nature ecology & evolution

SUPPLEMENTARY INFORMATION

VOLUME: 1 | ARTICLE NUMBER: 0139

In the format provided by the authors and unedited.

The palaeogenetics of cat dispersal in the ancient world

Claudio Ottoni[1,2†*], Wim Van Neer[3,4], Bea De Cupere[3], Julien Daligault[2], Silvia Guimaraes[2], Joris Peters[5,6], Nikolai Spassov[7], Mary E. Prendergast[8], Nicole Boivin[9], Arturo Morales-Muñiz[10], Adrian Bălășescu[11], Cornelia Becker[12], Norbert Benecke[13], Adina Boroneant[14], Hijlke Buitenhuis[15], Jwana Chahoud[16,17], Alison Crowther[18], Laura Llorente[10†], Nina Manaseryan[19], Hervé Monchot[20], Vedat Onar[21], Marta Osypińska[22], Olivier Putelat[23], Eréndira M. Quintana Morales[24], Jacqueline Studer[25], Ursula Wierer[26], Ronny Decorte[1], Thierry Grange[2‡*] and Eva-Maria Geigl[2‡*]

Table of contents

Precautions taken to minimize contamination 2

Preparation of archaeological samples 2

DNA extraction and purification 3

Ancient samples 3

Modern samples 3

Selection of SNPs and primer design 3

Mitochondrial DNA 3

Nuclear SNPs in the Taqpep gene 5

Amplification of DNA 5

Pyrosequencing of PCR products 5

Library preparation and Ion Torrent sequencing of amplified products in the Genomic Platform of the IJM. 6

Analysis of sequencing data 7

Authentication of aDNA data 8

Iconographic evidence in Europe 9

References 10

Supplementary datasets 11

Supplementary figures 12

Supplementary tables 16

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Precautions taken to minimize contamination

The genetic analyses were performed in the laboratory of Paleogenomics and Epigenomics of the Institute Jacques Monod (IJM) in Paris and in the Laboratory of Forensic Genetics and Molecular Archaeology in Leuven (Department of Imaging & Pathology, University of Leuven, Belgium). Both laboratories are equipped with dedicated pre-PCR ancient DNA (aDNA) facilities physically separated from post-PCR facilities. The aDNA facility in Paris is under positive air pressure, its equipment and the procedures followed are described in Bennett *et al.,* (2014)[1]. In Leuven, access to the pre-PCR laboratory was restricted to a limited number of people and only after wearing clean overalls, gloves, over-shoes, surgical facemasks, plastic spectacles, and following an irreversible sequence of work steps to avoid contamination. Access to the pre-PCR lab in Leuven was not permitted if PCR products had been handled the same day.

The aDNA facilities were routinely cleaned with bleach and RNAse Away (Molecular BioProducts, San Diego, CA, USA) and every item entering the room was extensively washed with bleach or RNAse Away and when possible UV-irradiated.

Buffer, $MgCl_2$ and BSA used for preparation of amplification reactions were UV-irradiated to minimize the risk of contamination by animal DNA in reagents[2] and either γ-irradiated water or autoclaved nuclease-free water (Promega) were used. In addition, carry-over prevention strategies based on the use of uracil-N-glycosylase (UNG) were adopted[3].

In Paris, extractions were done in a UV-irradiated laminar flow hood and amplification reactions were prepared in a dedicated room of the high containment laboratory in a UV-irradiated workstation (Template-Tamer PCR workstation, Qbiogene, Cambridge, U.K.). In Leuven, extractions were performed in a UV-irradiated workstation whereas preparation of amplification reactions was carried out in a UV-irradiated laminar flow cabinet (Esco, Breukelen, Netherlands).

In both laboratories, when the two-step nested PCR strategy was adopted (see below), strict temporal and spatial separation of the various steps was adopted to minimize contamination[4]. Preparation of second-step nested PCR amplification reactions was carried out in pre-PCR facilities normally used for modern DNA samples, under UV-irradiated workstations or laminar flow cabinets. Multiplex-PCR products of the first amplification step were loaded into the second-step amplification reactions in a laboratory physically separated from pre- and post-PCRs facilities inside a UV-irradiated workstation.

Extraction of modern DNA samples was performed in Leuven by a different operator and carried out in pre-PCR laboratories physically separated from the aDNA facilities.

Preparation of archaeological samples

To extract DNA from teeth and bones, one sample was prepared at a time. Samples were subjected to the following decontamination procedures. The outer surface of bone and teeth samples was removed through sterile blades or, in Leuven, with a Dremel drill (Dremel, Racine, WI, USA).

Additionally, the surfaces of the teeth were gently wiped with 10% bleach and rinsed with bidistilled water. Bone and teeth samples were subsequently ground into a fine powder with a Dremel drill, a mortar or in a 6750 Freezer Mill (SPEX CertiPrep, Metuchen, NJ, USA) and stored at 4°C until DNA extraction. Grinding vials were washed and decontaminated using RNAse Away (Molecular BioProducts, San Diego, CA, USA) and subsequent UV-irradiation (254 nm in cross-liker). In some instances, residues of wrapping tissue were present in skin and hair sampled in Egyptian mummies and they were removed with sterile blades and tweezers.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

DNA extraction and purification

Ancient samples

DNA extractions and purifications were performed on aliquots of 100-300 mg of bone or tooth powder, hair tuft (16-102 mg) and skin (up to 250 mg). Bone or tooth powder was incubated for 24-48h at 37°C on a rotating wheel in 1.8 mL digestion buffer containing 0.5 M EDTA (Sigma Aldrich), 0.25 M Na_2HPO4^{3-} (Sigma Aldrich) and 1% 2-mercaptoethanol (Sigma Aldrich) at pH 8, or in 1.8 mL digestion solution of 0.5 M EDTA pH 8 (Invitrogen, Carlsbad, CA, USA) and 0.25 mg/mL proteinase K (Roche, Penzberg, Germany). Skin and hair tuft samples of Egyptian mummies were incubated for 24h at 56°C on a rotating wheel in 1.8-3.6 mL digestion solution of Tris-HCl 100 mM pH 8, NaCl 100 mM, $CaCl_2$ 3 mM, N-Lauroylsarcosine 2%, DTT 40 mM and 0.4 mg/mL proteinase K[5].

After pelleting, DNA was purified following a protocol based on the QIAquick Gel Extraction kit (Qiagen, Hilden, Germany), including additional washing steps with 2 mL QG binding buffer and 2 mL PE wash buffer (Qiagen)[1]. Total volumes of 8–36 mL of extract in binding buffer were passed through the silica columns on a vacuum manifold (Qiagen) using 15-25 mL tube extenders (Qiagen). Final DNA elution was done in two steps, each using 27 µL EB elution buffer (Qiagen) heated to 65°C. Each independent extraction batch contained one blank control for every five archaeological samples.

