by removing low quality bases from their beginning (LEADING:3) and end (TRAILING:3), and by removing reads shorter than 50 bp (MINLEN:50). Quality was measured for sliding windows of four base pairs and had to be greater than 15 on average (SLIDINGWINDOW:4:15). For MinION sequencing, basecalling of fast5 files was performed using Guppy v3.1.5 (developed by ONT) with the *high accuracy* option, which takes longer but is more accurate than the standard *fast* model (**Appendix 1 – Figure 1**). Long-read adapters were removed using Porechop v0.2.3 (<https://github.com/rrwick/Porechop>). To take advantage of both the high accuracy of Illumina short reads sequencing and the size of MinION long reads, assemblies were performed using the MaSuRCA hybrid genome assembler (Zimin et al., 2013). This method transforms large numbers of paired-end reads into a much smaller number of longer ‘super-reads’ and permits assembling Illumina reads of differing lengths together with longer ONT reads. To illustrate the advantage of using short reads and long reads conjointly, assemblies were also performed with short reads only using SOAP-denovo (Luo et al., 2012) (kmer size=31, default parameters) and gaps between contigs were closed using the abundant paired relationships of short reads with GapCloser 1.12 (Luo et al., 2012). To evaluate genome quality, traditional measures, like the number of scaffolds and contig N50, the mean and maximum lengths were evaluated for 503 mammalian genome assemblies retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/assembly>) on August 13<sup>th</sup>, 2019 with filters: “Exclude derived from surveillance project”, “Exclude anomalous”, “Exclude partial”, and using only the RefSeq assembly for *Homo sapiens*. Finally, we assessed the gene completeness of our



assemblies by comparison with the 63 carnivoran assemblies available at NCBI on August 13th, 2019 using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3 (Waterhouse et al., 2018) with the Mammalia OrthoDB 9 BUSCO gene set (Zdobnov et al., 2017) through the gVolante web server (Nishimura et al., 2017).

## **Comparative species delineation based on genomic data**

### *Sampling and resequencing*

To assess the genetic diversity in *P. cristatus*, we sampled an additional roadkill individual of the South African subspecies *P. c. cristatus* (NMB 12667) and an individual of the East African subspecies *P. c. septentrionalis* (NMS.Z.2018.54) born in a zoo from wild Tanzanian parents (**Table 1**). A similar sampling was done for *O. megalotis*, with an additional roadkill individual of the South African subspecies *O. m. megalotis* (NMB 12640) and an individual of the East African subspecies *O. m. virgatus* (FMNH 158128) from Tanzania (**Table 1**). DNA extractions were performed with the DNeasy Blood and Tissue Kit (Qiagen), following manufacturer's instructions and a total amount of 1.0µg DNA per sample was outsourced to Novogene Europe (Cambridge, UK) for Illumina library preparation and Illumina paired-end 250 bp sequencing on HiSeqX10 and NovaSeq instruments to obtain about 200 Gb per sample (genome depth-of-coverage of about 80x). The resulting reads were cleaned using Trimmomatic 0.33 with the same parameters as described above.

### *Heterozygosity and genetic differentiation estimation*

In a panmictic population alleles observed in one individual are shared randomly with other individuals of the same population and the frequencies of homozygous and heterozygous alleles should follow Hardy-Weinberg expectations. However, any structure in subpopulations leads to a deficiency of heterozygotes (relative to Hardy-Weinberg expectations) in these subpopulations due to inbreeding (Holsinger and Weir, 2009;

Walhund, 2010) and thus decreases the polymorphism within the inbred subpopulations with respect to the polymorphism of the global population. Given that, Hudson et al. (1992) defined the  $F_{ST}$  as a measure of polymorphism reduction in two subdivided populations ( $p$  within) compared to the population at large ( $p$  between).

To assess the  $p$  within and  $p$  between of the two subspecies of each species (*P. cristatus* and *O. megalotis*), we compared the heterozygous alleles (SNPs) of two individuals of the same subspecies and the SNPs of two individuals of different subspecies by computing a  $F_{ST}$ -like statistic (hereafter called Genetic Differentiation Index: GDI) (**Appendix 3 – Figure 1**). In fact, polymorphic sites can be discriminated in four categories: (1) fixed in one individual (*e.g.* AA/TT); (2) shared with both individuals (*e.g.* AT/AT); (3) specific to individual 1 (*e.g.* AT/AA); and (4) specific to individual 2 (*e.g.* AA/AT). Using these four categories, it is possible to estimate the polymorphism of each individual 1 and 2 and thus estimate a GDI between two individuals of the same population A and the GDI between two individuals of different populations A and B as follows:

$$GDI_{intra A} = 1 - \frac{(\pi_{A1} + \pi_{A2})/2}{\pi_{totA}}$$

$$GDI_{intra B} = 1 - \frac{(\pi_{B1} + \pi_{B2})/2}{\pi_{totB}}$$

For each species cleaned short reads of all individuals (the one used to construct the reference genome and the two resequenced from each population) were aligned with their reference genome using BWA-MEM (Li, 2013). BAM files were created and merged using SAMtools (Li et al., 2009). Likely contaminant contigs identified using BlobTools (Laetsch and Blaxter, 2017) (**Appendix 4 – Figure 1, Supplementary Files 10-11**) and contigs likely belonging to the X chromosome following LASTZ (Rahmani et al., 2011) alignments were removed (contigs that align with cat or dog autosomes and not to X chromosome have been selected). Then, 100 regions of 100,000 bp were randomly sampled among contigs longer

885 than 100,000 bp and 10 replicates of this sampling were performed (*i.e.* 10 x 100 x 100,000  
886 bp = 100 Mb) to assess statistical variance in the estimates. Genotyping of these regions was  
887 performed with freebayes v1.3.1-16 (git commit id: g85d7bfc) (Garrison and Marth, 2012)  
888 using the parallel mode (Tange, 2011). Only SNPs with freebayes-estimated quality higher  
889 than 10 were considered for further analyses. A first GDI estimation comparing the average  
890 of the private polymorphisms of the two southern individuals (*p within A*) and the total  
891 polymorphism of the two individuals (*p between A*) was estimated to control that no genetic  
892 structure was observed in the Southern subspecies. Then a global GDI comparing the private  
893 polymorphisms of individuals from the two populations (*p within AB*) and the total  
894 polymorphism of the species (the two populations, *p between AB*) was estimated with one  
895 individual from each population (**Appendix 3 – Figure 1**). Finally, the two GDI were  
896 compared to check if the Southern populations were more structured than the entire  
897 populations.

898 To contextualize these results, the same GDI measures were estimated for well-  
899 defined species of Carnivora. The species pairs used to make the comparison and thus help  
900 gauge the taxonomic status of the bat-eared fox and aardwolf subspecies were selected  
901 according to the following criteria: (1) the two species had to be as closely related as  
902 possible, (2) they had both reference genomes and short reads available, (3) their estimated  
903 coverage for the two species had to be greater than 15x, and (4) short-read sequencing data  
904 had to be available for two individuals for one species of the pair. Given that, four species  
905 pairs were selected: (1) *Canis lupus* / *Canis aureus* (*Canis lupus*: SRR8926747,  
906 SRR8926748; *Canis aureus*: SRR7976426; vonHoldt et al., 2016; reference genome:  
907 GCF\_000002285.3 ; Lindblad-Toh et al., 2005); (2) *Ursus maritimus* / *Ursus arctos* (*Ursus*  
908 *maritimus* PB43: SRR942203, SRR942290, SRR942298; *Ursus maritimus* PB28:  
909 SRR942211, SRR942287, SRR942295; *Ursus arctos*: SRR935591, SRR935625,

SRR935627; Liu et al., 2014); (3) *Lynx pardinus* / *Lynx lynx* (*Lynx pardinus* LYNX11 :  
ERR1255591-ERR1255594; *Lynx lynx* LYNX8: ERR1255579-ERR1255582; *Lynx lynx*  
LYNX23: ERR1255540-ERR1255549; Abascal et al., 2016); and (4) *Panthera leo* /  
*Panthera pardus* (*Panthera leo*: SRR10009886, SRR836361; *Panthera pardus*:  
SRR3041424; Kim et al., 2016). Raw reads for the three individuals of each species pair were  
downloaded, cleaned and mapped as described above. Then, the same GDI estimation  
protocol was applied to each species pair by estimating the GDI within species, using two  
individuals of the same species, and the GDI between species, using one individual of each  
species of the pair.

To check the robustness of the genetic differentiation index estimation, two additional  
analyses were conducted. First, given that the estimation could be biased by the depth-of-  
coverage used for the genotype calling, the reads used for all individuals were randomly  
subsampled to obtain a homogenised depth-of-coverage of about 15x. Based on these new  
datasets, genetic differentiation indices were re-estimated for each group. Second, to show  
the consistency of the results, when few individuals are used for the estimates, a permuted  
subsampling approach, drawing from a larger dataset, was performed. Using the species pairs  
*Ursus maritimus*/*Ursus arctos*, for which sequencing data were available for 10 individuals  
of each species, genetic differentiation indices were estimated using all possible  
combinations, using either two individuals for *Ursus arctos* or one individual for each species  
(i.e. 45 *Ursus arctos*/*Ursus arctos* and 100 *Ursus arctos*/*Ursus maritimus*). Given the number  
of possible combinations, estimates were performed on only five replicates (instead of 10) of  
100 regions of 100,000bp for each combination (**Figure 3 – Figure supplement 2**).

## **Demographic analyses**

Historical demographic variations in effective population size were estimated using the Pairwise Sequentially Markovian Coalescent (PSMC) model implemented in the software PSMC (<https://github.com/lh3/psmc>) (Li and Durbin, 2011). As described above, cleaned short reads were mapped against the corresponding reference genome using BWA-MEM (Li, 2013) and genotyping was performed using Freebayes v1.3.1-16 (git commit id: g85d7bfc) (Garrison and Marth, 2012) for the three individuals of each species. VCF files were converted to fasta format using a custom python script, excluding positions with quality below 20 and a depth-of-coverage below 10x or higher than 200x. Diploid sequences in fasta format were converted into PSMC fasta format using a C++ program written using the BIO++ library (Guéguen et al., 2013) with a block length of 100bp and excluding blocks containing more than 20% missing data as implemented in “fq2psmcfa” (<https://github.com/lh3/psmc>).

PSMC analyses were run for all other populations, testing several -t and -p parameters including -p "4+30\*2+4+6+10" (Nadachowska-Brzyska et al., 2013) and -p "4+25\*2+4+6" (Kim et al., 2016) but also -p "4+10\*3+4", -p "4+20\*2+4" and -p "4+20\*3+4". Overall, the tendencies were similar, but some parameters led to unrealistic differences between the two individuals from the South African population of *Otocyon megalotis*. We chose to present the results obtained using the parameters -t15 -r4 -p "4+10\*3+4". For this parameter setting the variance in ancestral effective population size was estimated by bootstrapping the scaffolds 100 times. To scale PSMC results, based on several previous studies on large mammals, a mutation rate of  $10^{-8}$  mutation/site/generation (Ekblom et al., 2018; Gopalakrishnan et al., 2017) and a generation time of two years (Clark, 2005; Koehler and Richardson, 1990; van Jaarsveld, 1993) were selected. Results were plotted in Rv3.63 (R core Team, 2020) using the function “psmc.results” (<https://doi.org/10.5061/dryad.0618v/4>) (Liu and Hansen, 2017) modified using ggplot2 (Wickham, 2016) and cowplot (Wilke, 2016).

959

## 960 **Phylogenomic inferences**

961 To infer the Carnivora phylogenetic relationships, all carnivoran genomes available on  
962 Genbank, the DNAZoo website (<https://www.dnazoo.org>), and the OrthoMaM database  
963 (Scornavacca et al., 2019) as of February 11th, 2020 were downloaded (**Supplementary File**  
964 **12**). In cases where more than one genome was available per species, the assembly with the  
965 best BUSCO scores was selected. Then, we annotated our two reference genome assemblies  
966 and the other unannotated assemblies using MAKER2 (Holt and Yandell, 2011) following  
967 the recommendations of the DNA Zoo protocol ([https://www.dnazoo.org/post/the-first-](https://www.dnazoo.org/post/the-first-million-genes-are-the-hardest-to-make-r)  
968 [million-genes-are-the-hardest-to-make-r](https://www.dnazoo.org/post/the-first-million-genes-are-the-hardest-to-make-r)). In the absence of available transcriptomic data, this  
969 method leveraged the power of homology combined with the thorough knowledge  
970 accumulated on the gene content of mammalian genomes. As advised, a mammal-specific  
971 subset of UniProtKB/Swiss-Prot, a manually annotated, non-redundant protein sequence  
972 database, was used as a reference for this annotation step (Boutet et al., 2016). Finally, the  
973 annotated coding sequences (CDSs) recovered for the Southern aardwolf (*P. c. cristatus*)  
974 were used to assemble those of the Eastern aardwolf (*P. c. septentrionalis*) by mapping the  
975 resequenced Illumina reads using BWA-MEM (Li, 2013).

