

RESEARCH

Open Access



# Mitogenome sequences of domestic cats demonstrate lineage expansions and dynamic mutation processes in a mitochondrial minisatellite

Emily C. Patterson<sup>1</sup>, Gurdeep Matharu Lall<sup>1</sup>, Rita Neumann<sup>1</sup>, Barbara Ottolini<sup>1,2</sup>, Chiara Batini<sup>1,3,4</sup>, Federico Sacchini<sup>5</sup>, Aiden P. Foster<sup>6</sup>, Jon H. Wetton<sup>1\*</sup> and Mark A. Jobling<sup>1\*</sup>

## Abstract

**Background** As a population genetic tool, mitochondrial DNA is commonly divided into the ~1-kb control region (CR), in which single nucleotide variant (SNV) diversity is relatively high, and the coding region, in which selective constraint is greater and diversity lower, but which provides an informative phylogeny. In some species, the CR contains variable tandemly repeated sequences that are understudied due to heteroplasmy. Domestic cats (*Felis catus*) have a recent origin and therefore traditional CR-based analysis of populations yields only a small number of haplotypes.

**Results** To increase resolution we used Nanopore sequencing to analyse 119 cat mitogenomes via a long-amplification approach. This greatly improves discrimination (from 15 to 87 distinct haplotypes in our dataset) and defines a phylogeny showing similar starlike topologies within all major clades (haplogroups), likely reflecting post-domestication expansion. We sequenced RS2, a CR tandem array of 80-bp repeat units, placing RS2 array structures within the phylogeny and increasing overall haplotype diversity. Repeat number varies between 3 and 12 (median: 4) with over 30 different repeat unit types differing largely by SNVs. Five SNVs show evidence of independent recurrence within the phylogeny, and seven are involved in at least 11 instances of rapid spread along repeat arrays within haplogroups.

**Conclusions** In defining mitogenome variation our study provides key information for the forensic genetic analysis of cat hair evidence, and for the first time a phylogenetically informed picture of tandem repeat variation that reveals remarkably dynamic mutation processes at work in the mitochondrion.

**Keywords** Domestic cat, Mitochondrial DNA, Haplogroup, Control region, Tandem repeat, Nanopore sequencing

\*Correspondence:

Jon H. Wetton  
jw418@le.ac.uk  
Mark A. Jobling  
maj4@le.ac.uk

Full list of author information is available at the end of the article



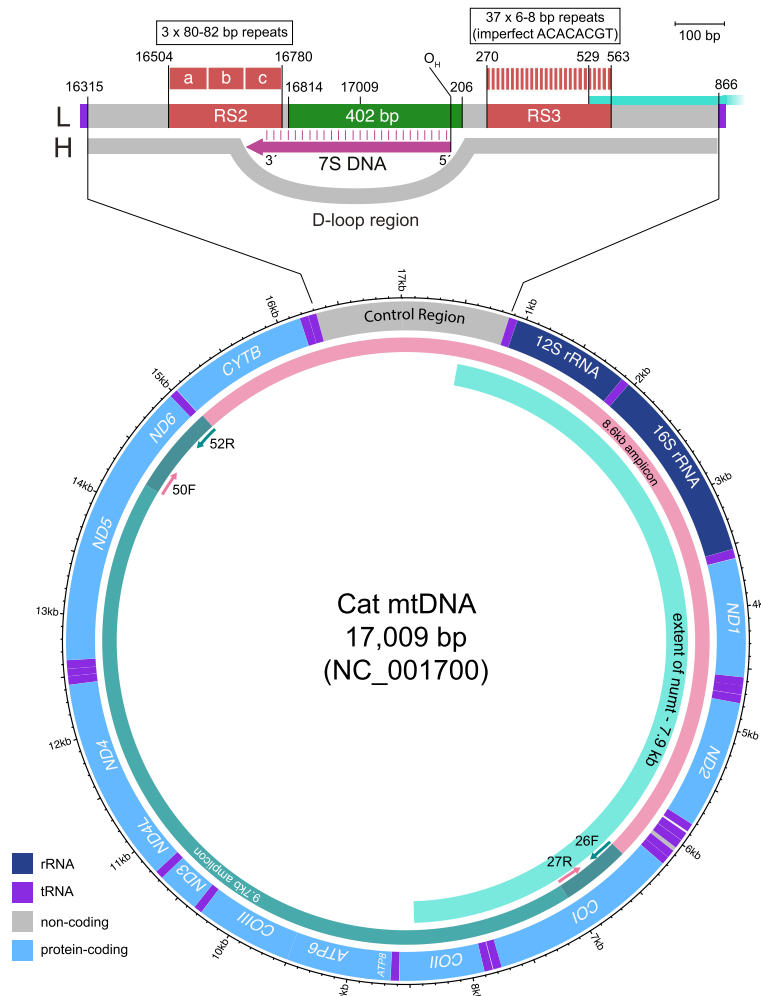
© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

### Background

Mitochondrial DNA (mtDNA) has long been a widely used tool for the genetic analysis of animal populations [1, 2] because of its unusual properties compared to the nuclear genome: a relatively high mutation rate that contributes to high diversity; a lack of recombination that allows a simple phylogeny to be constructed; and uniparental inheritance that provides information about female population histories. Its high copy number also makes it useful in animal forensic applications [3, 4], with highly degraded or trace materials often yielding mtDNA data when nuclear DNA analysis is not possible.

Most animal mitochondrial genomes are 16–17 kb in size (Fig. 1) with ~95% of this being genic, typically comprising 13 protein-coding, 22 tRNA, and two rRNA

genes. However, one segment, the ~1-kb control region (CR) includes no coding elements, instead containing sequences that control replication and transcription, and the site of the D-loop, a triple-stranded region formed by stable incorporation of a short third DNA strand (7S DNA; [5]). Constraint on sequence evolution in the CR is weaker than in the coding region and consequently it shows higher sequence diversity [6, 7], making it a traditional focus of mtDNA diversity studies. As well as single-nucleotide variation, in many species the CR contains tandemly-repeated sequences [8, 9] that can vary in copy number among individuals. Such sequences provide a potentially useful source of variation, but interpretation is complicated by the fact that they tend to display heteroplasmy [9]—the



**Fig. 1** Organisation of the domestic cat mitogenome. Structure of cat mtDNA based on the NC\_001700 reference sequence [10], showing genes and the extent of the large numt. Primers for the two long overlapping amplicons used here are indicated by short coloured arrows. Above is an expanded view of the control region, including the positions of the RS2 and RS3 repeat arrays, the 402-bp segment usually sequenced, the origin of heavy strand replication ( $O_H$ ), the D-loop and the 7S DNA. L: light strand; H: heavy strand

presence of multiple different length variants within a cell, tissue or individual.