All ancient samples were analyzed either by pyrosequencing after nested PCR with biotinylated primers or with Ion Torrent sequencing (see below).

Modern samples

DNA extractions and purifications of skin and claw samples from modern wildcats were done in the Laboratory of Forensic Genetics and Molecular Archaeology in Leuven in laboratories that are physically separated from the aDNA facilities and in workflows temporally separated from the aDNA analyses. Samples were directly incubated for 1 hour at 56 °C in 100 µL of incubation buffer (Tissue and Hair Extraction kit, Promega), 0.1 M DTT (Invitrogen) and 2 mg/mL proteinase K (Roche, Penzberg, Germany). After removing the claw or the piece of skin, 300 µL of lysis buffer (DNA IQ Casework Pro Kit for Maxwell 16, Promega) and 3 µL of DTT 1M were added to the solution and purified in a Maxwell 16 (Promega) with a DNA IQ Casework Pro Kit.

All modern samples were analyzed in Leuven by pyrosequencing after nested PCR amplification with biotinylated primers (see below).

Selection of SNPs and primer design

Mitochondrial DNA

For primer design, we relied on the most up-to-date worldwide cat phylogeny described by the 2007 study of Driscoll *et al*.[6], in which 2604 bp of the ND5, ND6 and CytB genes of the mtDNA were sequenced revealing five different clades (I to V) that correspond to five *Felis silvestris* subspecies (Supplementary Fig. S1a). The haplotypes from Driscoll's study were downloaded from GenBank, aligned in Geneious 6 (http://www.geneious.com[7]), against a reference sequence[8], and a maximum-likelihood tree was constructed using PhyML[9] in Geneious. With the aim to design a molecular assay that was suitable for the analysis of ancient degraded samples and at the same time preserved a fairly good phylogenetic resolution at the level of the major clades and subclades, we selected 42 informative SNPs distributed across nine short regions of the ND5, ND6 and CytB genes of the mtDNA. Primer design and multiplex testing were performed following the strategy of the aMPlex Torrent workflow[10]. We used the software Primer3 within Geneious to design primers which hybridize to conserved sequences flanking short phylogenetically informative sequences containing one or several diagnostic SNPs. For primer optimization, we also used the software Oligo 6 (Molecular Biology Insights, Inc) to select primers with a temperature of dissociation (Td) of 65 to 67°C (nearest neighbor method) and for

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

which there was no predicted 3'-dimer having a ΔG below -1.6 kcal/mol. In this way, we designed various candidate primer pairs targeting nine regions (<130 bp) containing the diagnostic SNPs selected and tested them for efficiency and dimer formation using quantitative real-time PCR (qPCR) with a LightCycler 480 (Roche). qPCR amplification was carried out with 2 µL of cat DNA (extracted from hairs of a modern cat) at five different concentrations (serial dilutions from 12,5 ng to 20 pg), or 2 µL of water for non-template controls (6 replications per primer pair), in a final volume of 7 µL containing SYBR Green I Master Mix 1x and 0.5 µM of each primer in a 384-well plate with an EpMotion 5070 robot (Eppendorf). The cycling program used was as follows: 5 minutes of polymerase activation at 95°C, followed by 60 cycles at 95°C for 10 seconds and 60°C for 40 seconds. A melting curve was established at the end of the run through a melting step (95°C for 5 seconds, then 65°C for 1 minute and a temperature increase of 0.11°C with continuous fluorescence measurement). We selected primer pairs whose efficiency was above 90 % and that yielded no primer-dimers before 40 cycles. The final primers selected amplified fragments of 60 to 111 bp in length (see Supplementary Table S1).

In order to choose the best combination of primers to use in multiplex PCRs, we predicted *in silico* their tendency to form primer-dimers using the PriDimerCheck software (http://biocompute.bmi.ac.cn/MPprimer/primer_dimer.html)[11]. Based on the output of the program we divided the nine primer pairs in two different multiplex reactions to minimize the predicted stability of 3'-dimers. Each multiplex was optimized by testing different $MgCl_2$ concentrations (2 mM, 3 mM and 4 mM) and was carried out with 5 µL of 10 pg/µL cat DNA in a 50 µL reaction volume with a final composition of 1X FastStart High Fidelity Reaction Buffer and $MgCl_2$ 3 mM (Roche), 0.25 mM each of dGTP, dATP, dCTP, 0.5 mM of dUTP (Bioline), BSA 1 mg/mL, 0.3 units of Uracil-N-glycosylase (UNG, ArcticZymes, Tromso), 1.7 units of FastStart Taq polymerase (Roche), 0.15 µM of each primer (Sigma, IDT) and, when appropriate, additional $MgCl_2$ to achieve a final concentration of either 3 or 4 mM. The $MgCl_2$ solution, reaction buffer and BSA were decontaminated by exposing the solutions in UV-pervious tubes (Qubit®, Life Technologies) to UV light for 10 minutes at a short distance as described elsewhere[2]. Negative controls for each multiplex reaction were processed in the same way as the samples throughout the whole experimental procedure. The cycling program consisted of 15 minutes at 37°C (carry-over contamination prevention through digestion by UNG of dUTP-containing amplicons), 95°C for 10 minutes (inactivation of UNG and activation of the Fast Start DNA polymerase), followed by 35 cycles at 95°C for 10 seconds and 60°C for 1 minute and a final extension step at 72°C for 10 minutes.

After the multiplex PCR amplification, qPCR analyses of the amplification products were performed using each primer pair individually, to measure the production of each PCR product and of possible primer-dimers involving the specific tested primer pair. Each simplex qPCR was carried out as described above using 2 µL of the multiplex product diluted to 1/50th. In order to minimize carry-over contamination, the dilution of the PCR product was performed in a dedicated laboratory, physically separated from the high containment laboratory, the modern DNA facility, the post-PCR facility and the genomic platform. In this way, we established that a $MgCl_2$ concentration of 3 mM was optimal for the multiplex amplification. The sensitivity of the two multiplex PCRs was finally tested by amplifying 20 pg of cat DNA, showing amplification of all fragments, and 5 pg of DNA, in which amplification of some fragments was not obtained.