976 Orthologous genes were extracted following the orthology delineation process of the  
977 OrthoMaM database (OMM) (Scornavacca et al., 2019). First, for each orthologous-gene  
978 alignment of OMM, a HMM profile was created via hmmbuild, using default parameters of  
979 the HMMER toolkit (Eddy, 2011), and all HMM profiles were concatenated and summarised  
980 using hmmcompress to construct a HMM database. Then, for each CDS newly annotated by  
981 MAKER, hmmscan was used on the HMM database to retrieve the best hits among the  
982 orthologous gene alignments. For each orthologous gene alignment, the most similar  
983 sequences for each species were detected via *hmmsearch*. Outputs from *hmmsearch* and

*hmmscan* were discarded, if the first-hit score was not substantially better than the second (hit<sub>2</sub> < 0.9 hit<sub>1</sub>). This ensures our orthology predictions for the newly annotated CDSs to be robust. Then, the cleaning procedure of the OrthoMaM database was applied to the set of orthologous genes obtained. This process, implemented in a singularity image (Kurtzer et al., 2017) named *OMM\_MACSE.sif* (Ranwez et al., 2021), is composed of several steps including nucleotide sequence alignment at the amino acid level with MAFFT (Katoh and Standley, 2013), refining alignments to handle frameshifts with MACSE v2 (Ranwez et al., 2018), cleaning of non-homologous sequences, and masking of erroneous/dubious parts of gene sequences with HMMcleaner (Di Franco et al., 2019). Finally, the last step of the cleaning process was to remove sequences that generated abnormally long branches during gene tree inferences. This was done by reconstructing gene trees using IQ-TREEv1.6.8 (Nguyen et al., 2015) with the MFP option to select the best-fitting model for each gene. Then, the sequences generating abnormally long branches were identified and removed by *PhylteR* (<https://github.com/damiendevienne/phylter>). This software allows detection and removal of outliers in phylogenomic datasets by iteratively removing taxa in genes and optimising a concordance score between individual distance matrices.

Phylogenomic analyses were performed using maximum likelihood (ML) using IQ-TREE 1.6.8 (Nguyen et al., 2015) on the supermatrix resulting from the concatenation of all orthologous genes previously recovered with the TESTNEW option to select the best-fitting model for each partition. Two partitions per gene were defined to separate the first two codon positions from the third codon positions. Node supports were estimated with 100 non-parametric bootstrap replicates. Furthermore, gene concordant (gCF) and site concordant (sCF) factors were measured to complement traditional bootstrap node-support measures as recommended in Minh et al. (2020). For each orthologous gene alignment a gene tree was inferred using IQ-TREE with model selection and gCF and sCF were calculated using the

specific option -scf and -gcf in IQ-TREE (Minh et al., 2020). The gene trees obtained with this analysis were also used to perform a coalescent-based species tree inference using ASTRAL-III (Zhang et al., 2018).

## **Data access**

Genome assemblies, associated Illumina and Nanopore sequence reads, and mitogenomes have been submitted to the National Center for Biotechnology Information (NCBI) and will be available after publication under BioProject number PRJNA681015. The full analytical pipeline, phylogenetic datasets (mitogenomic and genomic), corresponding trees, and other supplementary materials are available from zenodo.org (DOI: 10.5281/zenodo.4479226).

## **Disclosure declaration**

The authors declare that they have no competing interests.

## **Funding**

This work was supported by grants from the European Research Council (ERC-2015-CoG-683257 ConvergeAnt project), Investissements d'Avenir of the Agence Nationale de la Recherche (CEMEB: ANR-10-LABX-0004; ReNaBi-IFB: ANR-11-INBS-0013; MGX: ANR-10-INBS-09), and the National Research Foundation of South Africa (Grant specific unique reference number 86321).

## **Acknowledgements**

We are indebted to the Broad Institute ([www.broadinstitute.org](http://www.broadinstitute.org)), the DNA Zoo ([www.dnazoo.org](http://www.dnazoo.org)), and numerous other sequencing centres and institutions for making their



mammalian genomic data publically available. We would like to thank Rachid Koual and Amandine Magdeleine for technical help with DNA extractions and library preparations, Aude Caizergues and Nathalie Delsuc for fieldwork assistance, Christian Fontaine, Jean-Christophe Vié (Faune Sauvage, French Guiana), Corine Esser (Fauverie du Mont Faron, Toulon, France), François Catzefflis (ISEM Mammalian Tissue Collection), Adam Ferguson and Bruce Patterson (Field Museum of Natural History, Chicago, USA), and Lily Crowley and Andrew Swales (Hamerton Zoo Park, UK) for access to tissue samples. The National Museum (Bloemfontein, Free State, South Africa) is thanked for their collaboration and for making tissues from the Mammal Collection available for the study. ACK thanks the Negaunee Foundation for their generous support of a curatorial preparator who sampled the East African aardwolf used in this study. We also acknowledge Pierre-Alexandre Gagnaire for helpful discussion on the genetic differentiation index, Brian Chase for providing references on African paleoclimate, and Sérgio Ferreira-Cardoso for taking measurements of aardwolf skulls. Robb Cadd kindly made available his aardwolf photographs taken at Hamerton Zoo Park. We thank the Montpellier GenomiX Platform (MGX) part of the France Génomique National Infrastructure for sequencing data generation. Computational analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) computing platform. We are also grateful to the Institut Français de Bioinformatique and the Roscoff Bioinformatics platform ABiMS (<http://abims.sb-roscoff.fr>) for providing help for computing and storage resources. This is contribution ISEM 2021-XXX-SUD of the Institut des Sciences de l'Evolution de Montpellier.

## References

Abascal F, Corvelo A, Cruz F, Villanueva-Cañas JL, Vlasova A, Marcet-Houben M, Martínez-Cruz B, Cheng JY, Prieto P, Quesada V, Quilez J, Li G, García F, Rubio-Camarillo M, Frias L, Ribeca P, Capella-Gutiérrez S, Rodríguez JM, Câmara F, Lowy E, Cozzuto L, Erb I, Tress ML, Rodríguez-Ales JL, Ruiz-Orera J, Reverter F, Casas-Marce M, Soriano L, Arango JR, Derdak S, Galán B, Blanc J, Gut M, Lorente-Galdos B, Andrés-Nieto M, López-Otín C, Valencia A, Gut I, García JL, Guigó R, Murphy WJ, Ruiz-Herrera A, Marques-Bonet T, Roma G, Notredame C,

1058 Mailund T, Albà MM, Gabaldón T, Alioto T, Godoy JA. 2016. Extreme genomic erosion after  
1059 recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol* **17**:251.  
1060 doi:10.1186/s13059-016-1090-1

1061 Agnarsson I, Kuntner M, May-Collado LJ. 2010. Dogs, cats, and kin: A molecular species-level  
1062 phylogeny of Carnivora. *Mol Phylogenet Evol* **54**:726–745. doi:10.1016/J.YMPEV.2009.10.033

1063 Allen JA, Tjader R, Lang H. 1909. Mammals from British East Africa, collected by the Tjäder  
1064 Expedition of 1906. *Bull AMNH* **26**.

1065 Allen R, Ryan H, Davis BW, King C, Frantz L, Irving-Pease E, Barnett R, Linderholm A, Loog L,  
1066 Haile J, Lebrasseur O, White M, Kitchener AC, Murphy WJ, Larson G. 2020. A mitochondrial  
1067 genetic divergence proxy predicts the reproductive compatibility of mammalian hybrids. *Proc R*  
1068 *Soc B Biol Sci* **287**:20200690. doi:10.1098/rspb.2020.0690

1069 Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. 2020a. MitoFinder:  
1070 Efficient automated large-scale extraction of mitogenomic data in target enrichment  
1071 phylogenomics. *Mol Ecol Resour* **17**:1755–1760. doi:10.1111/1755-0998.13160

1072 Allio R, Scornavacca C, Nabholz B, Clamens A-L, Sperling FA, Condamine FL. 2020b. Whole  
1073 genome shotgun phylogenomics resolves the pattern and timing of swallowtail butterfly  
1074 evolution. *Syst Biol* **69**:38–60. doi:10.1093/sysbio/syz030

1075 Armstrong EE, Taylor RW, Miller DE, Kaelin CB, Barsh GS, Hadly EA, Petrov D. 2020. Long live  
1076 the king: chromosome-level assembly of the lion (*Panthera leo*) using linked-read, Hi-C, and  
1077 long-read data. *BMC Biol* **18**:3. doi:10.1186/s12915-019-0734-5

1078 Arnason U, Gullberg A, Janke A, Kullberg M. 2007. Mitogenomic analyses of caniform relationships.  
1079 *Mol Phylogenet Evol* **45**:863–874. doi:10.1016/J.YMPEV.2007.06.019

1080 Atickem A, Stenseth NC, Drouilly M, Bock S, Roos C, Zinner D. 2018. Deep divergence among  
1081 mitochondrial lineages in African jackals. *Zool Scr* **47**:1–8. doi:10.1111/zsc.12257

1082 Barnett R, Yamaguchi N, Barnes I, Cooper A. 2006. The origin, current diversity and future  
1083 conservation of the modern lion ( *Panthera leo* ). *Proc R Soc B Biol Sci* **273**:2119–2125.  
1084 doi:10.1098/rspb.2006.3555

1085 Batra SS, Levy-Sakin M, Robinson J, Guillory J, Durinck S, Kwok P-Y, Cox LA, Seshagiri S, Song  
1086 YS, Wall JD. 2019. Accurate assembly of the olive baboon (*Papio anubis*) genome using long-  
1087 -read and Hi-C data. *bioRxiv* 678771. doi:10.1101/678771

1088 Blaimer BB, LaPolla JS, Branstetter MG, Lloyd MW, Brady SG. 2016. Phylogenomics, biogeography  
1089 and diversification of obligate mealybug-tending ants in the genus *Acropyga*. *Mol Phylogenet*  
1090 *Evol* **102**:20–29. doi:10.1016/J.YMPEV.2016.05.030

1091 Blanco MB, Greene LK, Williams RC, Andrianandrasana L, Yoder AD, Larsen PA. 2019. Next-  
1092 generation in situ conservation and educational outreach in Madagascar using a mobile genetics  
1093 lab. *bioRxiv* 650614. doi:10.1101/650614

1094 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data.  
1095 *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

1096 Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L,  
1097 Xenarios I. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt  
1098 KnowledgeBase: How to use the entry view. Humana Press, New York, NY. pp. 23–54.  
1099 doi:10.1007/978-1-4939-3167-5\_2

1100 Brown CR, Bomberger Brown M. 2013. Where has all the road kill gone? *Curr Biol*.

doi:10.1016/j.cub.2013.02.023

Cabrera A. 1910. LI.—On two new Carnivora from North-east Africa. *Ann Mag Nat Hist* **6**:461–465.

Caley T, Extier T, Collins JA, Schefuß E, Dupont L, Malaizé B, Rossignol L, Souron A, McClymont EL, Jimenez-Espejo FJ, García-Comas C, Eynaud F, Martinez P, Roche DM, Jorry SJ, Charlier K, Wary M, Gourves PY, Billy I, Giraudeau J. 2018. A two-million-year-long hydroclimatic context for hominin evolution in southeastern Africa. *Nature* **560**:76–79. doi:10.1038/s41586-018-0309-6

Casas-Marce M, Soriano L, López-Bao J V., Godoy JA. 2013. Genetics at the verge of extinction: insights from the Iberian lynx. *Mol Ecol* **22**:5503–5515. doi:10.1111/mec.12498

Chase BM, Niedermeyer EM, Boom A, Carr AS, Chevalier M, He F, Meadows ME, Ogle N, Reimer PJ. 2019. Orbital controls on Namib Desert hydroclimate over the past 50,000 years. *Geology* **47**:867–871. doi:10.1130/G46334.1

Chevalier M, Chase BM. 2015. Southeast African records reveal a coherent shift from high- to low-latitude forcing mechanisms along the east African margin across last glacial-interglacial transition. *Quat Sci Rev* **125**:117–130. doi:10.1016/j.quascirev.2015.07.009

Clark HO. 2005. *Otocyon megalotis*. *Mamm Species* 1–5. doi:10.1644/1545-1410(2005)766[0001:OM]2.0.CO;2

De Queiroz K. 2007. Species concepts and species delimitation. *Syst Biol* **56**:879–886. doi:10.1080/10635150701701083

Dehghani R, Wanntorp L, Pagani P, Källersjö M, Werdelin L, Veron G. 2008. Phylogeography of the white-tailed mongoose (Herpestidae, Carnivora, Mammalia) based on partial sequences of the mtDNA control region. *J Zool* **276**:385–393. doi:10.1111/j.1469-7998.2008.00502.x

Delisle I, Strobeck C. 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. *Mol Phylogenet Evol* **37**:192–201. doi:10.1016/J.YMPEV.2005.04.025

deMenocal PB. 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth Planet Sci Lett* **220**:3–24. doi:10.1016/S0012-821X(04)00003-2

deMenocal PB. 1995. Plio-Pleistocene African climate. *Science* (80- ). doi:10.1126/science.270.5233.53

Di Genova A, Ruz GA, Sagot M-F, Maass A. 2018. Fast-SG: an alignment-free algorithm for hybrid assembly. *Gigascience* **7**. doi:10.1093/gigascience/giy048

Di Franco A, Poujol R, Baurain D, Philippe H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol Biol* **19**:21. doi:10.1186/s12862-019-1350-2

Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. 2018. Quantifying temporal genomic erosion in endangered species. *Trends Ecol Evol* **33**:176–185. doi:10.1016/J.TREE.2017.12.002

Doronina L, Churakov G, Shi J, Brosius J, Baertsch R, Clawson H, Schmitz J. 2015. Exploring massive incomplete lineage sorting in Arctoids (Laurasiatheria, Carnivora). *Mol Biol Evol* **32**:msv188. doi:10.1093/molbev/msv188

Drake-Brockman RE. 1910. The mammals of Somaliland. *Hurst and Blackett*.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. De novo assembly of the *Aedes aegypti* genome using