Sequencing the entire molecule is clearly the best approach to maximising the information from mtDNA and over time this approach has been overtaking CR-based studies. One strategy is to design multiple primer pairs that amplify the mitogenome as a set of short overlapping fragments that can be sequenced either using Sanger technology or, more recently, massively parallel sequencing methods [11]. However, a problem with this approach is the presence in animal nuclear genomes of numts (nuclear mitochondrial segments [12, 13]), some of which exist in high copy number. If primer sequences match numt copies, then these may be amplified alongside, or instead of, the true mtDNA, providing misleading information.

Maximising the detection of mtDNA diversity is particularly important in domesticated animals, whose histories are dominated by founder events and bottlenecks reflecting small numbers of domestication events and selective breeding. The domestic cat (*Felis catus*) has a relatively recent history, with the earliest claimed evidence for domestication dating to 9.5 KYA (thousand years ago) in Cyprus [14], followed by more extensive evidence from Egypt some 4000 years later [15]. Population studies focusing on a 402-bp region of the CR (Fig. 1) identify only twelve haplotypes (also known as 'mitotypes') designated A-L, with just four of these (A-D) representing 60–70% of individuals in worldwide cat populations [16]. The sequenced segment is flanked by tandem repeat arrays known as RS2 and RS3 [10]; in the cat mtDNA reference sequence (NC\_001700 [10]) RS2 contains three copies of a ~80–82-bp repeat, while RS3 contains 37 copies of a 6–8-bp repeat. Polymorphism and heteroplasmy have been reported for RS2 in big cats [17] and for RS3 in domestic cats [18]. This complex variability leads to these sequences being generally avoided in diversity studies. An additional complicating factor in domestic cat mitogenome sequencing is the presence on nuclear chromosome D2 of a tandemly repeated 7.9-kb numt (Fig. 1) that corresponds to a mitochondrial segment spanning from the CR 3' end to the *COII* gene [12].

The amplification of mtDNA as two large overlapping amplicons via long PCR provides a means to avoid common multi-copy numts, by the judicious placement of primer sites. The resulting amplicons are also amenable to sequencing using long-read technologies such as Nanopore sequencing, which provides a cost-effective way to survey whole mitogenome variation. In this study, we use this technology to generate whole mitogenome sequences for a population of domestic cats, providing a greater understanding of diversity, illuminating RS2 repeat sequence evolution in a phylogenetic context, and

contributing data on the potential discriminatory power of mitogenomes in forensic applications.

## Results

To assess whole mitogenome variation in domestic cats we undertook Nanopore sequencing of two overlapping mtDNA amplicons (8.6 and 9.7 kb in length) in blood DNA samples from 93 unrelated individuals (Table S1). Altogether, sequencing generated 3.2 million reads, of which ~60% were successfully allocated to one of 12 native barcodes; of those reads, 43% were then demultiplexed based on their custom barcodes, allowing assignment to individual cats. After filtering for quality and read-length, a total of 706,528 reads were taken for further analysis (an average of 5709 reads per individual for the 8.6-kb amplicon [minimum: 842 reads], and an average of 1887 reads for the 9.7-kb amplicon [minimum: 166]). We also generated 35 mitogenome sequences using the same PCR primers via Illumina sequencing; nine mitogenomes were sequenced with both technologies. A consensus sequence was generated for each mitogenome and the Nanopore sequences were combined with the Illumina-generated sequences and aligned.

Although Nanopore sequencing provides access to long reads, a disadvantage of the LSK109 chemistry used for this project was its relatively high error rate [19], particularly within homopolymeric runs. Among the 93 domestic cats, ten apparent single-base indels were identified within homopolymeric runs that were 5–8 nucleotides in length according to the reference sequence (Table S3a). Four of these lie within protein-coding regions and, if genuine, would represent frameshift mutations, suggesting that they are artefacts; all four were absent from the Illumina domestic cat sequences. Homopolymer variation in Illumina data suggests that four of the non-coding single-base deletions observed in Nanopore data may be genuine (Table S3a) but to be conservative we exclude all such deletions from analysis. With the exception of the homopolymer indels, and the repetitive RS2 and RS3 regions (discussed below) a general comparison between Nanopore and Illumina data showed complete concordance.

### Single-nucleotide variation within the cat mitogenome assessed by long-amplicon sequencing

In considering sequence variation in the cat mitogenomes, we excluded the repetitive region RS3 (Fig. 1a) since its short (6–8-bp) repeat units and overall length (~300 bp in the reference sequence) make it difficult to interpret (Figure S1). In the reference sequence, the repetitive region RS2 (Fig. 1a) contains fewer, longer repeat units (three copies of an 80–82 bp element). However, our sequence data demonstrate that this

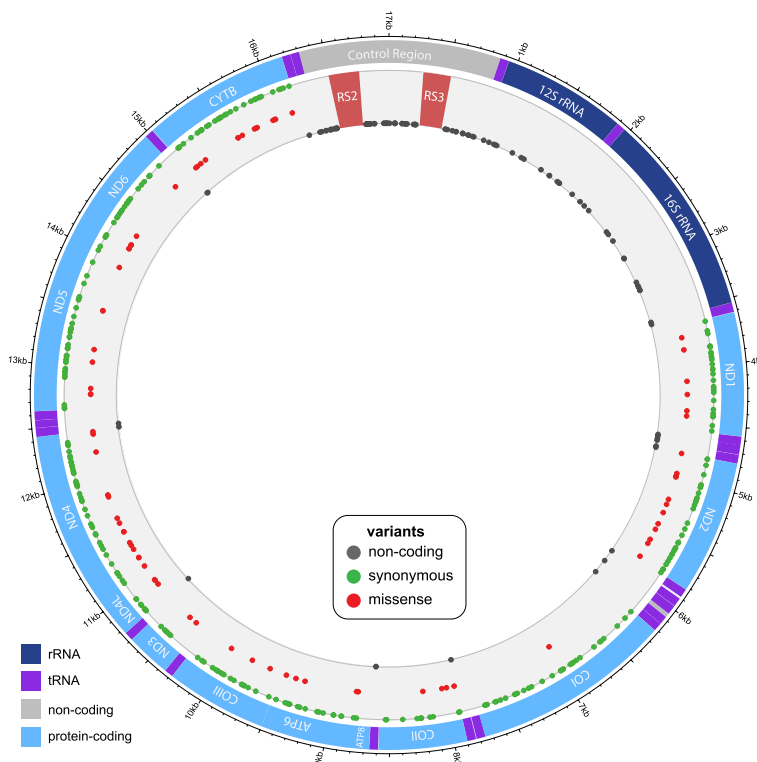
array contains variation both in repeat sequence and copy number which we consider in the following section. In this section, we restrict our attention to single-nucleotide variation in the mitogenome excluding RS2 and RS3: we can therefore compare variants among cats over a total of 16,438 bp (coordinates: 1–269, 564–16503, 16781–17009).