To assess the phylogenetic consistency of our molecular assay, we constructed a Bayesian tree from the haplotype data of Driscoll et al.[6] reduced to the minimum sequence length generated by our assays (i.e. 286 bp through pyrosequencing, see below) using Mr Bayes[12]. The tree preserved the same topology as that previously described in Driscoll *et al*.'s study in all its major clades and subclades with fairly high statistical support in the main nodes (see Supplementary Fig. S1). The same results were observed with the Maximum-Likelihood method as implemented in PHYML[9] adopting various statistical supports for node definition.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Nuclear SNPs in the Taqpep gene

During felid skin development, the loss of function of the *Taqpep* gene, which encodes the protein Tabulin, causes a loss of coat color pattern periodicity, determining the phenotypic variation in the shape of tabby patterns – mackerel (Ta^M/Ta^M, Ta^M/Ta^b) and blotched (Ta^b/Ta^b). Blotched cats were found to carry a nonsense mutation (W841X, G-A transition) in exon 17 of the *Taqpep* gene[13]. A second less frequent Ta^b allele (D228N, G-A transition) was found to cosegregate with the blotched phenotype in a research colony, whereas a variant at codon 139 (T139N, C-A transversion) was associated to an atypical swirled coat pattern, a phenotype that is similar to the spotting pattern of *Felis nigripes* [13].

The three nuclear SNPs associated to a phenotypic variation in the shape of tabby patters – mackerel (Ta^M/Ta^M, Ta^M/Ta^b) and blotched (Ta^b/Ta^b) [13] – in the nuclear gene *Transmembrane Aminopeptidase Q* (*Taqpep*) were analyzed by amplifying in one multiplex PCR three fragments ranging from 64-110 bp in length (see Supplementary Table S1). Primer pairs were designed and tested as described above for the mtDNA.

Amplification of DNA

Amplification of the nine mtDNA fragments was preceded by the elimination of carry-over contamination based on the dUTP/UNG system[3] and performed in two separate mixes in a final volume of 30-50 µL, at conditions tested and optimized as described above – 1X FastStart High Fidelity Reaction Buffer and $MgCl_2$ 3 mM (Roche), 0.25 mM each of dGTP, dATP, dCTP, 0.5 mM of dUTP (Bioline), BSA 1 mg/mL, 0.3 units of Uracil-DNA glycosylase (UNG, ArcticZymes), 1.7 units of FastStart Taq polymerase (Roche), 0.15 µM of each primer (Sigma, IDT) – using 3-6 µL of aDNA extract. The following cycle conditions were used: 37°C for 15 min, 95°C for 10 min, 35 cycles of 95°C for 30 sec, 60°C for 1 min, and a final step of 72°C for 4 min.

In a first phase of the project, positive amplification products were reamplified in nested SYBR Green qPCR and submitted to pyrosequencing assays (see below).

In a second phase of the project, amplicon libraries were constructed and sequenced on the PGM Ion Torrent platform of the IJM following the aMPlex Torrent workflow[10] (see below). In order to obtain an approximately even distribution of final reads we tested various primer concentrations for each multiplex. These tests were done with 3.5-350 pg of cat DNA (from FTA blood samples) and were carried out with the same strategy described above by measuring the production of each PCR product through qPCR analyses with each primer pair used individually. The final optimized primer concentrations are reported in Supplementary Table S1.

The amplification of the three nuclear fragments of the *Taqpep* gene, was done in one separate multiplex PCR with the same conditions as the multiplexe PCRs used for the mtDNA amplification except for a higher concentration of $MgCl_2$ (4 mM), a denaturation time of 30 sec in the amplification step was used, and 40 amplification cycles. All nuclear amplicons were sequenced exclusively with the PGM Ion Torrent platform.

When necessary, simplex PCR reactions of missing or poorly represented mtDNA fragments were performed using the same PCR conditions as in the multiplex PCR and were sequenced either with pyrosequencing or in the PGM Ion Torrent.

Pyrosequencing of PCR products

A first analytical phase of the project relied on pyrosequencing assays of 5'-biotinylated PCR products to screen 42 informative SNPs within the mtDNA fragments targeted. Nested PCR primers, sequencing primers and nucleotide dispensation orders (Supplementary Table S1, S2, S3) were automatically generated with the Assay Design Software v1.0 (Biotage). Pyrosequencing was performed after nested

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

simplex amplification of the nine mtDNA amplicons obtained from two multiplex PCRs. Amplification products were diluted up to 200 times and 2 µL of diluted PCR product were reamplified in nested SYBR Green qPCR simplex reactions using separately ten primer pairs with one 5'-biotinylated primer – one multiplex PCR fragment was covered by two different overlapping primer pairs – in a final volume of 25 µL containing either 1X GoTaq Green Master Mix (Promega) or 1X SYBR Green I Master Mix (Roche) and 0.76 µM of each primer (Supplementary Table S1). Amplifications were performed in a Lightcycler 480 (Roche) in Paris or in a 7500 Real Time PCR System (Applied Biosystems) in Leuven. The following cycle conditions were used: 95°C for 2 min, 30 cycles of 95°C for 15 sec, 60°C for 1 min, and a final step of 72°C for 4 min. Dimers were identified by subsequent melting curve analysis.

Positive PCR products from the SYBR Green qPCR simplex reactions were processed before the pyrosequencing reaction to yield single-stranded DNA fragments by attachment to streptavidin-coated beads, followed by strand separation and washing steps performed in a Vacuum Prep Workstation according to the manufacturer's instructions (Biotage PyroMark Q96 System and PyroMark Q24, Qiagen). The beads with single-stranded DNA fragments were then transferred to a buffer containing the sequencing primer (Sigma, IDT), which is annealed by heating to 80°C for 2 min and cooling for at least 5 min. Pyrosequencing was performed following the manufacturer's instructions in a Biotage PyroMark Q96 at the Institut de Biologie de l'Ecole Normale Supérieure in Paris and in a PyroMark Q24 (Qiagen) at the Laboratory of Forensic Genetics and Molecular Archaeology in Leuven. Results and sequences were analyzed in the PyroMark Q24 software (Qiagen) or the PyroMark Q96 software (Biotage). In three instances (museum specimens 2129, 2130 from Kenya and 17153 from Rwanda), out of phase shifts in the pyrograms of one fragment (i.e., fragment 3) were observed suggesting the presence of an additional variable site. To account for the additional SNP (G/A at np 13537), the dispensation order was changed and in all instances the correct phase was recovered.

Library preparation and Ion Torrent sequencing of amplified products in the Genomic Platform of the IJM.

In a second analytical phase of the project amplicon, libraries were prepared from the PCR products of the three multiplex PCR reactions of the ancient samples (nine fragments in two multiplexes of the mtDNA and three fragments in one multiplex of the nuclear *Taqpep* gene) and sequenced in a PGM Ion Torrent platform using the aMPlex Torrent workflow[10]. Amplifications, library preparations and Ion Torrent sequencing were all performed at the IJM in Paris.