1144 Hi-C yields chromosome-length scaffolds. *Science* (80- ) **356**:92–95.  
1145 doi:10.1126/science.aal3327

1146 Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* **7**.  
1147 doi:10.1371/journal.pcbi.1002195

1148 Eizirik E, Murphy WJ, Koepfli K-P, Johnson WE, Dragoo JW, Wayne RK, O'Brien SJ. 2010. Pattern  
1149 and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear  
1150 gene sequences. *Mol Phylogenet Evol* **56**:49–63. doi:10.1016/J.YMPEV.2010.01.033

1151 Ekblom R, Brechlin B, Persson J, Smeds L, Johansson M, Magnusson J, Flagstad Ø, Ellegren H.  
1152 2018. Genome sequencing and conservation genomics in the Scandinavian wolverine  
1153 population. *Conserv Biol* **32**:1301–1312. doi:10.1111/cobi.13157

1154 Etherington GJ, Heavens D, Baker D, Lister A, McNelly R, Garcia G, Clavijo B, Macaulay I, Haerty  
1155 W, Di Palma F. 2020. Sequencing smart: *De novo* sequencing and assembly approaches for a  
1156 non-model mammal. *Gigascience* **9**. doi:10.1093/GIGASCIENCE/GIAA045

1157 Faurby S, Eiserhardt WL, Svenning J. 2016. Strong effects of variation in taxonomic opinion on  
1158 diversification analyses. *Methods Ecol Evol* **7**:4–13. doi:10.1111/2041-210X.12449

1159 Flynn JJ, Finarelli JA, Spaulding M. 2010. Phylogeny of the Carnivora and Carnivoramorpha, and the  
1160 use of the fossil record to enhance understanding of evolutionary transformations In: Goswami  
1161 A, Friscia A, editors. *Carnivoran Evolution*. Cambridge: Cambridge University Press. pp. 25–63.  
1162 doi:10.1017/CBO9781139193436.003

1163 Flynn JJ, Finarelli JA, Zehr S, Hsu J, Nedbal MA. 2005. Molecular phylogeny of the Carnivora  
1164 (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships.  
1165 *Syst Biol* **54**:317–337. doi:10.1080/10635150590923326

1166 Flynn JJ, Nedbal MA. 1998. Phylogeny of the Carnivora (Mammalia): congruence vs incompatibility  
1167 among multiple data sets. *Mol Phylogenet Evol* **9**:414–426. doi:10.1006/MPEV.1998.0504

1168 Frankham R, Ballou JD, Dudash MR, Eldridge MDB, Fenster CB, Lacy RC, Mendelson JR, Porton  
1169 IJ, Ralls K, Ryder OA. 2012. Implications of different species concepts for conserving  
1170 biodiversity. *Biol Conserv* **153**:25–31. doi:10.1016/J.BIOCON.2012.04.034

1171 Galov A, Fabbri E, Caniglia R, Arbanasić H, Lapalombella S, Florijančić T, Bošković I, Galaverni M,  
1172 Randi E. 2015. First evidence of hybridization between golden jackal ( *Canis aureus* ) and  
1173 domestic dog ( *Canis familiaris* ) as revealed by genetic markers. *R Soc Open Sci* **2**:150450.  
1174 doi:10.1098/rsos.150450

1175 Galtier N. 2019. Delineating species in the speciation continuum: A proposal. *Evol Appl* **12**:657–663.  
1176 doi:10.1111/eva.12748

1177 Galtier N, Nabholz B, Glémin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular  
1178 diversity: a reappraisal. *Mol Ecol* **18**:4541–4550. doi:10.1111/j.1365-294X.2009.04380.x

1179 Gan HM, Falk S, Morales HE, Austin CM, Sunnucks P, Pavlova A. 2019. Genomic evidence of neo-  
1180 sex chromosomes in the eastern yellow robin. *Gigascience* **8**. doi:10.1093/gigascience/giz111

1181 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing.

1182 Gopalakrishnan S, Samaniego Castruita JA, Sinding M-HS, Kuderna LFK, Rääkkönen J, Petersen B,  
1183 Sicheritz-Ponten T, Larson G, Orlando L, Marques-Bonet T, Hansen AJ, Dalén L, Gilbert MTP.  
1184 2017. The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis*  
1185 spp. population genomics. *BMC Genomics* **18**:495. doi:10.1186/s12864-017-3883-3

1186 Gopalakrishnan S, Sinding M-HS, Ramos-Madrugal J, Niemann J, Samaniego Castruita JA, Vieira

1187 FG, Carøe C, Montero M de M, Kuderna L, Serres A, González-Basallote VM, Liu Y-H, Wang  
1188 G-D, Marques-Bonet T, Mirarab S, Fernandes C, Gaubert P, Koepfli K-P, Budd J, Rueness EK,  
1189 Sillero C, Heide-Jørgensen MP, Petersen B, Sicheritz-Ponten T, Bachmann L, Wiig Ø, Hansen  
1190 AJ, Gilbert MTP. 2018. Interspecific gene flow shaped the evolution of the genus *Canis*. *Curr*  
1191 *Biol* **28**:3441–3449.e5. doi:10.1016/J.CUB.2018.08.041

1192 Grant KM, Rohling EJ, Westerhold T, Zabel M, Heslop D, Konijnendijk T, Lourens L. 2017. A 3  
1193 million year index for North African humidity/aridity and the implication of potential pan-  
1194 African Humid periods. *Quat Sci Rev* **171**:100–118. doi:10.1016/j.quascirev.2017.07.005

1195 Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet  
1196 F, Cahais V, Bernard A, Scornavacca C, Nabholz B, Haudry A, Dachary L, Galtier N, Belkhir  
1197 K, Duthéil JY. 2013. Bio++: Efficient extensible libraries and tools for computational molecular  
1198 evolution. *Mol Biol Evol* **30**:1745–1750. doi:10.1093/molbev/mst097

1199 Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet  
1200 G, Nagy ZT, Lenglet G, Mayer F, Savolainen V. 2013. Next-generation museomics disentangles  
1201 one of the largest primate radiations. *Syst Biol* **62**:539–554. doi:10.1093/sysbio/syt018

1202 Hammer, Ø., Harper, D.A.T., and Ryan, P.D. 2001. PAST: Paleontological Statistics  
1203 Software Package for Education and Data Analysis. *Palaeontologia Electronica* 4(1):  
1204 9pp.

1205 Heller E. 1913. New antelopes and carnivores from British East Africa. *Smithson Misc Collect*.  
1206 Hollister N. 1918. East African mammals in the United States National Museum. Part I. Insectivora,  
1207 Chiroptera, and Carnivora. *Bull United States Natl Museum* **99**:1–194.

1208 Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating  
1209 and interpreting FST. *Nat Rev Genet* **10**:639–650. doi:10.1038/nrg2611

1210 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool  
1211 for second-generation genome projects. *BMC Bioinformatics* **12**:491. doi:10.1186/1471-2105-  
1212 12-491

1213 Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence  
1214 data. *Genetics* **132**.

1215 IUCN 2020. 2020. The IUCN Red List of Threatened Species. Version 2020-1.  
1216 <https://www.iucnredlist.org>.

1217 Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Diltthey AT, Fiddes  
1218 IT, Malla S, Marriott H, Nieto T, O’Grady J, Olsen HE, Pedersen BS, Rhie A, Richardson H,  
1219 Quinlan AR, Snutch TP, Tee L, Paten B, Phillippy AM, Simpson JT, Loman NJ, Loose M.  
1220 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat*  
1221 *Biotechnol* **36**:338–345. doi:10.1038/nbt.4060

1222 Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore  
1223 sequencing to the genomics community. *Genome Biol* **17**:1–11. doi:10.1186/s13059-016-1103-0

1224 Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence?  
1225 *Trends Genet* **22**:225–231. doi:10.1016/J.TIG.2006.02.003

1226 Jiang JB, Quattrini AM, Francis WR, Ryan JF, Rodríguez E, McFadden CS. 2019. A hybrid de novo  
1227 assembly of the sea pansy ( *Renilla muelleri* ) genome. *Gigascience* **8**.  
1228 doi:10.1093/gigascience/giz026

1229 Johnson TC, Werne JP, Brown ET, Abbott A, Berke M, Steinman BA, Halbur J, Contreras S,

1230 Grosshuesch S, Deino A, Scholz CA, Lyons RP, Schouten S, Damsté JSS. 2016. A  
1231 progressively wetter climate in southern East Africa over the past 1.3 million years. *Nature*  
1232 **537**:220–224. doi:10.1038/nature19065

1233 Kadobianskyi M, Schulze L, Schuelke M, Judkewitz B. 2019. Hybrid genome assembly and  
1234 annotation of *Danionella translucida*. *Sci Data* **6**:1–7. doi:10.1038/s41597-019-0161-z

1235 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast  
1236 model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589.  
1237 doi:10.1038/nmeth.4285

1238 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:  
1239 Improvements in performance and usability. *Mol Biol Evol* **30**:772–780.  
1240 doi:10.1093/molbev/mst010

1241 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,  
1242 Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic:  
1243 An integrated and extendable desktop software platform for the organization and analysis of  
1244 sequence data. *Bioinformatics* **28**:1647–1649. doi:10.1093/bioinformatics/bts199

1245 Kim S, Cho YS, Kim H-M, Chung O, Kim H, Jho S, Seomun H, Kim J, Bang WY, Kim C, An J, Bae  
1246 CH, Bhak Y, Jeon S, Yoon H, Kim Y, Jun J, Lee H, Cho S, Uphyrkina O, Kostyria A, Goodrich  
1247 J, Miquelle D, Roelke M, Lewis J, Yurchenko A, Bankevich A, Cho J, Lee S, Edwards JS,  
1248 Weber JA, Cook J, Kim S, Lee H, Manica A, Lee I, O’Brien SJ, Bhak J, Yeo J-H. 2016.  
1249 Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard  
1250 assembly. *Genome Biol* **17**:211. doi:10.1186/s13059-016-1071-4

1251 Koehler CE, Richardson PRK. 1990. *Proteles cristatus*. *Mamm Species* 1–6. doi:10.2307/3504197

1252 Koepfli K-P, Jenks SM, Eizirik E, Zahirpour T, Valkenburgh B Van, Wayne RK. 2006. Molecular  
1253 systematics of the Hyaenidae: Relationships of a relictual lineage resolved by a molecular  
1254 supermatrix. *Mol Phylogenet Evol* **38**:603–620. doi:10.1016/J.YMPEV.2005.10.017

1255 Koepfli K-P, Paten B, O’Brien SJ, O’Brien SJ. 2015. The Genome 10K Project: A way forward. *Annu*  
1256 *Rev Anim Biosci* **3**:57–111. doi:10.1146/annurev-animal-090414-014900

1257 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and  
1258 accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*  
1259 **27**:722–736. doi:10.1101/GR.215087.116

1260 Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. 2012. Statistics and truth in  
1261 phylogenomics. *Mol Biol Evol* **29**:457–472. doi:10.1093/molbev/msr202

1262 Kurtzer GM, Sochat V, Bauer MW. 2017. Singularity: Scientific containers for mobility of compute.  
1263 *PLoS One* **12**:e0177459. doi:10.1371/journal.pone.0177459

1264 Kwan HH, Culibrk L, Taylor GA, Leelakumari S, Tan R, Jackman SD, Tse K, MacLeod T, Cheng D,  
1265 Chuah E, Kirk H, Pandoh P, Carlsen R, Zhao Y, Mungall AJ, Moore R, Birol I, Marra MA,  
1266 Rosen DAS, Haulena M, Jones SJM, Kwan HH, Culibrk L, Taylor GA, Leelakumari S, Tan R,  
1267 Jackman SD, Tse K, MacLeod T, Cheng D, Chuah E, Kirk H, Pandoh P, Carlsen R, Zhao Y,  
1268 Mungall AJ, Moore R, Birol I, Marra MA, Rosen DAS, Haulena M, Jones SJM. 2019. The  
1269 Genome of the Steller Sea Lion (*Eumetopias jubatus*). *Genes (Basel)* **10**:486.  
1270 doi:10.3390/genes10070486

1271 Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research*  
1272 **6**:1287. doi:10.12688/f1000research.12232.1

1273 Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: Phylogenetic reconstruction with  
1274 infinite mixtures of profiles in a parallel environment. *Syst Biol* **62**:611–615.