To understand the variants in our generated sequences we aligned them to the cat reference sequence. In doing this we noted 40 base substitutions and one indel that appear fixed in our sequence dataset compared to the reference. These either represent true private variants or errors within the reference sequence, which was generated by Sanger technology [10]. The reference-specific variants show an unexpectedly strong (17:23) bias towards transversions, and 15/29 coding region variants among them represent apparent missense mutations (Table S4), which supports the idea that they are sequencing artefacts.

The high observed concordance between Nanopore and Illumina mitogenome sequence data allows us to combine the two datasets and consider this as a unified

set of 119 cat mitogenomes. Our Nanopore variant calling approach considers the majority nucleotide at any position and so is insensitive to heteroplasmy in the sequence data; we therefore do not consider heteroplasmy in this analysis.

The 119 mitogenomes contain a total of 438 variant sites (Table S5), of which 435 are base substitutions and three are indels. The indels (not in homopolymeric runs) are listed in Table S3b, with all three lying in non-coding regions. Among base substitutions, the transition:transversion ratio for the whole mitogenome is 19.45, and that for the control region (coordinates 16315–16503, 16781–269, 564–865) is 4.87. These values are similar to those found in dogs [20], where a ratio of 18.22 was observed in the mitogenome (excluding a VNTR), although a higher transition:transversion ratio of 10.86 was seen in the dog control region [21]. The 333 base substitutions within protein-coding genes include 270 synonymous and 63 missense variants; analysis with VEP indicates that none of these missense mutations is likely to have an impact on protein function that is more than moderate (data not shown). Figure 2 shows the locations of different classes of variants.



**Fig. 2** Observed mtDNA sequence variants within the dataset. The 438 variant sites (435 base substitutions and three indels) in mitogenome sequence data from 119 domestic cats, plus the variants in the NC\_001700 reference sequence [10] are indicated as filled circles in grey (within non-protein-coding regions), green (synonymous variants in protein-coding genes) or red (missense variants). Variants within the RS2 and RS3 repeat regions are not considered here

### Cat mitogenome phylogeny

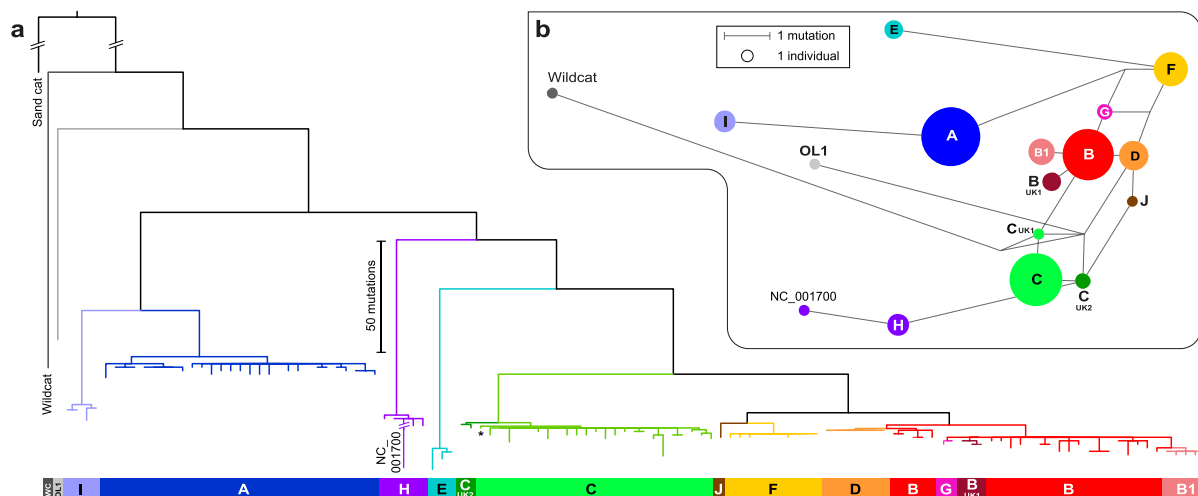
The 119 domestic cat mitogenome sequences, plus the reference sequence NC\_001700 and the wildcat sequence generated here, can be represented in a maximum parsimony tree (Fig. 3a; Figure S2), rooted to a sand cat (*F. margarita*) outgroup (a median-joining network is shown in Figure S3). The tree also includes the domestic cat reference sequence: as discussed above, there is good reason to believe that this sequence contains errors, which are likely responsible for its lying on a very elongated terminal branch (Figure S2). Sets of related haplotypes form clades within the tree, which we refer to as ‘haplogroups’ (hg): these are named according to the haplotypes (‘mito-types’) previously defined on the basis of CR sequence variation [16]. The single ‘OL1’ mitogenome, defined as a rare ‘outlier’ based on CR analysis [16, 22], lies basally in the tree and is more closely related to the wildcat sequence than to domestic cat sequences. This seems consistent with an origin via introgression from a wildcat population. All haplogroups are monophyletic, and CR-based definitions are highly congruent with the phylogenetic structure of the mitogenome-based tree.

With the exception of hg B1 our original study of 152 UK cats [22] did not detect any of the named sub-haplogroups or unique haplotypes which together comprised a quarter of the 1394 previously published in a CR database ([16]; Figure S4). These rare 402-bp CR variants all appear to be derived from the ten globally distributed clades A-J sequenced in our study, including the rare and

geographically restricted haplogroups K and L which fall within a cluster of named A sub-haplogroups [16]. While B-UK1, C-UK1 and C-UK2 represent derived sub-haplogroups so far only reported from the UK, the rare outlier haplogroups OL2 and OL3 were not detected in our dataset. CR-defined haplotypes were based on a relatively small number of variants; the addition of many more variants here shows that, while most established clades (A, C, E, H, I, J and F) are phylogenetically reasonable, some are not. For example, hgs B1, B-UK1 and G are completely embedded within the larger terminal portion of hg B and do not deserve special status based on our tree (Fig. 3a), and hg D and the smaller basal portion of hg B form a clear unit.

Among the 119 sequenced mitogenomes there are 87 distinct sequences; haplogroup D is notable in containing seven mitogenomes that show no observable sequence diversity. While 85% of our sampled cats of known breed are domestic short- or long-hair cats, three of the hg D individuals belong to the Burmese or the related Tonkinese breed, which in the western world are thought to derive from a female imported from Burma in 1930 [23]. This could therefore reflect a recent founder effect; no other breed-haplogroup relationships are evident in the dataset.