Amplification products from the three multiplexes were pooled for each individual and blank control in a 96-well plate. We constructed up to nine independent libraries. Each time, samples were amplified in batches of 44 duplicates from one extract, plus eight extraction and PCR controls, using all 96 Ion Torrent barcodes. In some instances, samples were amplified in triplicate or repeated in duplicate in separate amplification batches.

The following steps were performed using a Tecan Freedom Evo 100 robot equipped with a 4 channel liquid handling arm using disposable tips, a gripper to move objects, a double thermoblock and an automated solution for vacuum solid phase extraction.

We first performed end-repair with the NEBNext End Repair module (New England Biolabs) using 40 µL of pooled multiplex PCR (containing no more than 5 picomoles of amplicon products) in a 50 µL reaction volume with 0.1 µL of End Repair Enzymes. The reaction was incubated for 30 minutes at 25°C and then purified using the NucleoSpin 96 PCR clean-up kit as recommended by the manufacturer (Macherey-Nagel ref. 740658). DNA was eluted in 50 µL and 20 µL. Sample-specific Ion Torrent barcoded adaptors (1 µL of annealed A+P1, 20 µM) were ligated using the NEB Next ligation module (New England Biolabs) in a 30 µL reaction volume with 1 µL of Quick Ligase, incubated for 30 minutes at 16°C. After the ligation, 60 µL of NT binding buffer (Macherey-Nagel) was added and all samples

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

were pooled in a tube before purification on a silica column (Qiagen or Macherey-Nagel). Ten µL (out of 30 µL) of the pooled barcoded PCR products were then size selected using the Caliper Labchip XT (Perkin Elmer) by setting a size range of 144-197 bp in order to recover all the amplicons and remove adapter-dimers. Final recovery was done in 20 µL of collection buffer. Size-selected products were subjected to nick-repair and amplification in a 24 µL final volume reaction using NEB OneTaq Hot Start and containing Ion Torrent primers A and P1 (0.5 µM). The reaction was incubated for 20 min at 68°C (nick-repair step), then amplified using the following program: initial denaturation at 94°C for 5 min, (94°C for 15 sec, 60°C for 15 sec, 68°C for 40 sec) for 6 cycles, final elongation at 68°C for 5 min. Products were then purified with a Qiaquick PCR purification kit (Qiagen). The size distribution and concentration of the library was assessed on the Agilent 2100 Bioanalyzer. Emulsion PCR and Ion Sphere Particle enrichment were conducted with the Ion OneTouch System (Life Technologies) using the Ion OneTouch 200 Template kit v2 DL according to the manufacturer's protocol. Each of the nine DNA libraries was sequenced independently on the Ion Torrent Personal Genome Machine (PGM) Sequencer using the Ion PGM 200 Sequencing Kit and Ion 314 semiconductor sequencing chips (Life Technologies).

Analysis of sequencing data

Using the software provided in the Torrent suite on the Ion Torrent server, sequencing reads were demultiplexed and mapped with the Torrent Mapping Alignment Program (TMAP) to a reference fasta sequence composed of the 12 fragments (9 mitochondrial, 3 nuclear) separated by spacers of 100 Ns. The quality of the sequencing reaction and of the library analyzed was assessed within the analysis report of the Torrent Browser. Details about the quality of each Ion Torrent run are reported in Supplementary Table S4.

Downstream sequence analysis was performed by first renaming the BAM and BAI files to allocate simple descriptive names using bash commands. The files were then analyzed with a bash script previously developed[10] (Annex 1, SI) dependent on the following programs (and tested with the versions indicated between brackets): featureCounts (Subread package 1.4.6)[14], SAMtools and bcftools (1.2)[15]. BAM files were converted to SAM files with SAMtools. The number of reads per amplicon for each sample was estimated with featureCounts, which uses a GFF file (Annex 3, SI) describing the features of the fasta file (Annex 2, SI) used for mapping. In the gff file we defined the sequences of the PCR products and the primers as "miscellaneous features" to count the proper PCR products as well as the reads that contain either the forward or the reverse primers. This ensures the accurate estimation of the mapped reads while measuring potential primer dimers identified as supernumerary primer-only counts. Finally, SAMtools *mpileup* and *bcftools* were used to generate a consensus sequence of the 12 concatenated fragments obtained for each individually barcoded sample.

We used Geneious (http://www.geneious.com[7]) to import both the consensus fasta and the initial BAM files. Consensus sequences were aligned against a reference sequence (NC001700), with manual editing when necessary. The various outputs were verified by analyzing the initial sequencing reads within the BAM files displayed within Geneious.

By applying the authentication criteria described below, our sequencing strategy made it possible to gather in total 209 mtDNA profiles out of the 352 ancient cats analyzed (Supplementary Dataset S1, S2). We observed that success rates decreased with increasing environmental temperature. Indeed, only 12 out of 68 ancient Egyptian samples analyzed (Predynastic to Greco-Roman, including 52 mummies) and none of the 14 Neolithic and Bronze Age samples from Cyprus, Jordan and Syria yielded DNA sequences.

In 197 specimens, complete mtDNA profiles (i.e. the nine fragments analyzed) were obtained, with total length of sequence reads ranging from 286-299 bp in samples sequenced exclusively by pyrosequencing

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

respectively with the Pyromark Q96 (N=13) and Pyromark Q24 (N=23), up to 449 bp in samples entirely sequenced in the Ion Torrent platform (N=150), whereas the remaining samples were sequenced by a combination of both techniques (N=11) (Supplementary Dataset S2). In 12 samples, incomplete profiles were generated by successfully amplifying two to seven mtDNA fragments. In most instances (11/12 samples) the longest fragments ranging between 101-111 bp failed to amplify, as expected for ancient samples following diagenetic trajectories of biomolecular decay.

A tree of 66 unique 286 bp-long haplotypes (Supplementary Fig. S2) was constructed with MrBayes[12] implementing the HKY85+G model, as identified by ModelTest[16], from two runs of five million generations and four heated chains each. The posterior probabilities and consensus tree of 66 unique 286 bp-long haplotypes (Supplementary Fig. S2) constructed with MrBayes were visualized with FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/). The same topology of the tree was obtained with PhyML[9] with various branch support methods – bootstrap (1000 iterations) and SH.

One of three nuclear fragments could be amplified in 97 samples in at least one amplification reaction. The Tabby-T139A allele could be amplified and sequenced in 63 samples, the Tabby-D228N in 88 samples and the Tabby-W841X allele in 67 samples. The three fragments together were successfully amplified and sequenced in 48 samples.