- 1275       doi:10.1093/sysbio/syt022
- 1276   Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in  
1277       the genomes of living cats (Felidae). *Genome Res* **26**:1–11. doi:10.1101/GR.186668.114
- 1278   Li G, Figueiró H V, Eizirik E, Murphy WJ. 2019. Recombination-aware phylogenomics reveals the  
1279       structured genomic landscape of hybridizing cat species. *Mol Biol Evol* **36**:2111–2126.  
1280       doi:10.1093/molbev/msz139
- 1281   Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
- 1282   Li H, Durbin R. 2011. Inference of human population history from individual whole-genome  
1283       sequences. *Nature* **475**:493–496. doi:10.1038/nature10231
- 1284   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.  
1285       The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079.  
1286       doi:10.1093/bioinformatics/btp352
- 1287   Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, Zhang Z, Zhang Y, Wang W,  
1288       Li J, Wei F, Li H, Jian M, Li J, Zhang Z, Nielsen R, Li D, Gu W, Yang Z, Xuan Z, Ryder OA,  
1289       Leung FCC, Zhou Y, Cao J, Sun X, Fu Y, Fang X, Guo X, Wang B, Hou R, Shen F, Mu B, Ni  
1290       P, Lin R, Qian W, Wang G, Yu C, Nie W, Wang J, Wu Z, Liang H, Min J, Wu Q, Cheng S,  
1291       Ruan J, Wang M, Shi Z, Wen M, Liu B, Ren X, Zheng H, Dong D, Cook K, Shan G, Zhang H,  
1292       Kosiol C, Xie X, Lu Z, Zheng H, Li Y, Steiner CC, Lam TTY, Lin S, Zhang Q, Li G, Tian J,  
1293       Gong T, Liu H, Zhang D, Fang L, Ye C, Zhang J, Hu W, Xu A, Ren Y, Zhang G, Bruford MW,  
1294       Li Q, Ma L, Guo Y, An N, Hu Y, Zheng Y, Shi Y, Li Z, Liu Q, Chen Y, Zhao J, Qu N, Zhao S,  
1295       Tian F, Wang X, Wang H, Xu L, Liu X, Vinar T, Wang Y, Lam TW, Yiu SM, Liu S, Zhang H,  
1296       Li D, Huang Y, Wang X, Yang G, Jiang Z, Wang J, Qin N, Li L, Li J, Bolund L, Kristiansen K,  
1297       Wong GKS, Olson M, Zhang X, Li S, Yang H, Wang J, Wang J. 2010. The sequence and de  
1298       novo assembly of the giant panda genome. *Nature* **463**:311–317. doi:10.1038/nature08696
- 1299   Lim S, Chase BM, Chevalier M, Reimer PJ. 2016. 50,000 years of vegetation and climate change in  
1300       the southern Namib Desert, Pella, South Africa. *Palaeogeogr Palaeoclimatol Palaeoecol*  
1301       **451**:197–209. doi:10.1016/j.palaeo.2016.03.001
- 1302   Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL,  
1303       Kulbokas EJ, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP,  
1304       Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin  
1305       CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M,  
1306       Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP,  
1307       Searle SMJ, Sutter NB, Thomas R, Webber C, Lander ES. 2005. Genome sequence, comparative  
1308       analysis and haplotype structure of the domestic dog. *Nature* **438**:803–819.  
1309       doi:10.1038/nature04338
- 1310   Liu S, Hansen MM. 2017. PSMC (pairwise sequentially Markovian coalescent) analysis of RAD  
1311       (restriction site associated DNA) sequencing data. *Mol Ecol Resour* **17**:631–641.  
1312       doi:10.1111/1755-0998.12606
- 1313   Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M,  
1314       Babbitt C, Wray G, Li J, He W, Wang Z, Fu W, Xiang X, Morgan CC, Doherty A, O’Connell  
1315       MJ, McInerney JO, Born EW, Dalén L, Dietz R, Orlando L, Sonne C, Zhang G, Nielsen R,  
1316       Willerslev E, Wang J. 2014. Population genomics reveal recent speciation and rapid  
1317       evolutionary adaptation in polar bears. *Cell* **157**:785–794. doi:10.1016/J.CELL.2014.03.054
- 1318   Lorenzen ED, Heller R, Siegmund HR. 2012. Comparative phylogeography of African savannah  
1319       ungulates 1. *Mol Ecol* **21**:3656–3670. doi:10.1111/j.1365-294X.2012.05650.x
- 1320   Loughry WJ, McDonough CM. 1996. Are road kills valid indicators of armadillo population

1321 structure? *Am Midl Nat* **135**:53–59. doi:10.2307/2426871

1322 Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H,  
1323 Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G,  
1324 Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: An empirically  
1325 improved memory-efficient short-read de novo assembler. *Gigascience* **1**:18. doi:10.1186/2047-  
1326 217X-1-18

1327 Maigret TA. 2019. Snake scale clips as a source of high quality DNA suitable for RAD sequencing.  
1328 *Conserv Genet Resour* **11**:373–375. doi:10.1007/s12686-018-1019-y

1329 Maslin MA, Brierley CM, Milner AM, Shultz S, Trauth MH, Wilson KE. 2014. East african climate  
1330 pulses and early human evolution. *Quat Sci Rev*. doi:10.1016/j.quascirev.2014.06.012

1331 Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão  
1332 TLL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL,  
1333 Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, Murphy WJ. 2011.  
1334 Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification.  
1335 *Science (80- )* **334**:521–524. doi:10.1126/SCIENCE.1211028

1336 Meyer CP, Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol*  
1337 **3**:e422. doi:10.1371/journal.pbio.0030422

1338 Miller JM, Hallager S, Monfort SL, Newby J, Bishop K, Tidmus SA, Black P, Houston B, Matthee  
1339 CA, Fleischer RC, Hallager S, Monfort SL, Newby J, Bishop ÁK, Tidmus SA, Black P, Houston  
1340 ÁB, Matthee CA. 2011. Phylogeographic analysis of nuclear and mtDNA supports subspecies  
1341 designations in the ostrich (*Struthio camelus*). *Conserv Genet* **12**:423–431. doi:10.1007/s10592-  
1342 010-0149-x

1343 Minh BQ, Hahn MW, Lanfear R. 2020. New methods to calculate concordance factors for  
1344 phylogenomic datasets. *Mol Biol Evol*. doi:10.1093/molbev/msaa106

1345 Nadachowska-Brzyska K, Burri R, Olason PI, Kawakami T, Smeds L, Ellegren H. 2013.  
1346 Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-  
1347 genome re-sequencing data. *PLoS Genet* **9**:e1003942. doi:10.1371/journal.pgen.1003942

1348 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic  
1349 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268–274.  
1350 doi:10.1093/molbev/msu300

1351 Nicholls SM, Quick JC, Tang S, Loman NJ. 2019. Ultra-deep, long-read nanopore sequencing of  
1352 mock microbial community standards. *Gigascience* **8**. doi:10.1093/gigascience/giz043

1353 Nishimura O, Hara Y, Kuraku S. 2017. gVolante for standardizing completeness assessment of  
1354 genome and transcriptome assemblies. *Bioinformatics* **33**:3635–3637.  
1355 doi:10.1093/bioinformatics/btx445

1356 Parker J, Helmstetter AJ, Devey D, Wilkinson T, Papadopoulos AST. 2017. Field-based species  
1357 identification of closely-related plants using real-time nanopore sequencing. *Sci Rep* **7**:8345.  
1358 doi:10.1038/s41598-017-08461-5

1359 Peng R, Zeng B, Meng X, Yue B, Zhang Z, Zou F. 2007. The complete mitochondrial genome and  
1360 phylogenetic analysis of the giant panda (*Ailuropoda melanoleuca*). *Gene* **397**:76–83.  
1361 doi:10.1016/J.GENE.2007.04.009

1362 Périquet S, Roxburgh L, le Roux A, Collinson WJ. 2018. Testing the value of citizen science for  
1363 roadkill studies: A case study from South Africa. *Front Ecol Evol* **6**:15.  
1364 doi:10.3389/fevo.2018.00015



- 1365 Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Barrio-Amorós CL,  
1366 Salazar-Valenzuela D, Prost S. 2018. Real-time DNA barcoding in a rainforest using nanopore  
1367 sequencing: opportunities for rapid biodiversity assessments and local capacity building.  
1368 *Gigascience* **7**. doi:10.1093/gigascience/giy033
- 1369 R core Team. 2020. R: A language and environment for statistical computing.
- 1370 Rahmani AM, Liljeberg P, Plosila J, Tenhunen H. 2011. LastZ: An ultra optimized 3D networks-on-  
1371 chip architecture. *2011 14th Euromicro Conf Digit Syst Des* 173–180. doi:10.1109/DSD.2011.26
- 1372 Ranwez V, Chantret N, Delsuc F. 2021. Aligning protein-coding nucleotide sequences with MACSE.  
1373 *Methods Mol Biol* **2231**:51–70. doi:10.1007/978-1-0716-1036-7\_4
- 1374 Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the  
1375 alignment of coding sequences accounting for frameshifts and stop codons. *Mol Biol Evol*  
1376 **35**:2582–2584. doi:10.1093/molbev/msy159
- 1377 Ravinet M, Westram A, Johannesson K, Butlin R, André C, Panova M. 2016. Shared and nonshared  
1378 genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol Ecol* **25**:287–  
1379 305. doi:10.1111/mec.13332
- 1380 Roberts A. 1951. The mammals of South Africa. The Mammals of South Africa. *Mamm South Africa*  
1381 *Mamm South Africa*.
- 1382 Roberts A. 1932. Roberts, Austin. "Preliminary description of fifty-seven new forms of South African  
1383 mammals. *Ann Transvaal Museum* **15**:1–19.
- 1384 Rohland N, Pollack JL, Nagel D, Beauval C, Airvaux J, Pääbo S, Hofreiter M. 2005. The population  
1385 history of extant and extinct hyenas. *Mol Biol Evol* **22**:2435–2443. doi:10.1093/molbev/msi244
- 1386 Rothschild LW. 1902. Two new subspecies of *Proteles*. *Novit Zool* **9**:443.
- 1387 Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding light on the grey  
1388 zone of speciation along a continuum of genomic divergence. *PLOS Biol* **14**:e2000234.  
1389 doi:10.1371/journal.pbio.2000234
- 1390 Rybczynski N, Dawson MR, Tedford RH. 2009. A semi-aquatic Arctic mammalian carnivore from  
1391 the Miocene epoch and origin of Pinnipedia. *Nature* **458**:1021–1024. doi:10.1038/nature07985
- 1392 Sato JJ, Wolsan M, Minami S, Hosoda T, Sinaga MH, Hiyama K, Yamaguchi Y, Suzuki H. 2009.  
1393 Deciphering and dating the red panda's ancestry and early adaptive radiation of Musteloidea.  
1394 *Mol Phylogenet Evol* **53**:907–922. doi:10.1016/J.YMPEV.2009.08.019
- 1395 Sato JJ, Wolsan M, Suzuki H, Hosoda T, Yamaguchi Y, Hiyama K, Kobayashi M, Minami S. 2006.  
1396 Evidence from nuclear DNA sequences sheds light on the phylogenetic relationships of  
1397 Pinnipedia: single origin with affinity to Musteloidea. *Zoolog Sci* **23**:125–146.  
1398 doi:10.2108/zsj.23.125
- 1399 Schröder C, Bleidorn C, Hartmann S, Tiedemann R. 2009. Occurrence of Can-SINEs and intron  
1400 sequence evolution supports robust phylogeny of pinniped carnivores and their terrestrial  
1401 relatives. *Gene* **448**:221–226. doi:10.1016/J.GENE.2009.06.012
- 1402 Schwartz ALW, Shilling FM, Perkins SE. 2020. The value of monitoring wildlife roadkill. *Eur J*  
1403 *Wildl Res* **66**:1–12. doi:10.1007/s10344-019-1357-4
- 1404 Scornavacca C, Belkhir K, Lopez J, Dernat R, Delsuc F, Douzery EJP, Ranwez V. 2019. OrthoMaM  
1405 v10: Scaling-up orthologous coding sequence and exon alignments with more than one hundred  
1406 mammalian genomes. *Mol Biol Evol* **36**:861–862. doi:10.1093/molbev/msz015
- 1407 Scott AD, Zimin A V., Puiu D, Workman R, Britton M, Zaman S, Caballero M, Read AC, Bogdanove

1408 AJ, Burns E, Wegrzyn J, Timp W, Salzberg SL, Neale DB. 2020. The giant sequoia genome and  
1409 proliferation of disease resistance genes. *bioRxiv* 2020.03.17.995944.  
1410 doi:10.1101/2020.03.17.995944

1411 Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K,  
1412 Maurer N, Koren S, Sedlazeck FJ, Marschall T, Mayes S, Costa V, Zook JM, Liu KJ, Kilburn D,  
1413 Sorensen M, Munson KM, Vollger MR, Monlong J, Garrison E, Eichler EE, Salama S, Haussler  
1414 D, Green RE, Akeson M, Phillippy A, Miga KH, Carnevali P, Jain M, Paten B. 2020. Nanopore  
1415 sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes.  
1416 *Nat Biotechnol* 1–10. doi:10.1038/s41587-020-0503-6

1417 Shilling F, Perkins SE, Collinson W. 2015. Wildlife/Roadkill Observation and Reporting  
1418 Systems Handbook of Road Ecology. Chichester, UK: John Wiley & Sons, Ltd. pp. 492–501.  
1419 doi:10.1002/9781118568170.ch62

1420 Srivathsan A, Baloglu B, Wang W, Tan WX, Bertrand D, Ng AHQ, Boey EJH, Koh JJY, Nagarajan  
1421 N, Meier R. 2018. A MinION<sup>TM</sup>-based pipeline for fast and cost-effective DNA barcoding. *Mol*  
1422 *Ecol Resour* **18**:1035–1049. doi:10.1111/1755-0998.12890

1423 Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM. 2018. Finding Nemo: hybrid  
1424 assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish  
1425 (*Amphiprion ocellaris*) genome assembly. *Gigascience* **7**. doi:10.1093/gigascience/gix137

1426 Tange O. 2011. Gnu parallel-the command-line power tool. *USENIX Mag* **36**:42–47.

1427 Tilak M-K, Justy F, Debais-Thibaud M, Botero-Castro F, Delsuc F, Douzery EJP. 2015. A cost-  
1428 effective straightforward protocol for shotgun Illumina libraries designed to assemble complete  
1429 mitogenomes from non-model species. *Conserv Genet Resour* **7**:37–40. doi:10.1007/s12686-  
1430 014-0338-x