The deep structure of the cat mitogenome phylogeny contains no features indicative of ancient expansions (Fig. 3a): internal branches are long, involving no polytomies. The tips, however, present a highly consistent



**Fig. 3** Maximum parsimony phylogeny based on 119 cat mitogenome sequences compared to the CR variant network. **a** Maximum parsimony tree rooted to a sand cat mitogenome outgroup (*Felis margarita*; NC\_028308) with the addition of a wildcat mitogenome and the *F. catus* reference sequence (NC\_001700; [10]). Clades are labelled with the original (A-L, and variants thereof) haplogroup names based on control region variant sites. Note the long (here, truncated) terminal branch leading to the NC\_001700 reference sequence, which reflects likely sequencing errors. Figure S2 shows the same tree including sample names. **b** Median-joining network based on variants within the control region in the 119 cat mitogenomes, plus the wildcat and NC\_001700 reference sequence [10]. Circles represent haplotypes, with area proportional to sample size, and lines between haplotypes represent mutational steps as shown in the key

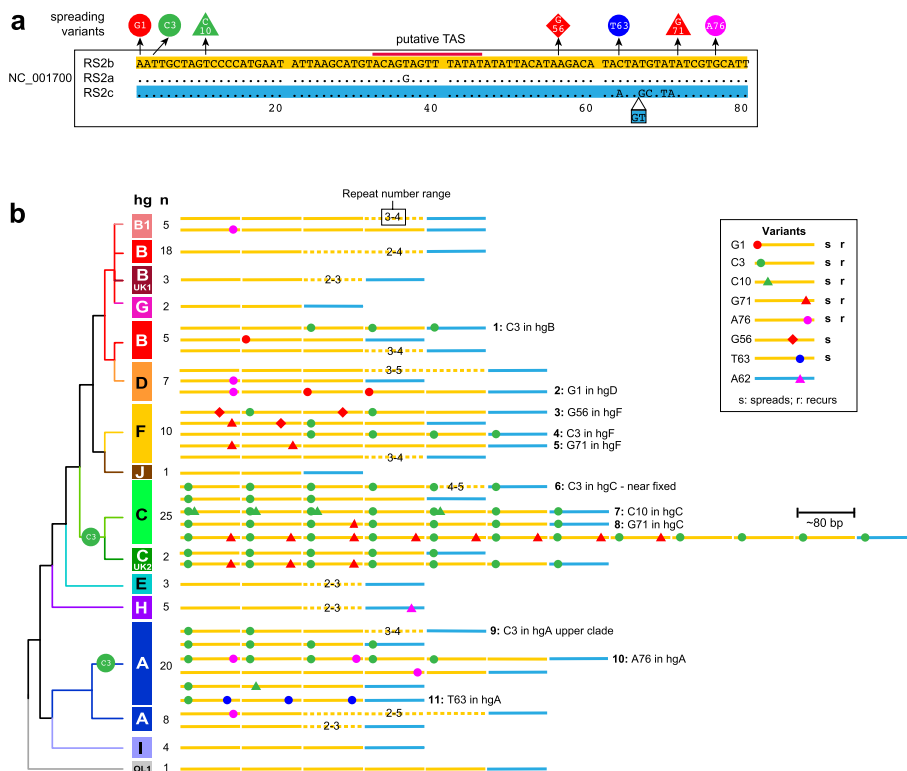
appearance showing star-like patterns of similar mutational depths in all lineages, compatible with domestication-related expansions across the entire tree. To estimate the expansion times of these clades we need a measure of mutation rate. The sand cat/domestic cat split is dated at 1.92 MYA (0.39–4.64 MYA) [24] which, based on the 276 substitutions in the 15,449-bp coding region between sand cat reference sequence and domestic cat (using cat67, closely related to the reference sequence but without its putative errors) implies a mutation rate of 4.65E-09 (1.93E-09–2.29E-08) per base per year. We used the rho statistic [25], i.e. the mean number of mutations to the root, to estimate the average clade expansion date. For the coding region, average rho for the star-like expansion clades A, B/D, C, E, F, H and I is 2.24, which equates to an average clade age of 31.2 KYA (6.3–75.3 KYA). We expect the expansions to post-date the onset of domestication, which, given the archaeological evidence, has been placed between 9.5 and 5.7 KYA [14, 15]. This period lies well towards the lower end of our age estimate range, suggesting that our phylogenetic mutation rate estimate may be artificially low; this could reflect issues

with the fossil record, hybridisation between lineages [24] or domestication-related relaxation of selection on mtDNA, as suggested for domestic dogs [26].

**Variation within the RS2 repeat array**

Within the cat mitochondrial reference sequence [10] the RS2 repeat array contains three copies of an ~80-bp repeat unit (Fig. 4a). Alignment of long-amplicon sequence reads to the reference sequence across this region revealed evidence that many of the sequenced mitogenomes contain different numbers of repeat units (data not shown); this included apparent mixed nucleotides produced through misalignment, similar to the RS3 region (Figure S1).

To investigate this variation, we amplified a smaller segment of mtDNA encompassing RS2 (755 bp based on the reference sequence) and analysed the products by agarose gel electrophoresis in a subset of cats (Figure S5). Each cat presented heterogeneous PCR fragments of different lengths and intensities, with periodicity around 80 bp; given the length of the repeat unit and the consistency of amplification product



**Fig. 4** Sequence variation in the RS2 repeat region. **a** The three RS2 repeats within the reference sequence (NC\_001700). Seven single-nucleotide variants that spread within RS2 arrays are shown, as is the position of a putative termination associated sequence (TAS; [17]). **b** Simplified phylogeny showing examples of RS2 repeat array structures and the eleven inferred variant spreading events (labelled 1–11). The full set of 118 analysed arrays can be seen in Figure S6 and Table S6. Within schematic repeat arrays, numbers over dotted lines indicate repeat number ranges for a given repeat type, e.g. ‘3–4’ indicates a range of 3–4 repeat units within that set of structures

lengths between replicate PCRs (data not shown), it seems unlikely that PCR slippage is responsible for the observed heterogeneity and it more probably represents heteroplasmic variation in copy number of the RS2 repeat unit among mitogenome copies. The length of the predominant PCR fragment varied among cats, suggesting underlying repeat unit copy number variation in RS2.

By filtering sequence reads on the basis of binned lengths separated by ~80 bp, it was possible to capture pools of reads corresponding to PCR products of different lengths and RS2 repeat copy numbers (Figure S5). This allowed the repeat copy number of the major molecular species in each cat to be identified and the sequence determined using a reference-free approach (Supplementary methods). In this way we could analyse both copy number and sequence of RS2 repeat regions in 118/119 of the Nanopore-sequenced cat mitogenomes and place this repeat array variation in the context of the phylogeny (Figs. 4 and S5; Table S6).