Authentication of aDNA data

1) In the nine Ion Torrent sequencing runs carried out, 50 blank extractions and PCR controls were processed in parallel with the samples, providing an average coverage of 0.54, ranging from 0.09 to 1.52, for the mtDNA amplicons (with a maximum of five reads observed), and 0.09 ranging from 0 to 0.33 for the nuclear DNA amplicons (with a maximum of two reads). Only in one instance, an extraction control provided from 7 to 20 reads in six mtDNA fragments, a result that nevertheless was not replicated in a second control from the same extraction batch. All successful samples from that sequencing run showed a larger number of reads (>150) and the sequences were replicated in independent experiments. Furthermore, all the successful samples from that extraction batch provided results replicated at high average coverage (>150 reads) from multiple amplification and/or extractions or showed a different sequence from that in the blank control. In the nested PCR approach no extraction or PCR control was ever amplified.

2) In all instances, mtDNA haplotypes obtained by Ion Torrent sequencing or pyrosequencing were reproduced in multiple experiments, from at least two and up to eight independent multiplex PCR experiments from one or up to three DNA extracts (e.g., samples BM06 and OXY3). When possible, independent DNA extractions were performed from anatomically distant bone samples or from different tissues (e.g. skin, hair and bone for the mummified samples).

Given the lower chance of successful amplification of nuclear fragments and the extremely low number of reads observed in the negative controls (average read number 0.09), in some instances results could not be replicated in independent amplification and yet considered authentic when at least 20 reads were obtained (see Supplementary Dataset S2). For this reason and given the degraded nature of aDNA, we cannot exclude that, due to allelic drop-out, our strategy for nuclear SNPs analysis might have underestimated the number of alleles observed for the three SNPs examined. Therefore, we adopted a conservative strategy accounting only for the minimum number of alleles observed.

3) Samples in which one or more mtDNA fragments failed to amplify or were poorly represented after multiple attempts of multiplex PCR amplification were re-analyzed using either *i)* simplex PCRs to increase the chances of successful amplification of the missing or poorly represented fragments and sequenced in the Ion Torrent, or *ii)* independent simplex PCRs, followed by nested qPCR with a 5'-biotinylated primer and pyrosequencing.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Samples in which all fragments failed to amplify or that were covered by <10 reads and those in which inconsistencies of the sequences were found across multiple replicates were discarded and considered unsuccessful.

4) Of the 150 successful samples with a complete 449 bp-long profile sequenced with the Ion Torrent, in 33 the average coverage obtained from independent replicates of the multiple PCRs was >1000 reads, in 28 samples it was between 500 and 1000, in 71 samples between 50 and 500, and in 18 samples it was between 14 and 50. In the latter case, to ensure higher reliability of poorly covered fragments, simplex PCRs were performed that allowed obtaining up to thousand reads (Supplementary Dataset S2). In samples where both Ion Torrent and pyrosequencing techniques were applied on the same fragment, the results were always consistent.
5) Nuclear fragments were represented in all samples by >10 reads in two or more independent PCR amplifications, or by >20 reads in samples where a single amplification was obtained (Supplementary Dataset S2). In 73 samples the average coverage of the successful nuclear fragments amplified was >100 reads.
6) DNA damage profiles could not be used to authenticate the ancient nature of the obtained sequences since: (i) DNA damage at the ends of the initial ancient DNA molecules are lost when targeted PCR is used; (ii) our systematic use of UNG to prevent carry-over contamination simultaneously lowers the contribution of post-mortem DNA modifications to a negligible level[17]. Furthermore, to ensure that the validated DNA sequences were not erroneous due to diagenetic transformations we used three strategies: (i) the aforementioned UNG treatment; (ii) bulk sequencing of PCR products; (iii) independent PCR replication.

Iconographic evidence in Europe

Representations of cats in domestic contexts are known from various artifacts and funerary monuments in the Greek world during the 6th c. BC. Here we report a list of some of the iconographic depictions.

- The Arkesilas Cup found in the Etruscan site of Vulci and dated to about 565/560 BC (Cabinet des médailles of the Bibliothèque nationale de France in Paris (inv. 189);
- The funerary statue of Themistocle 510 BC (National Archaeological Museum, Athens); (see also Luce, J-M. (2015) Les chats dans l'Antiquité grecque. In "Chiens & Chats dans la Préhistoire et l'Antiquité", Bellier, C., Cattelain, L. et Cattelain, P. (eds), Editions du Cedarc, Bruxelles.)
- The red-figure vases of the "Cat and Dog Painter" (The Beazley Archive Pottery Database, BAPD, http://www.beazley.ox.ac.uk/XDB/ASP/default.asp)
- Greek Coins (Kraay, Colin M (1976), Archaic and Classical Greek Coins, New York : Sanford J. Durst); for images see : Lauwers, C. "Des monnaies entre chien et loup" In "Chiens & Chats dans la Préhistoire et l'Antiquité", Bellier, C., Cattelain, L. et Cattelain, P. (eds), Editions du Cedarc, Bruxelles)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

References

1. Bennett, E.A. *et al.* Library construction for ancient genomics: single strand or double strand? *Biotechniques* **56**, 289-90, 292-6, 298, passim (2014).
2. Champlot, S. *et al.* An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* **5**, e13042 (2010).
3. Pruvost, M., Grange, T. & Geigl, E.-M. Minimizing DNA-contamination by using UNG-coupled quantitative real-time PCR (UQPCR) on degraded DNA samples: application to ancient DNA studies. *Biotechniques* **38**, 569-575 (2005).
4. Rompler, H. *et al.* Multiplex amplification of ancient DNA. *Nat Protoc* **1**, 720-8 (2006).
5. Pfeiffer, I., Völkel, I., Täubert, H. & Brenig, B. Forensic DNA-typing of dog hair: DNA-extraction and PCR amplification. *Forensic Science International* **141**, 149-151 (2004).
6. Driscoll, C.A. *et al.* The Near Eastern Origin of Cat Domestication. *Science* **317**, 519-523 (2007).
7. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).
8. Lopez, J.V., Cevario, S. & O'Brien, S.J. Complete Nucleotide Sequences of the Domestic Cat (Felis catus) Mitochondrial Genome and a Transposed mtDNA Tandem Repeat (Numt) in the Nuclear Genome. *Genomics* **33**, 229-246 (1996).
9. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).
10. Guimaraes, S. *et al.* A cost-effective high-throughput metabarcoding approach powerful enough to genotype ~44 000 year-old rodent remains from Northern Africa. *Mol Ecol Resour*, doi: 10.1111/1755-0998.12565 (2016).
11. Shen, Z. *et al.* MPprimer: a program for reliable multiplex PCR primer design. *BMC Bioinformatics* **11**, 143 (2010).
12. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539-42 (2012).
13. Kaelin, C.B. *et al.* Specifying and sustaining pigmentation patterns in domestic and wild cats. *Science* **337**, 1536-41 (2012).
14. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
15. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
16. Posada, D. & Crandall, K.A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-8 (1998).
17. Pruvost, M. *et al.* DNA diagenesis and palaeogenetic analysis: critical assessment and methodological progress. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **266**, 211-219 (2008).
18. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Supplementary datasets

Dataset S1. List of all the ancient and modern samples analyzed in this study. Unsuccessful and successful samples with a complete and incomplete mtDNA profile are indicated in red, green and yellow respectively.