1431 Tilak M-K, Allio R, Delsuc F. 2020. An optimized protocol for sequencing mammalian roadkill  
1432 tissues with Oxford Nanopore Technology (ONT). doi:10.17504/PROTOCOLS.IO.BEIXJCFN

1433 Trauth MH, Larrasoana JC, Mudelsee M. 2009. Trends, rhythms and events in Plio-Pleistocene  
1434 African climate. *Quat Sci Rev* **28**:399–411. doi:10.1016/j.quascirev.2008.11.003

1435 van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J,  
1436 Lander ES. 2010. Hi-C: A method to study the three-dimensional architecture of genomes. *J Vis*  
1437 *Exp* e1869. doi:10.3791/1869

1438 van Jaarsveld AS. 1993. A comparative investigation of hyaena and aardwolf life-histories, with notes  
1439 on spotted hyaena mortality patterns. *Trans R Soc South Africa* **48**:219–232.  
1440 doi:10.1080/00359199309520272

1441 Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long  
1442 uncorrected reads. *Genome Res* **27**:737–746. doi:10.1101/GR.214270.116

1443 vonHoldt BM, Kays R, Pollinger JP, Wayne RK. 2016. Admixture mapping identifies introgressed  
1444 genomic regions in North American canids. *Mol Ecol* **25**:2443–2453. doi:10.1111/mec.13667

1445 Waetjen DP, Shilling FM. 2017. Large extent volunteer roadkill and wildlife observation systems as  
1446 sources of reliable data. *Front Ecol Evol* **5**:89. doi:10.3389/fevo.2017.00089

1447 Walhund S. 2010. Zusammensetzung von populationen und korrelationserscheinungen vom  
1448 standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**:65–106. doi:10.1111/j.1601-  
1449 5223.1928.tb02483.x

1450 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman  
1451 J, Young SK, Earl AM. 2014. Pilon: an Integrated tool for comprehensive microbial variant

1452 detection and genome assembly improvement. *PLoS One* **9**:e112963.  
1453 doi:10.1371/journal.pone.0112963

1454 Walton LR, Joly DO. 2003. *Canis mesomelas*. *Mamm Species* **715**:1–9. doi:10.1644/715  
1455

1456 Wang W, Das A, Kainer D, Schalamun M, Morales-Suarez A, Schwessinger B, Lanfear R. 2020. The  
1457 draft nuclear genome assembly of *Eucalyptus pauciflora*: a pipeline for comparing de novo  
1458 assemblies. *Gigascience* **9**. doi:10.1093/gigascience/giz160

1459 Wang W, Das A, Kainer D, Schalamun M, Morales-Suarez A, Schwessinger B, Lanfear R. 2019. The  
1460 draft nuclear genome assembly of *Eucalyptus pauciflora*: new approaches to comparing de novo  
1461 assemblies. *bioRxiv* 678730. doi:10.1101/678730

1462 Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva E V,  
1463 Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and  
1464 phylogenomics. *Mol Biol Evol* **35**:543–548. doi:10.1093/molbev/msx319

1465 Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, Holmes L, Sogoloff B, Tabbaa D, Williams L,  
1466 Russ C, Nusbaum C, Lander ES, Maccallum I, Jaffe DB. 2014. Comprehensive variation  
1467 discovery in single human genomes. *Nat Genet* **46**:1350–1355. doi:10.1038/ng.3121

1468 Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A,  
1469 Kolesnikov A, Olson ND, Töpfer A, Alonge M, Mahmoud M, Qian Y, Chin C-S, Phillippy AM,  
1470 Schatz MC, Myers G, DePristo MA, Ruan J, Marschall T, Sedlazeck FJ, Zook JM, Li H, Koren  
1471 S, Carroll A, Rank DR, Hunkapiller MW. 2019. Accurate circular consensus long-read  
1472 sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol*  
1473 **37**:1155–1162. doi:10.1038/s41587-019-0217-9

1474 Westbury M V, Hartmann S, Barlow A, Wiesel I, Leo V, Welch R, Parker DM, Sicks F, Ludwig A,  
1475 Dalén L, Hofreiter M. 2018. Extended and continuous decline in effective population size results  
1476 in low genomic diversity in the world's rarest hyena species, the brown hyena. *Mol Biol Evol*  
1477 **35**:1225–1237. doi:10.1093/molbev/msy037

1478 Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford  
1479 Nanopore sequencing. *Genome Biol* **20**:129. doi:10.1186/s13059-019-1727-y

1480 Wickham H. 2016. *Ggplot2: elegant graphics for data analysis*. Springer.

1481 Wilke CO. 2016. *cowplot: streamlined plot theme and plot annotations for 'ggplot2.'* *CRAN Repos*.

1482 Wilson DE, Mittermeier RA, Cavallini P. 2009. *Handbook of the mammals of the world, Vol. 1. ed.*  
1483 Barcelona: Lynx Edicions.

1484 Yu L, Zhang Y. 2006. Phylogeny of the caniform Carnivora: evidence from multiple genes. *Genetica*  
1485 **127**. doi:10.1007/S10709-005-2482-4

1486 Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M,  
1487 Loetscher A, Kriventseva E V. 2017. OrthoDB v9.1: cataloging evolutionary and functional  
1488 annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res*  
1489 **45**:D744–D749. doi:10.1093/nar/gkw1119

1490 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree  
1491 reconstruction from partially resolved gene trees. *BMC Bioinforma* **19**:15–30.  
1492 doi:10.1186/s12859-018-2129-y

1493 Zimin A V., Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome  
1494 assembler. *Bioinformatics* **29**:2669–2677. doi:10.1093/bioinformatics/btt476

1495 Zimin A V., Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL. 2017.

Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res* **27**:787–792. doi:10.1101/GR.213405.116

Zoonomia consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**:240–245. doi:10.1038/s41586-020-2876-6

## Additional files

**Figure 3 – Figure supplement 1:** Genetic differentiation indices obtained from a comparison of intraspecific (orange) and interspecific (red) polymorphisms after having homogenized the coverage of all species (at about 15x). The estimates were calculated for four pairs of well-defined Carnivora species and for the subspecies of aardwolf (*Proteles cristatus*) and bat-eared fox (*Otocyon megalotis*) (grey). Silhouettes from <http://phylopic.org/>.

**Figure 3 – Figure supplement 2:** Genetic differentiation indices obtained from the comparison of intraspecific (orange) and interspecific (red) polymorphisms for the pair *Ursus arctos*/*Ursus maritimus* (~10 replicates per species). GDI is estimated for each pair of individuals. This result demonstrates that randomly picking only three individuals (out of 10) is sufficient to accurately estimate the level of genetic differentiation between the two species.

**Supplementary File 1:** Pairwise patristic distances estimated for the 142 species based on branch lengths of the phylogenetic tree inferred with the 15 mitochondrial loci (2 rRNAs and 13 protein-coding genes).

**Supplementary File 2:** Results of Bayesian dating for the two nodes leading to the *Proteles cristatus* spp. and the *Otocyon megalotis* spp.. Divergence time estimates based on UGAM and LN models are reported with associated 95% credibility intervals for each MCMC chain.

**Supplementary File 3:** Sample details and assembly statistics (Number of contigs/scaffolds and associated N50 values) for the 503 mammalian assemblies retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/assembly>) on August 13th, 2019 with filters: “Exclude derived from surveillance project”, “Exclude anomalous”, “Exclude partial”, and using only the RefSeq assembly for *Homo sapiens*.

**Supplementary File 4:** Genome completeness assessment of MaSuRCA and SOAPdenovo assemblies obtained for *Proteles cristatus cristatus* and *Otocyon megalotis megalotis* together with the 63 carnivoran assemblies available at NCBI on August 13th, 2019 using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3 with the Mammalia OrthoDB 9 BUSCO gene set.

**Supplementary File 5:** Annotation summary and supermatrix composition statistics of the 53 species used to infer the genome-scale Carnivora phylogeny.

**Supplementary File 6:** Statistics on morphological measures of the current subspecies of *Proteles cristatus*.

**Supplementary File 7:** Skull measurements of *Proteles* taxa from museum specimens and the literature (Allen 1909, Heller 1913 Hollister 1918, Roberts 1932, 1951)

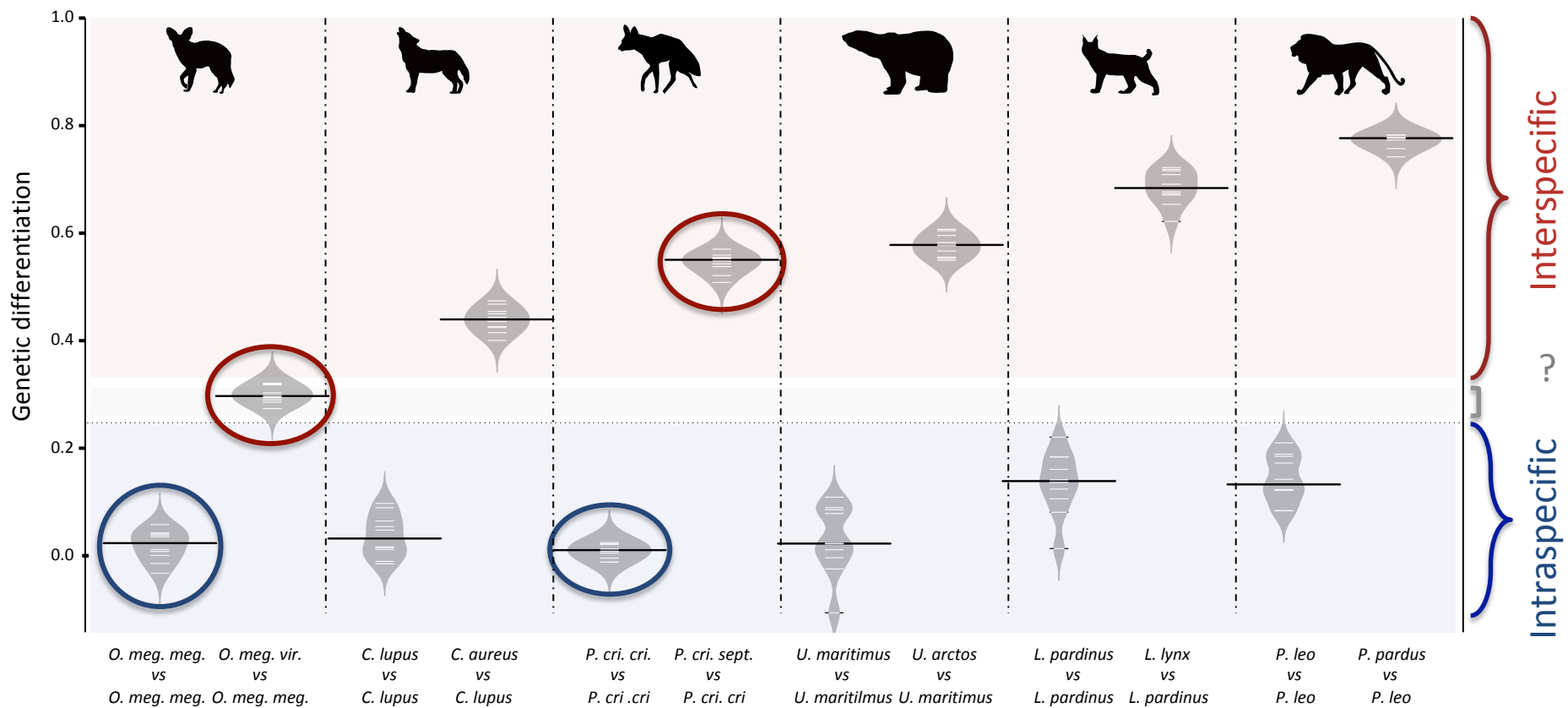
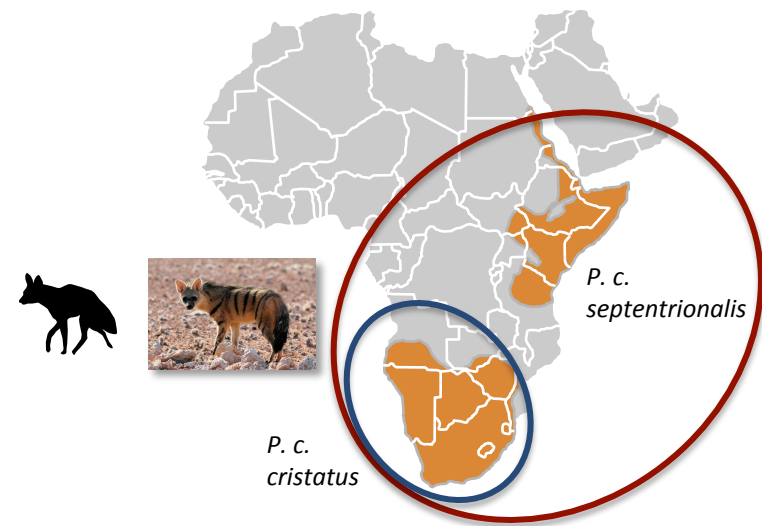
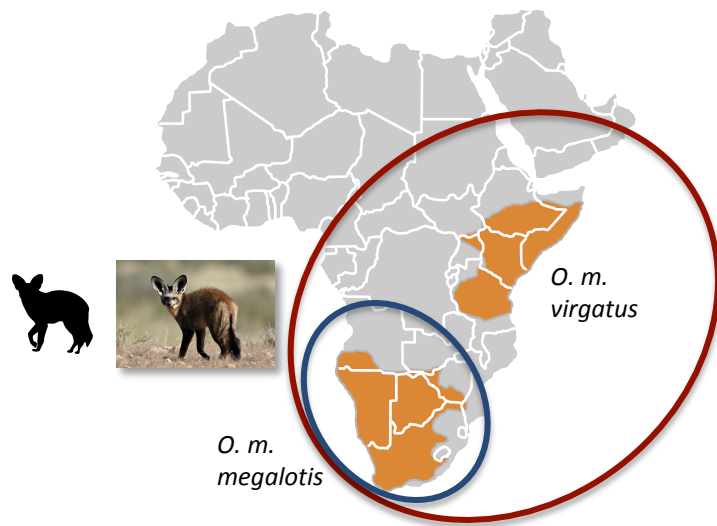
**Supplementary File 8:** Sample details and assembly statistics of the 13 newly assembled carnivoran mitochondrial genomes.