Within the cat reference mtDNA sequence, the first two repeats (80 bp; designated RS2a and b [10]) differ by one base substitution, while the third (82 bp; RS2c) is somewhat divergent, containing five base substitutions and a 2-bp insertion in its last 20 bp (Fig. 4a). The 118 sequenced RS2 arrays follow the same general pattern (Fig. 4b, S5; Table S6), with a single RS2c-like repeat preceded by a set of between 2 and 11 repeat unit copies of RS2b-like repeats; overall median repeat number is 4. The RS2b repeat is found in all haplogroups of the phylogeny including the deepest rooting branch, OL1, and seems likely to represent the ancestral repeat type. However, among RS2b-like repeats across the phylogeny there are 30 variant repeat sequences that differ largely by single-nucleotide variants. Nine such variants are singletons, but the remainder are shared between haplotypes in a way that shows clear evidence of multiple independent recurrence of variants, and the dynamic diffusion of variants along repeat arrays within haplogroups. Examples of such variant spreading events are highlighted in Fig. 4b, and the full set of observed RS2 structures is illustrated in Figure S6 and described in Table S6. A maximum parsimony approach indicates that there have been at least 11 different variant spreading events within the phylogeny.

Adding unambiguous RS2 SNVs to the sequence variation observed in the rest of the mitogenome (Table S5) increases the number of different haplotypes from 87 (Figs. 3; S2) to 103. When repeat array length is also considered, this is further increased to 110 different haplotypes among the 118 studied cats.

## Discussion

Cats are widely believed to have become domesticated in the Near East through association with the development of agriculture (since 12 KYA) and the concomitant increase in rodent pests linked to farming and food storage. Dates from archaeological studies are consistent with this, the earliest claimed evidence dating to 9.5 KYA [14]. Subsequent human migration and trade are likely to have spread domesticated cats widely and rapidly, and this is reflected in a lack of feline population structure as detected by analysis of multiple short-tandem repeat and single-nucleotide polymorphisms in a large sample of mostly Eurasian cats [27]. Likewise, mtDNA CR haplotypes are widely shared among populations, albeit at different frequencies [22] and a few common haplotypes predominate.

Here we have used Nanopore and Illumina sequencing of long PCR amplicons to analyse the mitogenomes of 119 domestic cats, producing a highly resolved SNP-based phylogeny based on 16,438 bp of sequence. The analysed cats are all from the UK and therefore cannot be claimed to be representative of the global population; nonetheless, they include all major clades as observed in previous studies [16, 28, 29] of more geographically widespread samples, reflecting the rapid expansion and spread of cats. The deep structure of the phylogeny (Fig. 3a) contains no polytomies and no evidence of ancient expansions of domestic cat matrilineages. One basal branch (OL1) represents a possible wildcat introgression. The phylogeny contains five haplogroups (A, C, E, H and I) that are separated by deep splits; haplogroups in the remaining well-separated superhaplogroup (encompassing B, D, G, F and J) are more closely related. Within clades, we observe star-like topologies of similar depths, which suggests domestication-related expansions across the tree. However, dating based on the median sand cat/domestic cat divergence time produces unexpectedly ancient expansion ages, suggesting possible issues with the fossil dates, discordance due to ancient hybridisation (widespread among the Felidae [24]) or recent increases in effective mutation rates via relaxed selection [26]. One notable feature in the phylogeny is the extended branch leading to the reference sequence (NC\_001700); we interpret this as evidence of errors in the original sequence, and suggest use of an amended reference as an alternative, based on the cat67 mitogenome, which is the most closely related sequence to NC\_001700 in our phylogeny (Figure S2).

Vertebrate mitochondrial DNA is often portrayed as a paragon of economy, with its conserved and minimal gene set, and its lack of wasteful introns or intergenic sequences. In this context, the presence in some species of repeat arrays at five different positions in the CR [9]

is surprising. Diversity in such repeat regions has been understudied because of the problems of heteroplasmy, and past Sanger-based analysis has often relied on cloning [9] which in turn may have introduced artefacts. Current sequencing approaches such as Nanopore technology offer the opportunity to use bioinformatic filtering to examine individual populations of reads [30], and thereby throw light on complex structural variation. The cat RS3 array clearly contains variation in repeat sequence that appears to be phylogenetically structured (Figure S1) as well as variation in copy number, but remains challenging to analyse; here, we have focused our attention on RS2.

Although RS2 variation has previously been studied between related species (e.g. big cats [17] and ruminants [31]) or within a sample of a species (e.g. Japanese sika deer [32]), our study of the structures of the cat RS2 repeat array presents the first (to our knowledge) picture of the diversity of a mitochondrial minisatellite-like sequence within the framework of a detailed intraspecific SNV-based phylogeny. This allows events that are identical by descent (ancestral) or identical by state (recurrent) to be distinguished, and provides evidence of a remarkably dynamic process of recurrent mutation and variant spread across arrays. The blood DNA samples studied here show evidence of a high degree of RS2 heteroplasmy in the form of multiple different length variants in individual cats, as observed in PCR products and length profiles of sequence reads (Figure S5). We have chosen to focus on the apparent majority length class in each cat in defining a representative RS2 array structure. However, this is not entirely objective and majority length could vary between tissues.

What can be deduced about underlying RS2 mutation processes? In early descriptions of mitochondrial repeat diversity (focused on RS3; [9]), repeat secondary structures in single-stranded DNA were invoked as drivers of intrahelical misalignment and replication slippage. This is also implied in the original description of cat RS2 [10], which illustrated a secondary structure involving two repeats. However, the structures presented were for the folding of RNA, rather than DNA; in terms of free energy, RNA secondary structures are much more stable than those of equivalent DNA sequences [33, 34] because of the additional hydrogen bonds permitted by 'wobble' base pairing and by the presence of the 2'-hydroxyl group of ribose. Reanalysis of RS2 using DNA-folding [35], by contrast, produces structures that show very weak intramolecular interactions (Figure S7) which seem unconvincing candidates for secondary structure formation. The RS2 repeat unit contains [10, 17] a putative termination associated sequence (TAS), which has been linked to the termination of replication of the 7S DNA molecule

(Fig. 1; [5]) and in human cells is reversibly bound by a helicase, TWINKLE [36]. The TAS was first identified in human-mouse comparisons [37] and lies some distance rightwards (in our orientation; Fig. 1) of the 3' end of the 7S DNA, as assessed from direct analyses of this molecule [37, 38]. It therefore seems likely that much of the heavy strand of the control region, including some RS2 repeat units, is single-stranded while the D-loop structure exists. If this were important in the generation and spread of new repeat variants in the RS2 array, it might be expected to lend some polarity to repeat array variation; however, the structures shown in Figs. 4 and S5 do not display this.