Dataset S2. List of samples successfully analyzed in this study and detailed information about the samples, the dating, the genotyping procedures followed, the mtDNA haplotypes and the polymorphic states of the three nuclear markers investigated. For samples analyzed with the aMPlex Torrent approach, the average number of reads obtained in all fragments amplified for the mtDNA and the nuDNA fragments is reported (in samples in which simplex amplifications were performed, the maximum and minimum number of reads obtained from different fragments and/or PCRs reactions is indicated in brackets). The color code associated to sample ID in column B is identical to that used in Dataset1 whereas that used for the clade/subclade in column J is identical to that used in the various trees depicted in Fig.1, Supplementary Fig. S1 and S2.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Supplementary figures

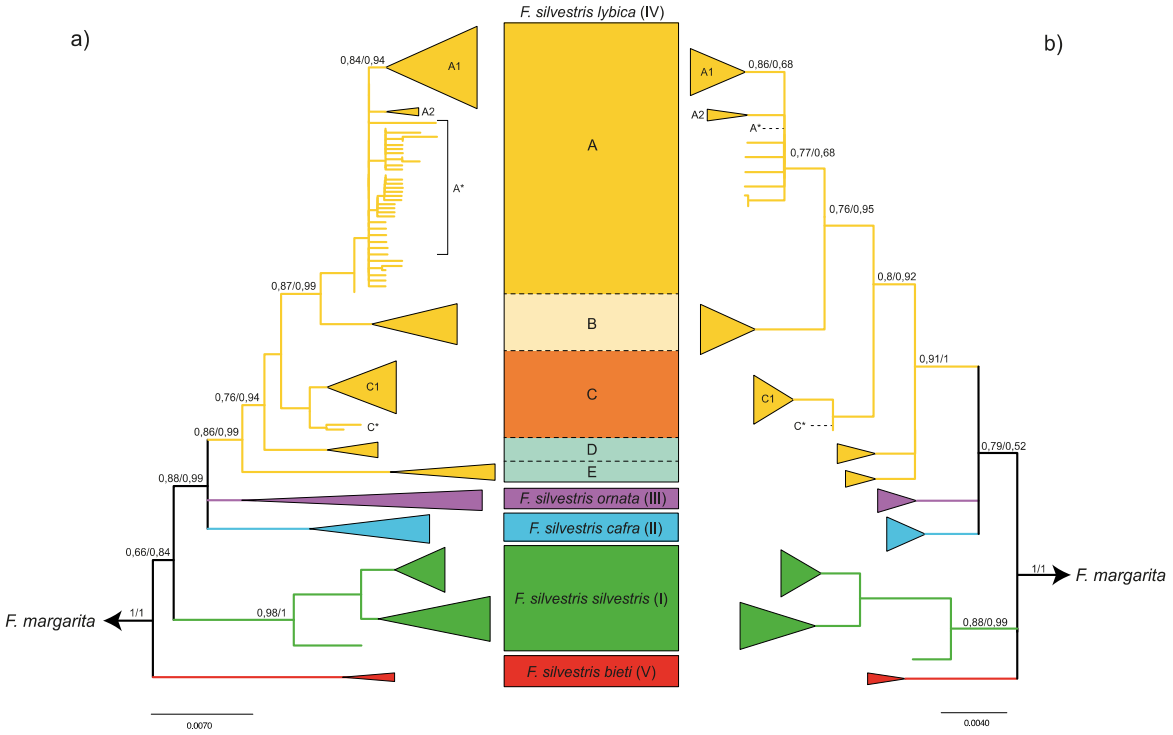

Figure S1. a) Maximum-likelihood (ML) trees of haplotypes from Driscoll *et* al.[6] and b) of the same haplotype data reduced to the minimum sequence length − 286 bp − generated by our assay. Subspecies and clade names as described by Driscoll et al. are reported in the rectangular shape between the trees, and colors of clades and subclades are as in haplotypes of Fig. 1b. Names of subclades within clade IV (*F. s. lybica*) dubbed in this study are reported in each tree. Some of the subclades and lineages observed in the 2007 study of Driscoll *et al*. were collapsed in a single root haplotype (IV-A*, IV-C*). The ML trees were generated using PHYML3.0 [9,18] and the Shimodaira-Hasegawa-like branch test (SH) was used to evaluate the statistical support of the nodes. A Bayesian tree generated using Mr Bayes had the same topology and was used to evaluate the posterior probabilities of the nodes. Statistical support for the main nodes is reported as SH/posterior probability values.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