**Supplementary File 9:** Node calibrations used for the Bayesian dating inferences based on mitogenomic data.

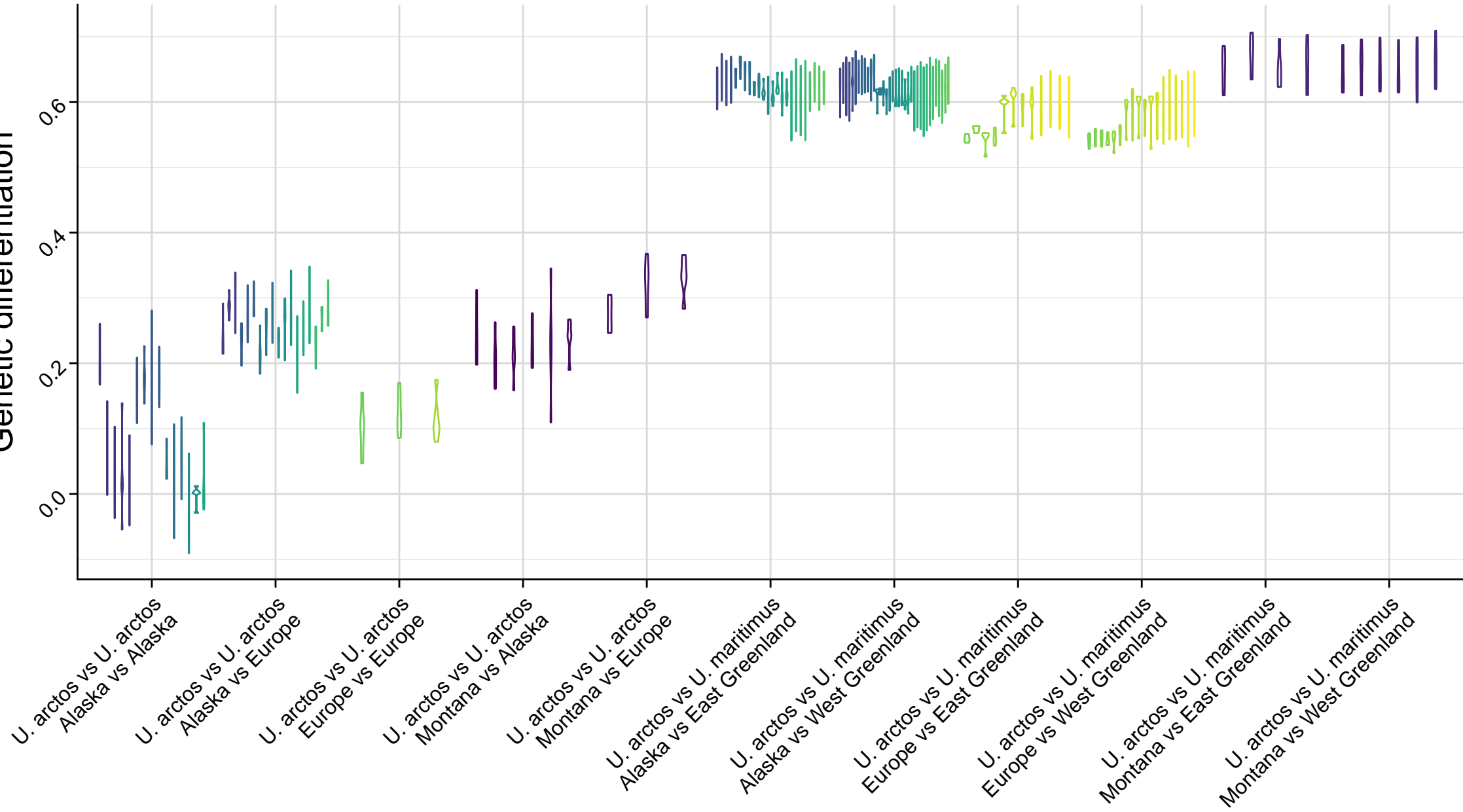
**Supplementary File 10:** Results of contamination analyses performed with BlobTools for the aardwolf (*Proteles cristatus cristatus*).

**Supplementary File 11:** Results of contamination analyses performed with BlobTools for the bat-eared fox (*Otocyon megalotis megalotis*).

**Supplementary File 12:** Summary information for the Carnivora genomes available either on GenBank, DNA Zoo and the OrthoMaM database as of February 11th, 2020. The “OMM” column indicates if the genome was available on OMM (yes) or not (no). The “Annotation” column indicates whether the genome was already annotated (yes) or not (no).



Genetic differentiation



## Appendix1

### Difference between *Fast* and *High accuracy* modes of Guppy basecaller

For MinION sequencing, basecalling of fast5 files was performed using Guppy v3.1.5 (developed by ONT) with the *high accuracy* option, which takes longer but is more accurate than the standard *fast* model

**Appendix 1 – Figure 1:** Plot of the quality of Nanopore long reads base-called with either the *fast* or the *high accuracy* option of Guppy v3.1.5. The quality of the base-calling step has a large impact on the final quality of the assemblies by reducing the number of contigs and increasing the N50 value.

### Genome quality assessments

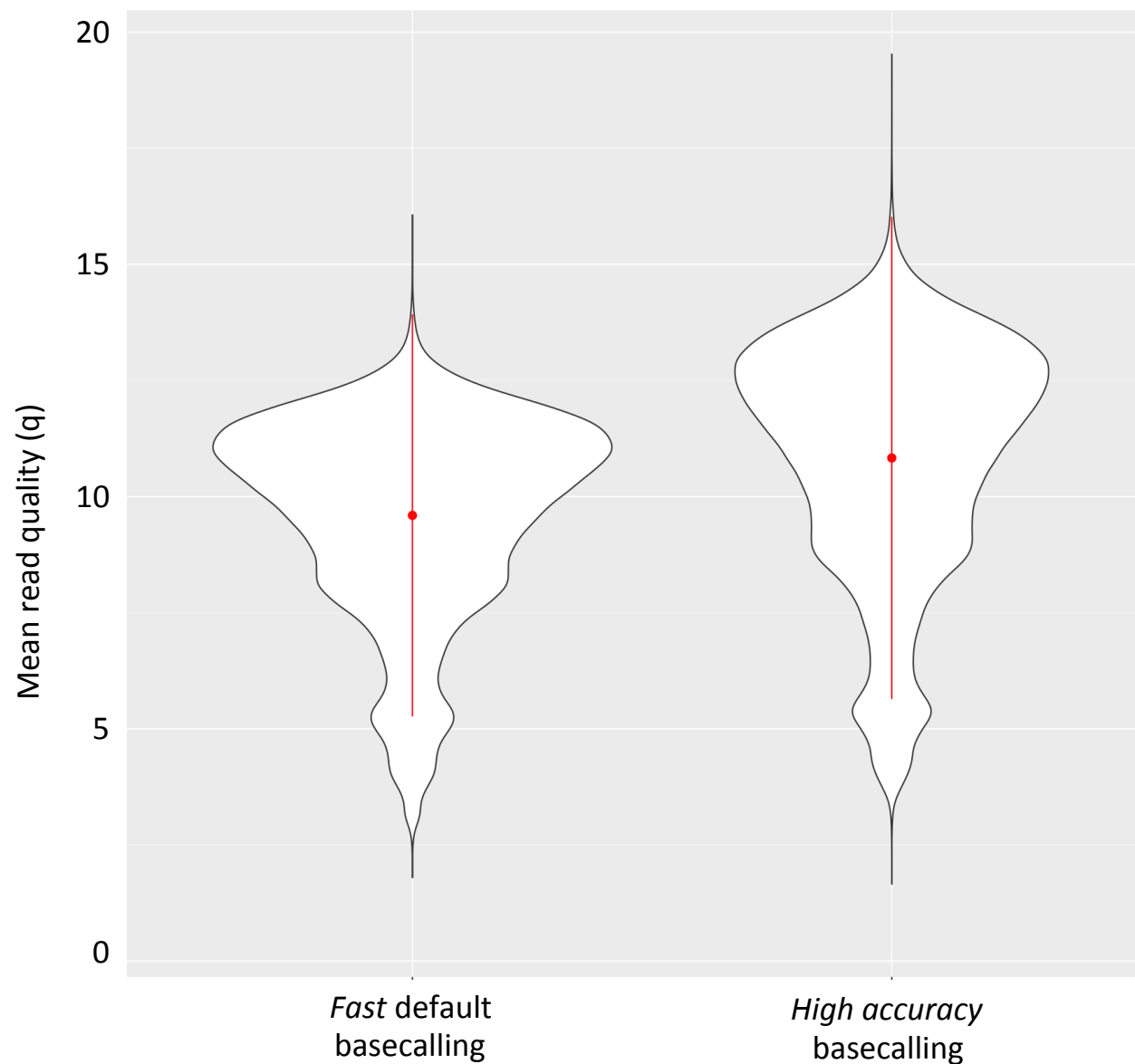
Exhaustive comparisons with 503 available mammalian assemblies revealed a large heterogeneity among taxonomic groups and a wide variance within groups in terms of both number of scaffolds and N50 values (**Figure 2, Supplementary File 3**). Xenarthra was the group with the lowest quality genome assemblies, with a median number of scaffolds of more than one million and a median N50 of only 15 kb. Conversely, Carnivora contained genome assemblies of much better quality, with a median number of scaffolds of 15,872 and a median N50 of 4.6 Mb, although a large variance was observed among assemblies for both metrics (**Figure 2 Supplementary File 3**). Our two new genomes compared favourably with the available carnivoran genome assemblies in terms of contiguity showing slightly less than the median N50 and a lower number of scaffolds than the majority of the other assemblies (**Figure 2, Supplementary File 3**). Comparison of two hybrid assemblies with Illumina-only assemblies obtained with SOAPdenovo illustrated the positive effect of introducing Nanopore long reads even at moderate coverage by reducing the number of scaffolds from 409,724 to 5,669 (aardwolf) and from 433,209 to 11,081 (bat-eared fox) while increasing the N50 from 17.3 kb to 1.3 Mb (aardwolf) and from 22.3 kb to 728 kb (bat-eared fox). With regard to completeness based on 4,104 single-copy mammalian BUSCO orthologues, our two hybrid assemblies are among the best assemblies with more than 90% complete BUSCO genes and less than 4% missing genes (**Figure 3, Supplementary File 4**). As expected, the two corresponding Illumina-only assemblies were much more fragmented and had globally much lower BUSCO scores (**Figure 3, Supplementary File 4**).

**Appendix 1 – Figure 2:** Comparison of 503 mammalian genome assemblies from 12 taxonomic groups using bean plots of the a) number of scaffolds, and b) scaffold N50 values ranked by median values. Thick black lines show the medians, dashed black lines represent individual data points, and polygons represent the estimated density of the data. Note the log scale on the Y axes. The bat-eared fox (*Otocyon megalotis megalotis*) and aardwolf (*Proteles*



*cristatus cristatus*) assemblies produced in this study using SOAPdenovo and MaSuRCA are indicated by asterisks. Bean plots were computed using BoxPlotR (Spitzer et al., 2014).

**Appendix 1 – Figure 3:** BUSCO completeness assessment of 67 Carnivora genome assemblies visualized as bar charts representing percentages of complete single-copy (light blue), complete duplicated (dark blue), fragmented (yellow), and missing (red) genes ordered by increasing percentage of total complete genes. The bat-eared fox (*Otocyon megalotis megalotis*) and aardwolf (*Proteles cristatus cristatus*) assemblies produced in this study using MaSuRCA and SOAPdenovo are indicated by asterisks.