The long-amplicon based mitogenome sequencing undertaken here greatly increases observed sequence diversity compared to the standard approach of CR analysis. The 16,438 bp sequenced (excluding RS2 and RS3) in 119 cats yields 87 distinct sequences, a marked increase from the 15 observed CR haplotypes. This represents a reduction of random match probability (RMP) from 0.16 to 0.018, and therefore it illustrates that mitogenome sequencing will have forensic utility. Since most forensic casework involves cat hairs [22, 39], in which DNA is likely to be degraded and long-amplicon approaches unsuccessful, it is necessary to develop suitable short-amplicon multiplexes (this we describe elsewhere [40]). If it were possible to add RS2 sequence diversity to assessments of forensic evidence, this would further decrease RMP, as has been suggested for RS3 [18]. However, the high degree of observed heteroplasmy is likely to present problems in the rigorous definition of matches or exclusions.

## Materials and methods

### DNA samples

We analysed samples from 119 domestic cats. These included 115 from a set of 152 individuals previously sampled [22] from 105 veterinary surgeries throughout England; the remaining four were from the Bristol Veterinary School. All blood samples were taken by veterinarians as part of a routine clinical examination and anonymised apart from breed information. Samples from a wildcat (*Felis silvestris*) and a sand cat (*F. margarita*) were respectively from Twycross Zoo and the laboratory collection at the University of Leicester. Table S1 lists all studied individuals, their sources and the analyses undertaken. In sub-sampling the 115 cats we used CR-based mitotype information from a previous Sanger sequencing study [22], including less common mitotypes as well as multiple individuals of the same mitotype. A population differentiation test [41] conducted between the full set of 152 and the 119 studied here indicated that the haplotype frequencies did not differ significantly ( $p=0.98$ ); hence,



the dataset presented here is representative. DNA was extracted from 200 µl of blood using the QIAamp DNA Mini Kit (Qiagen), and quantified using the NanoDrop 2000 (Thermo Scientific).

### Long-PCR and sequencing of mtDNA

The mitogenomes of 93 cats (of which 89 were from a previous study [22]), were each amplified in two overlapping segments of 8670 and 9789 bp using primer pairs 26F (5'-AATCGTCACTGCCCATGC-3') and 52R (5'-TCGAATGTTGGTCATTAAGTT-3'), and 50F (5'-AATGAGCCTAACTATGAGCCA-3') and 27R (5'-CCATGTAGCCAAAGGGTTC-3') respectively (Fig. 1a). Primer design was based on the domestic cat mitochondrial reference (NC\_001700 [10]). Primers were custom barcoded with a 13-bp 5' sequence and unique barcode combinations were used (Table S2). Barcodes were designed [40] to differ from each other even when accounting for Nanopore sequencing errors.

Amplification was carried out in a 10 µl volume containing 20 ng of template DNA, 0.3 µM of each primer, 0.9 µl of '11.1×buffer' [42] and 0.06 µl of a 20:1 mix of PCR BIO Taq Polymerase (PCRBiosystems; 5U/µl); Pfu DNA polymerase (Agilent Technologies; 2.5 U/µl). PCR cycling conditions for the amplification of cat mitogenome segments generally consisted of an initial denaturation at 96 °C for 1 min, followed by 28 cycles of denaturation at 96 °C for 20 s, annealing at 60 °C for 30 s, extension at 68 °C for 9 min for the 8.6-kb segment or 10 min for the 9.7-kb segment, followed by a final extension at 68 °C for 10 min.

PCR products were quantified by agarose gel electrophoresis. Equimolar concentrations of the two overlapping amplicons from up to eight cats, each amplified with primers bearing unique custom barcode combinations, were pooled prior to secondary barcoding at the library preparation stage. Each pool was then purified with a 1.0×volume of AppMag PCR Clean Up Beads (Appleton Woods Ltd.) and purified products quantified using the dsDNA HS Assay kit with the Qubit fluorometer 2.0. A Nanopore DNA library was prepared according to the ligation sequencing amplicons—native barcoding amplicon protocol (ONT version: NBA\_9093\_v109\_revF\_12Nov2019) using the native barcoding expansion 1–12 (EXP-NBD104) and ligation sequencing kit (SQK-LSK109). Sequencing was carried out using a MinION flow cell (FLO-MIN106, R9.4.1) with a Mk1b MinION sequencer and the software MinKNOW (v20.10.3) without the basecalling option.

Mitogenome fragments from a wildcat and sand cat were amplified (without custom barcodes), quantified and purified as above. A DNA library was prepared following the native barcoding genomic DNA

protocol for Flongle (ONT version: NBE\_9065\_v109\_revY\_14Aug2019) although the the NEBNext FFPE DNA repair mix and buffer were excluded during the DNA repair and End-prep stage. Sequencing was undertaken using a Flongle flow cell (FLO-FLG001) with the MinKNOW software (v20.10.3).

A total of 35 cat mitogenomes were sequenced using Illumina technology. Amplification used the same PCR reaction mixture and cycling conditions as above, except that 35 cycles were used rather than 28. For 34 samples, sequencing was carried out in-house using the Nextera XT DNA Library Preparation Kit PE 150-bp Illumina MiSeq [MiSeq Reagent Cartridge v2]. One sample (cat67; HgH) was sequenced by Novogene, Cambridge, to a read-depth of 60,337× on a NovaSeq 6000 SP platform. Nine cat mitogenomes were sequenced using both Nanopore and Illumina sequencing technologies to test concordance.

### Processing of sequence data

In processing Nanopore data, Guppy (v5.0.16) was used for basecalling using the high accuracy model with Qscore filtering disabled and for demultiplexing of Nanopore library native barcodes. Demultiplexing was performed based on the presence of the native barcode on both ends of a read for MinION flow cell runs, and on the presence of one barcode only for Flongle sequencing data (unless stated otherwise). Read quality statistics were assessed using NanoPlot (v1.33.1) [43] and reads filtered using NanoFilt (v2.8.0) [43]. Demultiplexing of reads based on our PCR custom barcodes was carried out before filtering. MiniBar (v 0.21) [44] was used to search for barcode and primer sequences within the first and last 150 bp of any given read, with these sequences then being trimmed. Up to two nucleotide differences were allowed in the barcode sequences and up to 11 in the primer sequences. NanoFilt (v2.8.0) [43] was used to remove reads with a quality score of <11, and a length of <6000 or >16,000 bp. Two hundred reads were subsampled for each amplicon for the flow cell sequencing data. Two domestic cats yielded <200 reads for the 9.7-kb amplicon, and therefore all reads were retained.