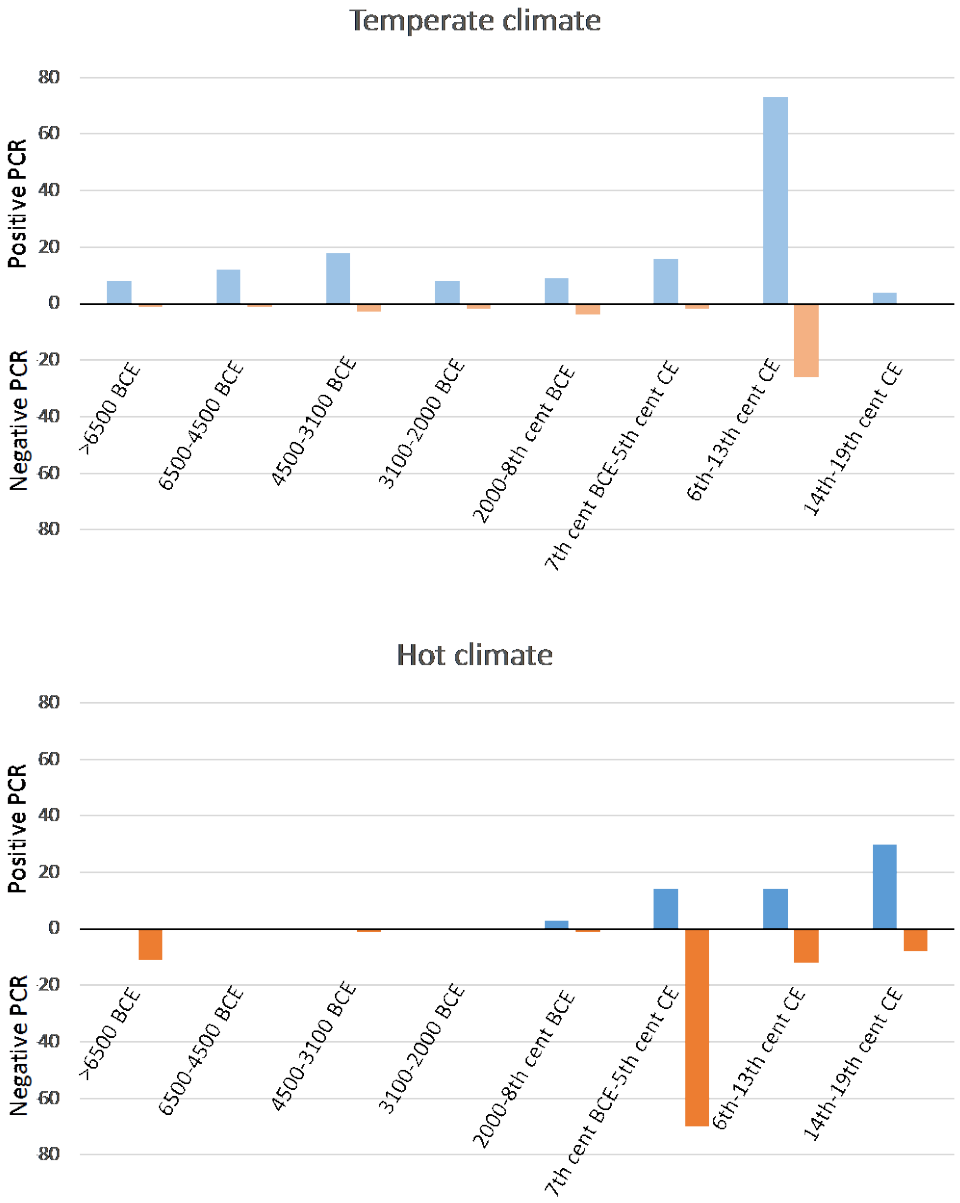

Figure S2. Distribution of the samples allowing or not PCR amplifications according to age and environmental temperature. Samples are classified in time bins as indicated and countries of origin are classified according to climate either as "temperate" on the top graph (Europe, Armenia, Iran and Turkey) or "hot" on the bottom graph (Africa, Jordan, Lebanon, Oman, Saudi Arabia, Syria, Egypt). The number of samples that yielded PCR products in each category are represented in blue above the zero line with a scale oriented upward, whereas the number of PCR negative samples are represented in orange below the zero line with a scale oriented downward. The percentage of success in samples from the temperate zone was about 80% (148 samples), ranging from 70 – 90% for all time bins. In contrast, in hot climate zones the success rates were overall below 40% (164 samples) ranging from 0 in samples older than 4,000 years to 80% in the most recent time bin.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

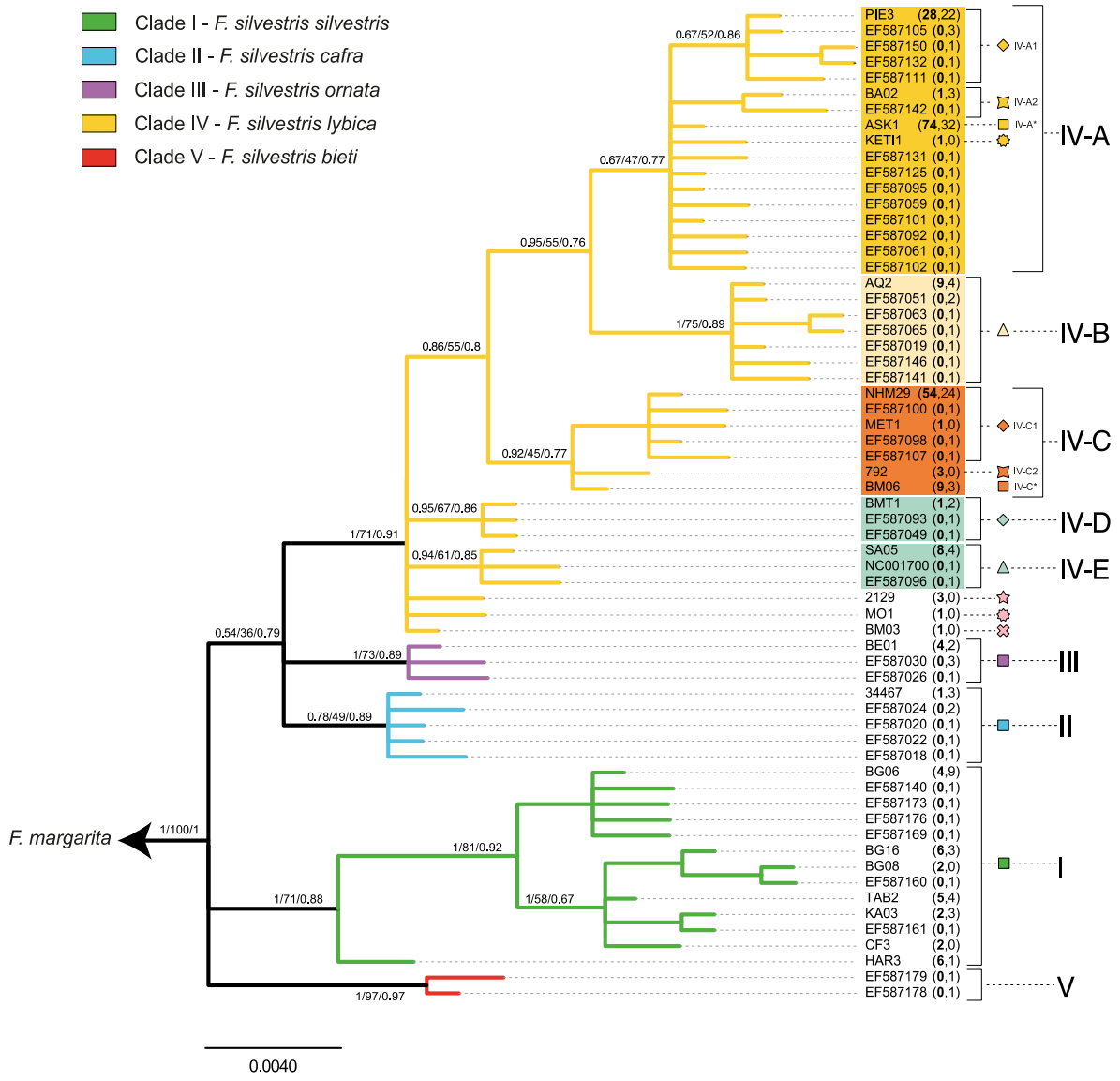

Figure S3. Bayesian tree of 286 bp-long haplotype data from this study and literature [6], represented in a schematic way in Fig. 1b. Colors of branches in the five *Felis silvestris* clades are as in figure 1. Statistical support values are reported as Bayesian posterior probability/bootstrap/SH-values obtained by three different methods (Mr Bayes, PHYML3.0, PHYML3.0, respectively) reproducing the same topology. Each haplotype is indicated by the name of one specimen (or a Genbank accession number for sequences from the literature not found in the dataset generated in this study). For each haplotype, the number of sequences from this study (in bold) and from the 2007 study of Driscoll *et al*. (haplotypes deposited in GenBank) sharing the same 286 bp-haplotype is indicated in brackets.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