### Aardwolf

*Fast default  
basecalling*

*High accuracy  
basecalling*

**Contigs:** 8,874



5,669

**N50:** 699 Kb



1.31 Mb

### Bat-eared fox

*Fast default  
basecalling*

*High accuracy  
basecalling*

**Contigs:** 12,735



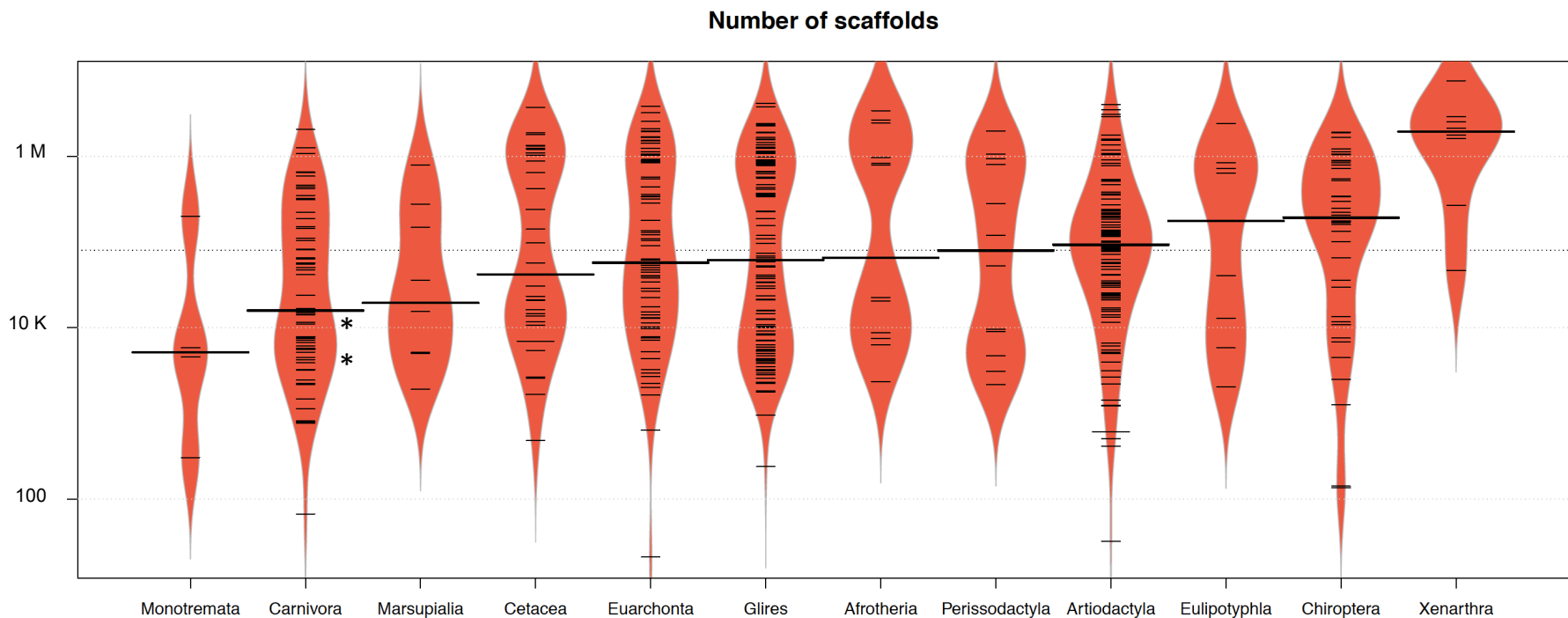
11,081

**N50:** 676 Kb

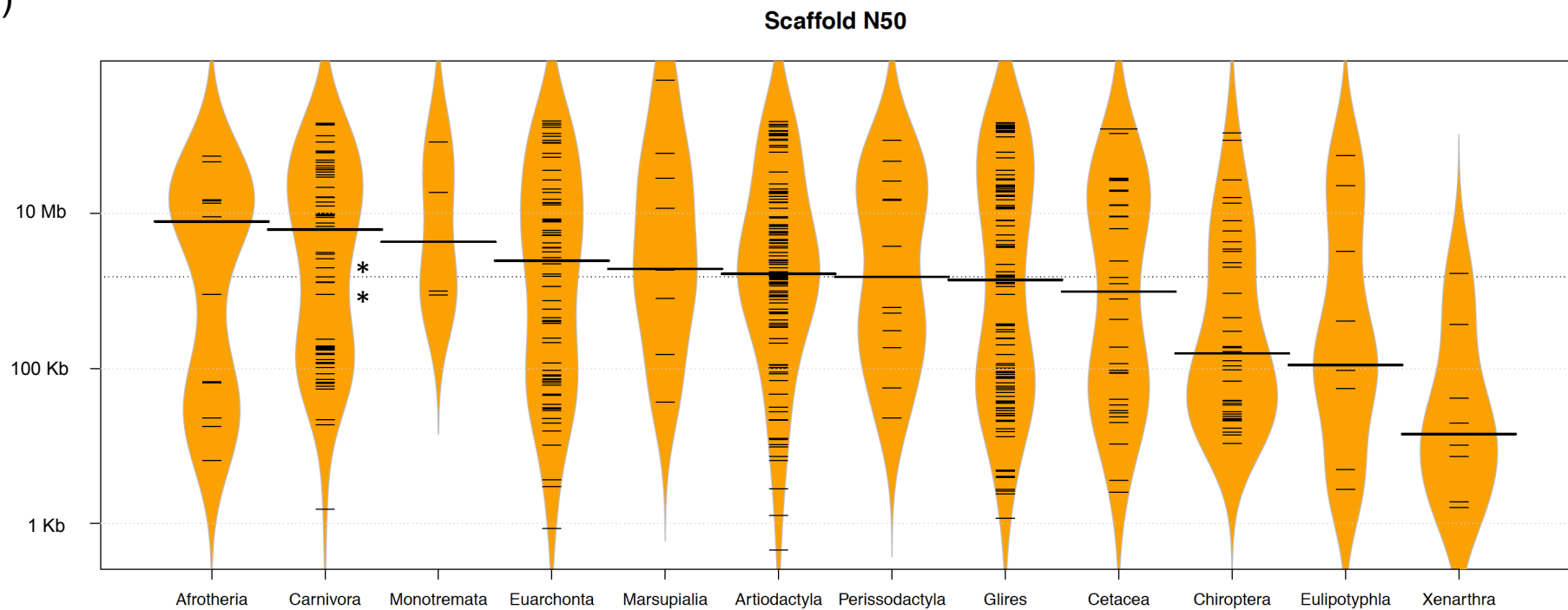


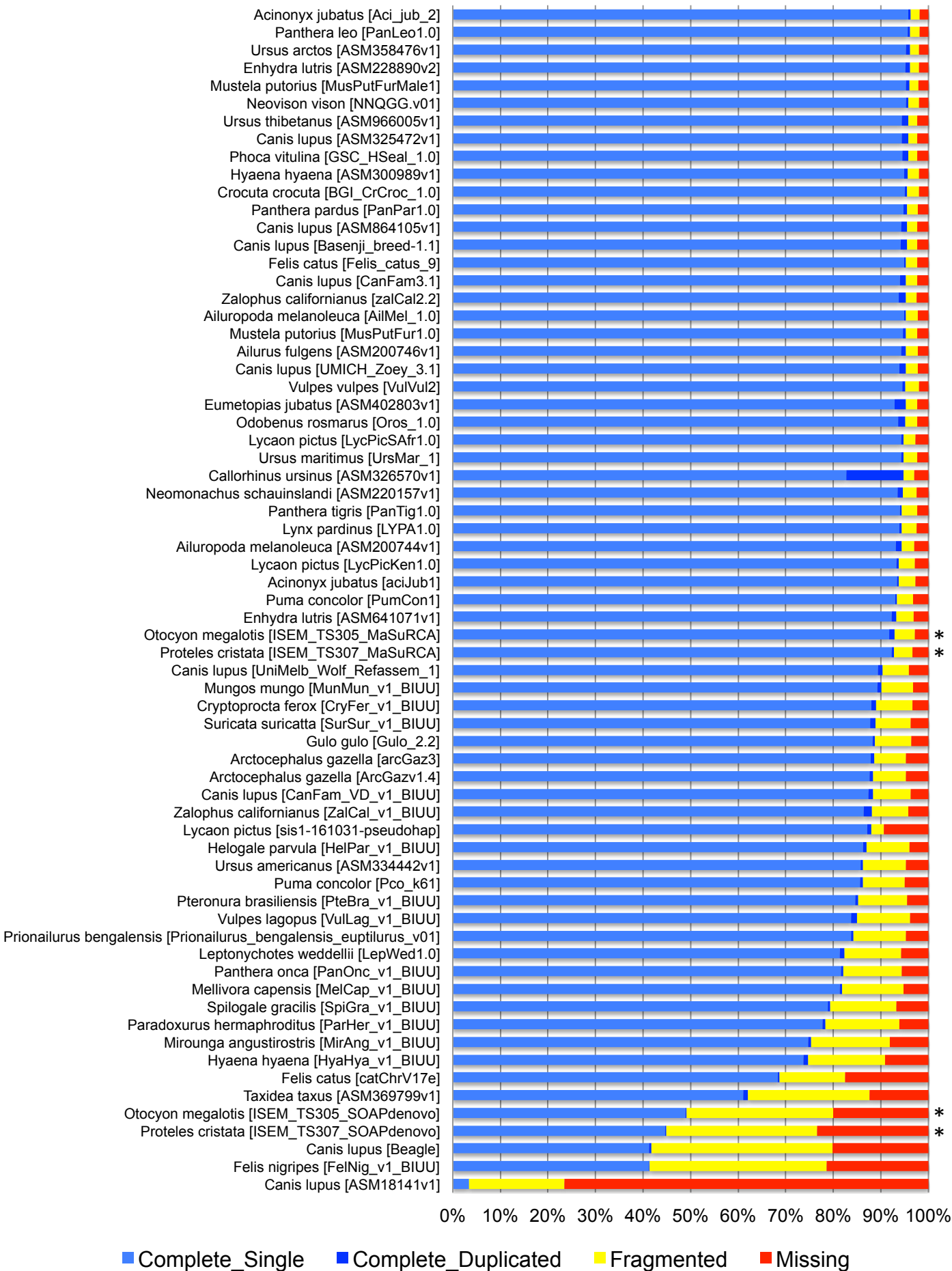
728 Kb

a)



b)





## Appendix 2 - Morphological differences between *Proteles* taxa

### 1. Differences in fur colouration and markings

Cabrera (1910) described how the fur of *pallidior* is unicolored and lacks the brown base of *cristatus*. This latter character appears to be consistent in an Ethiopian specimen in National Museums Scotland (NMS.Z.1877.15.5) compared with three skins of *cristatus* of Namibian and South African origin (NMS.Z.2020.44, NMS.Z.2020.46.1 and NMS.Z.2020.46.6) also in the collections of National Museums Scotland (Figure 1), although it would appear to be a difference in the coloration of the underfur. However, a Zimbabwean specimen (NMS.Z.1950.68) also had only pale underfur, which appears to contradict Cabrera (1910), so the usefulness of this character is in doubt.

**Appendix 2 – Figure 1:** Unicolored fur of an Eastern aardwolf from Ethiopia (NMS.Z.1877.15.5) (A) and bicoloured fur of a Southern aardwolf of South African origin (NMS.Z.2020.44) (B).

In reviewing georeferenced photographs of aardwolves from throughout the range, the striping pattern appeared to be variable, but overall East African specimens tended to be paler, with more contrasting stripes with a pale forehead compared with the longer, greyer or ochre-grey fur in Southern African specimens, which have broader less distinctive stripes (A.C.K. pers. obs.). However, fur length and hence stripe distinctiveness may just be a phenotypic response to lower temperatures at higher latitudes compared with equatorial East African specimens.

Additional preliminary observations were made on pelage coloration and markings based on the skins above and live specimens of both taxa kept at Hamerton Zoo Park, Cambridgeshire, UK. The live specimens offer a unique opportunity to examine these characters at the same latitude and environmental conditions, so that phenotypes should reflect genetic differences between taxa. Two pelage characters appear to be different between the two taxa. Firstly the stripes in *cristatus* tend to be broader and less well defined, whereas in *septrionalis* they are thinner, more contrasting and break up into spots on the neck. Secondly the forehead coloration is dark grizzled grey in *cristatus*, but lighter yellowish-grey or creamy-grey in *septrionalis*. Further investigation is required to examine pelage variation from throughout the ranges of both taxa to see if these characters are diagnostic and to determine additional diagnostic characters.