The two overlapping amplicons were merged and mapped to the domestic cat reference mtDNA sequence (NC\_001700) using minimap2 (v2.17) [45] (for the domestic cat, *Felis catus*; NC\_028310 for the wildcat, *F. silvestris*; and NC\_028308 for the sand cat, *F. margarita*). SAMtools (v1.13) [46] was used to process the mapping files and variants called using freebayes (v0.9.21) [47] with a ploidy of 1. SNPs were then filtered using a quality (QUAL) cut-off of 200. A consensus sequence was generated using BCFtools [48] and the resulting sequences were combined with the Illumina-generated

cat mitogenome sequences and aligned using MUSCLE [49]. Illumina sequence data generated on the MiSeq platform were processed as previously described [50] and variants were called using BCFtools mpileup with the  $-B$  flag and VarScan (v.2.3.9) mpileup2cns [51].

### Data analysis

Sequence data were visualised using IGV, the Integrative Genomics Viewer [52]. shinyCircos [53] was used to draw circular representations of mtDNA features and summary data. The effects of variants within coding regions were assessed using the Ensembl Variant Effect Predictor (v105) [54]. Maximum parsimony trees were constructed using PHYLIP v3.698 [55]. Three trees were created using DNAPARS, where the input order of sequences was randomised with a different seed number each time and jumbled ten times. A consensus tree was created using the Consense package and this was used to obtain the final tree in DNAPARS. Median-joining networks [56] were constructed using Network 10 (fluxus-engineering.com) and represented using Network Publisher. Population differentiation testing was carried out within Arlequin ver 3.5.2.2 [41]. Dating of clades in the mtDNA phylogeny was done in Network using the mean mutational distance to the relevant root haplotype ( $\rho$ ; [25]). We considered only the coding region and estimated mutation rate based on a mean sand cat/domestic cat divergence time of 1.92 MYA (range: 0.39–4.64 MYA [24]), assuming a constant evolutionary rate. The estimated mutation rate was 4.65E-09 substitutions per bp per year (range: 1.92E-09–2.29E-08). Secondary structure of RS2 repeats was assessed using mfold [35].

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09789-1>.

**Additional file 1: Table S1.** Details of all individuals analysed in this study, CR-predicted haplogroups, and sequencing methods. **Table S2.** Barcode combinations used for Nanopore sequencing. **Table S3.** Indels observed in Nanopore data from 93 domestic cat mitogenomes. **Table S4.** Variants specific to the domestic cat mitogenome reference sequence. **Table S5.** Variants list showing phylogenetic status. **Table S6.** Sequence variation in the RS2 repeat array.

**Additional file 2: Figure S1.** Complex variation within the RS3 repeat array. **Figure S2.** Maximum parsimony tree based on all mitogenome single-nucleotide variants, including sample names. **Figure S3.** Median joining networks based on cat mitogenome sequences. **Figure S4.** Maximum Likelihood tree based on the 402-bp CR variants reported by Grahn et al. [16]. **Figure S5a.** Correspondence of PCR products and sequence read-depth across the RS2 repeat region in a selection of cats. **Figure S5b.** Uncropped gel photograph on which Figure S5a is based. **Figure S6.** RS2 repeat variation across domestic cat phylogeny. **Figure S7.** Folding of RS2 repeats based on RNA and DNA sequences.

**Additional file 3.** Supplementary Methods: Investigation of RS2 region sequence variation.

### Acknowledgements

We gratefully acknowledge colleagues who contributed DNA samples, and NUCLEUS Genomic Services at the University of Leicester for access to Illumina MiSeq sequencing. This research used the SPECTRE High Performance Computing Facility at the University of Leicester for data analysis.

### Authors' contributions

JHW, MAJ conceived and supervised the study; FS, APF, JHW provided resources. ECP, GML, BO, and RN carried out experiments; ECP, CB, JHW carried out data analysis; ECP, JHW, MAJ designed methodology, visualized data and drafted the manuscript. All authors have read, edited, and approved the final manuscript.

### Funding

EP was supported by a BBSRC-MIBTP (grant no. BB/M01116X/1) iCASE studentship, co-sponsored by Twycross Zoo (East Midland Zoological Society) and Zoological Society of London. We thank Jane and Andrew Elliot for additional financial support.

### Availability of data and materials

Sequence data described in this paper are available from GenBank under the accession listed in Table S1. Sequences are also available as FASTA files and an alignment at [https://figshare.com/articles/dataset/Domestic\\_cat\\_mtDNA\\_sequences/23266712](https://figshare.com/articles/dataset/Domestic_cat_mtDNA_sequences/23266712).

### Declarations

#### Ethics approval and consent to participate

This research was approved by the University of Leicester's Animal Welfare and Ethical Review Body (ref. AWERB/2021/159). All methods were performed in accordance with the relevant guidelines and regulations, and the study is reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>). Consent to participate is not relevant.

#### Consent for publication

Not applicable.

#### Competing interests

BO is currently an Oxford Nanopore Technologies employee but was not at the time of data generation for this study. All other authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Genetics & Genome Biology, University of Leicester, University Road, Leicester LE1 7RH, UK. <sup>2</sup>Present Address: Oxford Nanopore Technologies Plc., Oxford Science Park, Edmund Halley Rd, Oxford OX4 4DQ, UK.

<sup>3</sup>Present Address: Department of Population Health Sciences, University of Leicester, Leicester, UK. <sup>4</sup>Biomedical Research Centre, Leicester National Institute for Health and Care Research, Glenfield Hospital, Leicester, UK. <sup>5</sup>IDEXX Laboratories Italia S.R.L., Via Guglielmo Silva, 36-20149 Milano, MI, Italy. <sup>6</sup>Bristol Veterinary School, University of Bristol, Langford House, Langford BS40 5DU, North Somerset, UK.

Received: 23 June 2023 Accepted: 6 November 2023

Published online: 17 November 2023

### References

- Harrison RG. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends Ecol Evol.* 1989;4(1):6–11.
- Bruford MW, Bradley DG, Luikart G. DNA markers reveal the complexity of livestock domestication. *Nat Rev Genet.* 2003;4(11):900–10.
- Kanthaswamy S. Review: domestic animal forensic genetics - biological evidence, genetic markers, analytical approaches and challenges. *Anim Genet.* 2015;46(5):473–84.
- Linacre A. Animal Forensic Genetics. *Genes (Basel).* 2021;12(4):515.
- Nicholls TJ, Minczuk M. In D-loop: 40 years of mitochondrial 7S DNA. *Exp Gerontol.* 2014;56:175–81.

6. Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisa E. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J Mol Biol.* 1986;192(3):503–11.
7. Saccone C, Pesole G, Sbisa E. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J Mol Evol.* 1991;33(1):83–91.
8. Fumagalli L, Taberlet P, Favre L, Hausser J. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Mol Biol Evol.* 1996;13(1):31–46.
9. Hoelzel AR, Lopez JV, Dover GA, O'Brien SJ. Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. *J Mol Evol.* 1994;39(2):191–9.
10. Lopez JV, Cevario S, O'Brien SJ. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. *Genomics.* 1996;33(2):229–46.
11. Parson W, Huber G, Moreno L, Madel MB, Brandhagen MD, Nagl S, Xavier C, Eduardoff M, Callaghan TC, Irwin JA. Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. *Forensic Sci Int Genet.* 2015;15:8–15.
12. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J Mol Evol.* 1994;39(2):174–90.
13. Calabrese FM, Balacco DL, Preste R, Diroma MA, Forino R, Ventura M, Attimonelli M. NumtS colonization in mammalian genomes. *Sci Rep.* 2017;7(1):16357.
14. Vigne JD, Guilaine J, Debue K, Haye L, Gerard P. Early taming of the cat in Cyprus. *Science.* 2004;304(5668):259.
15. Van Neer W, Linseele V, Friedman R, De Cupere B. More evidence for cat taming at the Predynastic elite cemetery of Hierakonpolis (Upper Egypt). *J Archaeol Sci.* 2014;45:103–11.
16. Grahm RA, Kurushima JD, Billings NC, Grahm JC, Halverson JL, Hammer E, Ho CK, Kun TJ, Levy JK, Lipinski MJ, et al. Feline non-repetitive mitochondrial DNA control region database for forensic evidence. *Forensic Sci Int Genet.* 2011;5(1):33–42.
17. Jae-Heup K, Eizirik E, O'Brien SJ, Johnson WE. Structure and patterns of sequence variation in the mitochondrial DNA control region of the great cats. *Mitochondrion.* 2001;1(3):279–92.
18. Fridez F, Rochat S, Coquoz R. Individual identification of cats and dogs using mitochondrial DNA tandem repeats? *Sci Justice.* 1999;39(3):167–71.
19. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE.* 2021;16(10):e0257521.
20. Verscheure S, Bacheljau T, Desmyter S. Dog mitochondrial genome sequencing to enhance dog mtDNA discrimination power in forensic casework. *Forensic Sci Int Genet.* 2014;12:60–8.
21. Duleba A, Skonieczna K, Bogdanowicz W, Malyarchuk B, Grzybowski T. Complete mitochondrial genome database and standardized classification system for *Canis lupus familiaris*. *Forensic Sci Int Genet.* 2015;19:123–9.
22. Ottolini B, Lall GM, Sacchini F, Jobling MA, Wetton JH. Application of a mitochondrial DNA control region frequency database for UK domestic cats. *Forensic Sci Int Genet.* 2017;27:149–55.
23. The Tonkinese; <https://cfa.org/tonkinese/>.
24. Li G, Davis BW, Eizirik E, Murphy WJ. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* 2016;26(1):1–11.
25. Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet.* 2000;67:718–26.
26. Bjørnerfeldt S, Webster MT, Vila C. Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.* 2006;16(8):990–4.
27. Nilson SM, Gandolfi B, Grahm RA, Kurushima JD, Lipinski MJ, Randi E, Waly NE, Driscoll C, Murua Escobar H, Schuster RK, et al. Genetics of randomly bred cats support the cradle of cat domestication being in the Near East. *Heredity (Edinb).* 2022;129(6):346–55.
28. Wesselink M, Bergwerff L, Hoogmoed D, Kloosterman AD, Kuiper I. Forensic utility of the feline mitochondrial control region - A Dutch perspective. *Forensic Sci Int Genet.* 2015;17:25–32.
29. Arcieri M, Agostinelli G, Gray Z, Spadaro A, Lyons LA, Webb KM. Establishing a database of Canadian feline mitotypes for forensic use. *Forensic Sci Int Genet.* 2016;22:169–74.
30. Marshall C, Parson W. Interpreting NUMTs in forensic genetics: Seeing the forest for the trees. *Forensic Sci Int Genet.* 2021;53:102497.
31. Hassanin A, Ropiquet A, Couloux A, Cruaud C. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). *J Mol Evol.* 2009;68(4):293–310.
32. Ba H, Wu L, Liu Z, Li C. An examination of the origin and evolution of additional tandem repeats in the mitochondrial DNA control region of Japanese sika deer (*Cervus nippon*). *Mitochondrial DNA A DNA Mapp Seq Anal.* 2016;27(1):276–81.
33. Bercy M, Bockelmann U. Hairpins under tension: RNA versus DNA. *Nucleic Acids Res.* 2015;43(20):9928–36.
34. Swadling JB, Ishii K, Tahara T, Kitao A. Origins of biological function in DNA and RNA hairpin loop motifs from replica exchange molecular dynamics simulation. *Phys Chem Chem Phys.* 2018;20(5):2990–3001.
35. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003;31(13):3406–15.
36. Jemt E, Persson O, Shi Y, Mehmedovic M, Uhler JP, Davila Lopez M, Freyer C, Gustafsson CM, Samuelsson T, Falkenberg M. Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Res.* 2015;43(19):9262–75.
37. Doda JN, Wright CT, Clayton DA. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci U S A.* 1981;78(10):6116–20.
38. Foran DR, Hixson JE, Brown WM. Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Res.* 1988;16(13):5841–61.
39. Lyons LA, Grahm RA, Kun TJ, Netzel LR, Wictum EE, Halverson JL. Acceptance of domestic cat mitochondrial DNA in a criminal proceeding. *Forensic Sci Int Genet.* 2014;13:61–7.
40. Patterson EC, Matharu Lall G, Neumann R, Ottolini B, Sacchini F, Foster AP, Jobling MA, Wetton JH. Defining cat mitogenome variation and accounting for numts via multiplex amplification and Nanopore sequencing. *Forensic Sci Int Genet.* 2023;102944. <https://doi.org/10.1016/j.fsigen.2023.102944>.
41. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10(3):564–7.
42. Jeffreys AJ, Neumann R, Wilson V. Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell.* 1990;60(3):473–85.
43. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34(15):2666–9.
44. Krehenwinkel H, Pomerantz A, Henderson JB, Kennedy SR, Lim JY, Swamy V, Shoobridge JD, Graham N, Patel NH, Gillespie RG, et al. Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *Gigascience.* 2019;8(5):giz006.
45. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100.
46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
47. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv 2012, 1207.3907*.
48. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156–8.
49. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
50. Batini C, Hallast P, Vagene AJ, Zadik D, Eriksen HA, Pamjav H, Sajantila A, Wetton JH, Jobling MA. Population resequencing of European mitochondrial genomes highlights sex-bias in Bronze Age demographic expansions. *Sci Rep.* 2017;7(1):12086.
51. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. VarScan: variant detection in massively

parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009;25(17):2283–5.

52. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24–6.
53. Yu Y, Ouyang Y, Yao W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics*. 2018;34(7):1229–31.
54. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122.
55. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Seattle, WA: Department of Genome Sciences, University of Washington; 2005.
56. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37–48.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