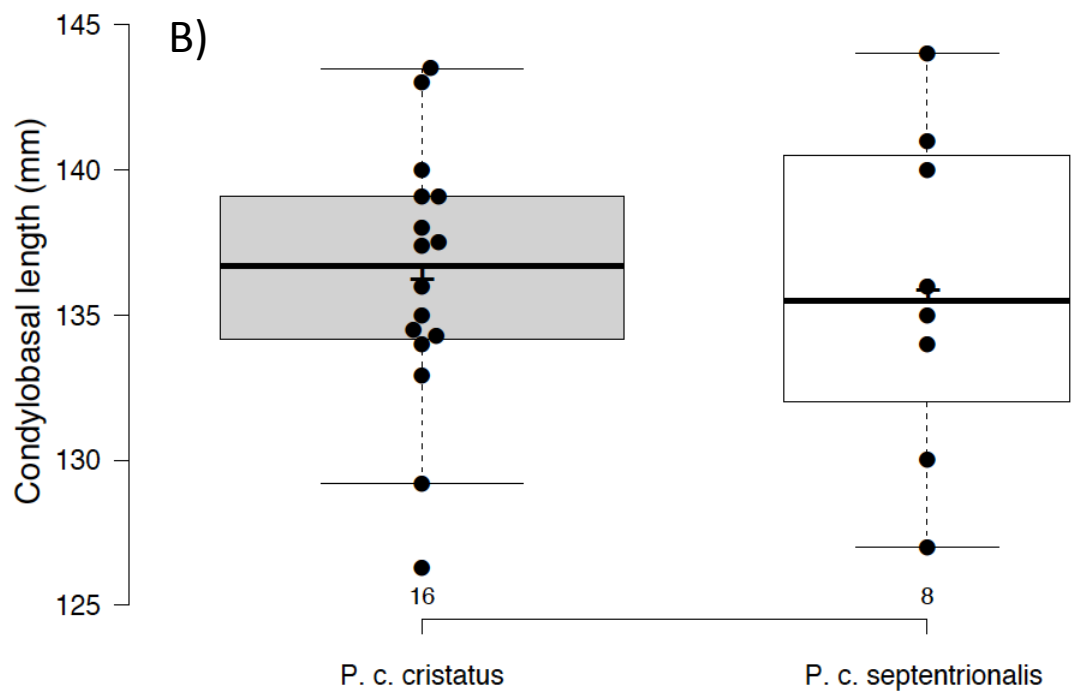
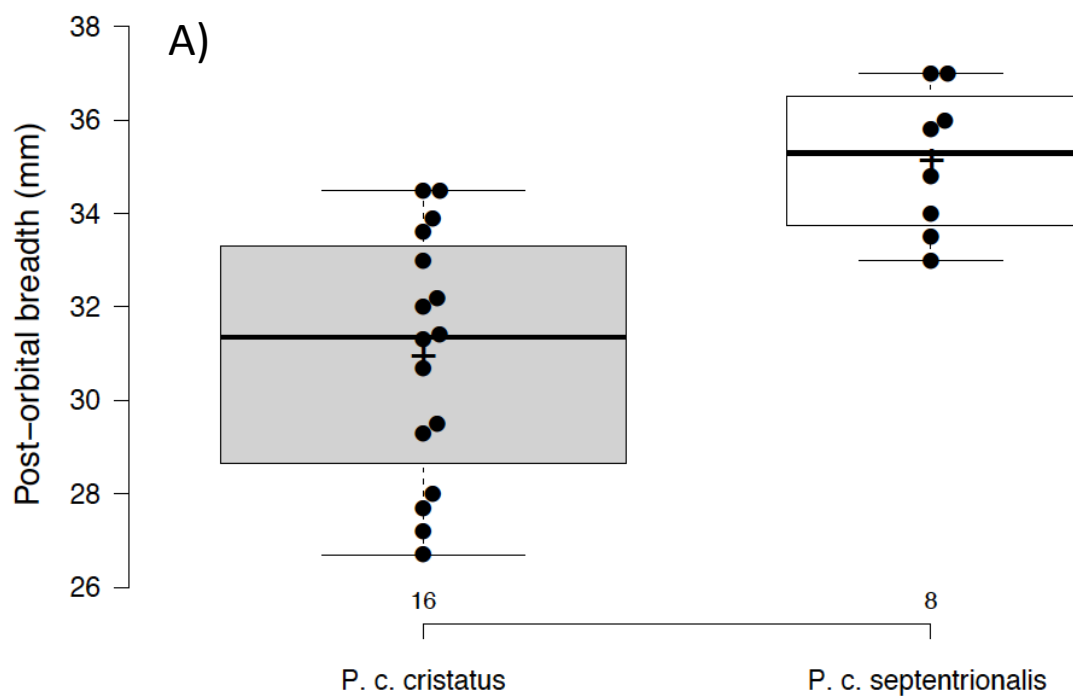
### 2. Skull morphometric analyses

In addition to skull measurements taken from specimens in the Natural History Museum, London (NHMUK), Museum of Vertebrate Zoology (MVZ) and National Museums Scotland (NMS), measurements of skulls were taken from the literature (Allen 1909, Heller 1913, Hollister 1918, Roberts 1932, 1951) (Table 1). Comparison of means confirmed that mean post-orbital breadth is significantly greater in *septrionalis* than in *cristatus* ( $t_{8,16}=4.10$ ,  $P<0.001$ ) (Figure 2). However, there are no differences between the means of other skull

measurements, including condylobasal length of skull (Figure 3), zygomatic width, inter-orbital breadth, brain-case width and mandible length ( all  $P>0.05$ ). As noted above with skins, sample sizes are small and thus the significant difference in mean post-orbital breadth between the two taxa remains tentative subject to examination of a larger sample.

**Appendix 2 - Figure 2:** Box and jitter plot of A) post-orbital breadths of *Proteles* taxa: *cristatus* (left) and *septrionalis* (right) and B) condylobasal lengths of skull of *Proteles* taxa: *cristatus* (left) and *septrionalis* (right). Graph generated with BoxPlotR (<http://shiny.chemgrid.org/boxplotr/>).







## F-Statistic:

$$F_{ST} = 1 - \frac{\pi_{\text{within}}}{\pi_{\text{between}}}$$

Hudson et al 1992

$F_{ST} = 1$  = Highly structured

$F_{ST} = 0$  = No structuration

## Genetic differentiation index (GDI, based on heterozygosity):

Fixed AA/TT	Private A AT/AA	Private B AA/AT	Shared AT/AT
----------------	--------------------	--------------------	-----------------

$$\pi_{AB} = \text{fixed} + \text{private}_A + \text{private}_B + \text{shared}_{AB}$$

$$\pi_A = \text{private}_A + \text{shared}_{AB}$$

$$\pi_B = \text{private}_B + \text{shared}_{AB}$$

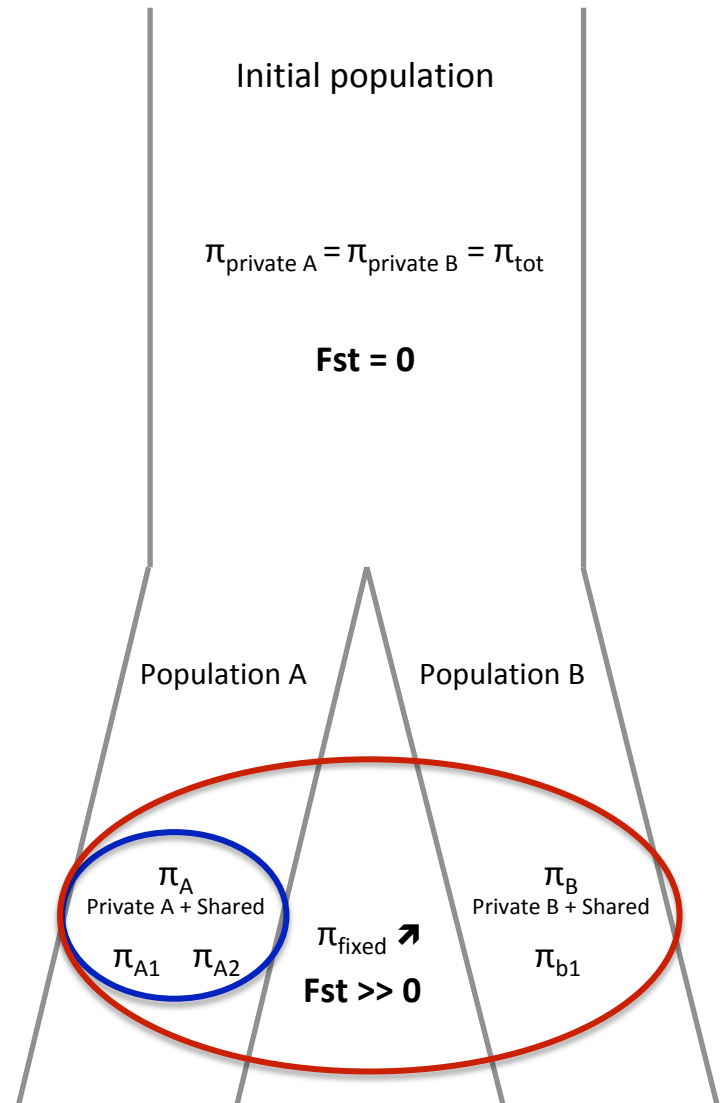
$$1 - \frac{(\pi_A + \pi_B) / 2}{\pi_{\text{tot AB}}}$$

## Comparing intra vs extra population GDIs:

$$1 - \frac{(\pi_{A1} + \pi_{A2}) / 2}{\pi_{\text{tot A}}} \quad \text{vs} \quad 1 - \frac{(\pi_{A1} + \pi_{B1}) / 2}{\pi_{\text{tot A1B1}}}$$

GDI within pop A  
(control)

global GDI

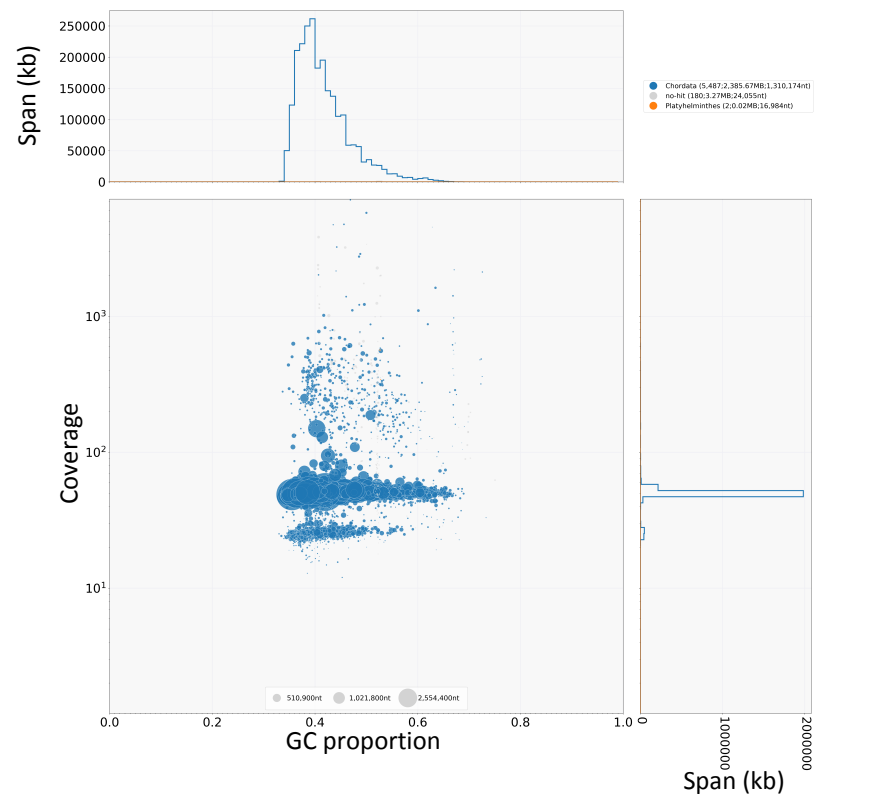


### **Appendix 3 – Genetic differentiation index**

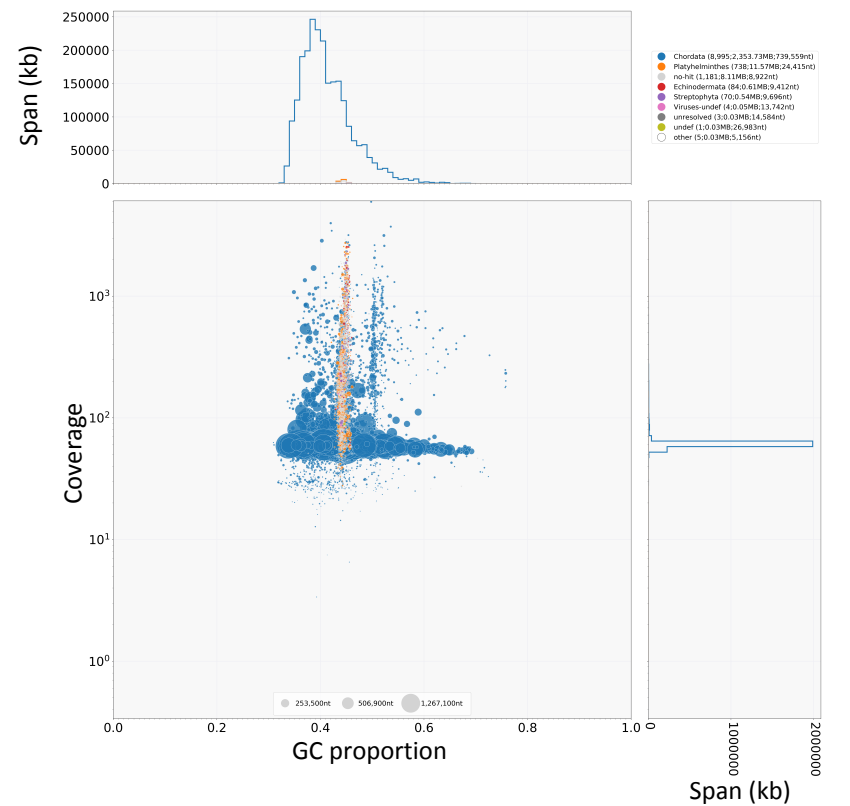
To estimate the level of genetic differentiation between two populations, we developed a new index based on the heterozygosity of at least one individual of each population (**Appendix 3 – Figure 1**).

**Appendix 3 – Figure 1:** Definition of the genetic differentiation index (GDI) based on the F-statistic (FST). The main difference between these two indexes is the use of heterozygous allele states for GDI rather than real polymorphism for the FST. Green =  $\pi_{\text{within}}$ , Orange =  $\pi_{\text{between}}$ , Blue = Population A, Red = Population A+B.

a) BlobTools results for *Proteles cristatus*



b) BlobTools results for *Otocyon megalotis*



## Appendix 4 – Contigs selection for genetic differentiation analyses.

Using Blobtools (Laetsch and Blaxter, 2017), we were able to specifically select the Carnivora contigs for further analyses (**Appendix 4 – Figure 1, Supplementary Files 10-11**). Additionally, contigs likely belonging to X chromosome were identified and removed based on LASTZ (Rahmani et al., 2011) alignments (contigs that align with cat or dog autosomes and not to X chromosome have been selected).

**Appendix 4 – Figure 1.** Graphical representation (BlobPlot) of the results of contamination analyses performed with BlobTools for a) the aardwolf (*Proteles cristatus cristatus*) and b) the bat-eared fox (*Otocyon megalotis megalotis*) genome assemblies.