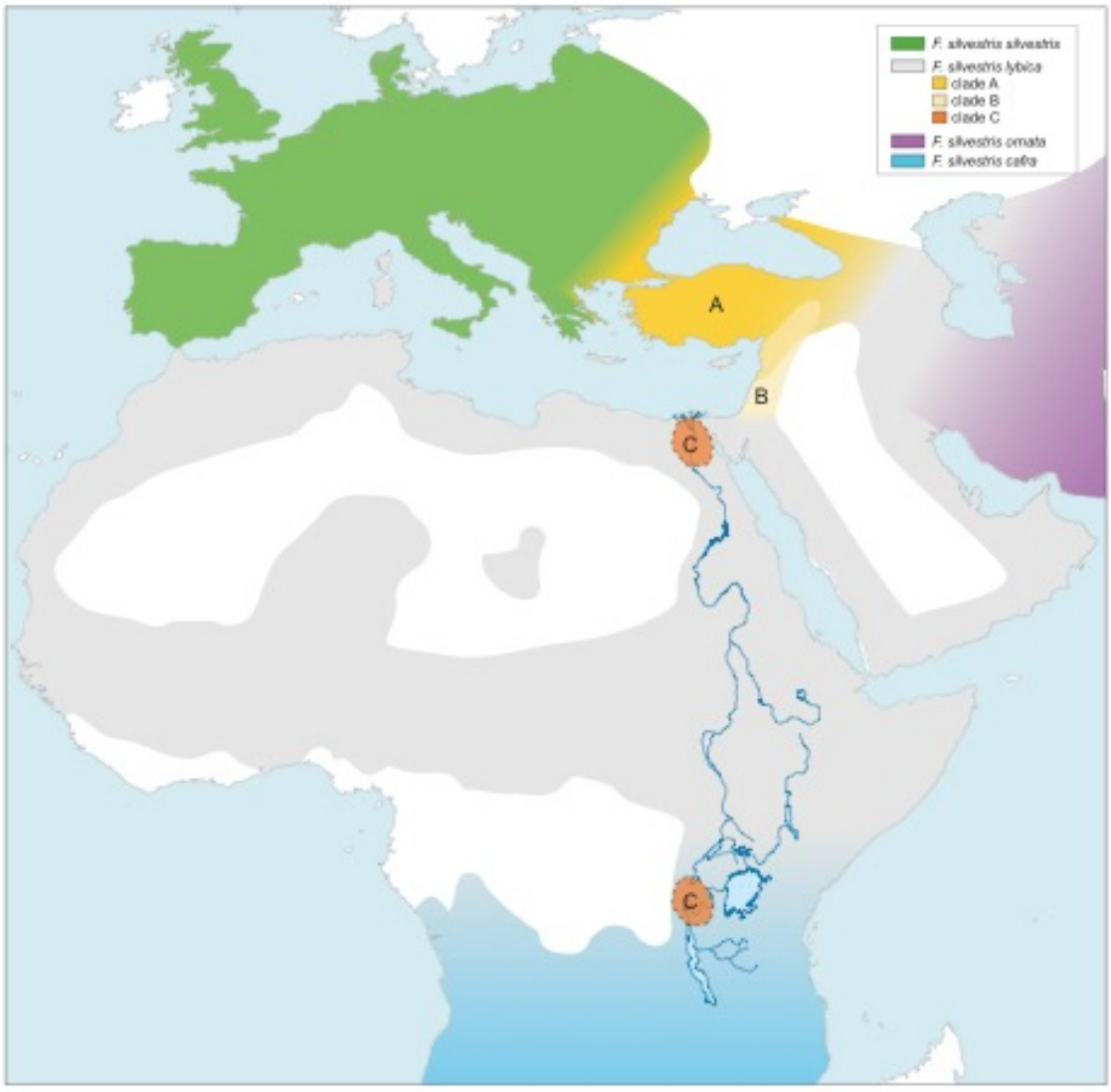

Figure S4. Representation of (pre)historical geographic range of *Felis silvestris* mtDNA clades in Western Eurasia and Africa as inferred in this study. The entire *F. s. lybica* (clade IV) range is in grey, and each specific clade distribution inferred from our ancient dataset has a specific color as in the tree of figure S2 and figure 1. Our dataset suggests a wide distribution of clade C on the African continent, but we reported only the areas where we directly observed it (delimited by a broken line).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Supplementary tables

Table S1. All PCR primers used in multiplex and simplex PCRs in this study

Primer	Primer seq (5'-3')	Final conc (nM)	Amplicon length (bp)	Gene-Region	PCR
Frag_1F	TGCTACTAGGACAACCACGATTC	100	83	ND5-mtDNA	Mplx-B
Frag_1R	GAGACGTTTAATGGAGTTGATGAG	100			
Frag_1F2	CTACTAGGACAACCACGATTCA	760	77	ND5-mtDNA	Splx-Fr1
Frag_1R10	[Btn]-CGTTTAATGGAGTTGATGAGG	760			
Frag_2F1	CCGCCAACAATCTATTTCAAC	50	89	ND5-mtDNA	Mplx-A
Frag_2R	TGCTCGGCCATATCATCATC	50			
Frag_2F1	GCCAACAATCTATTTCAACTATTC	760	80	ND5-mtDNA	Splx-Fr2
Frag_2R3	[Btn]-CCATATCATCATCCGATAAGT	760			
Frag_3F	TTGTTCACAGCTATTTGTGCTCTC	250	106	ND5-mtDNA	Mplx-A
Frag_3R	GGTTAATCCCAATGGTTACAATTAT	250			
Frag_3F5	ATTTGTGCTCTCACACAAAAT	760	92	ND5-mtDNA	Splx-Fr3
Frag_3R1	[Btn]-TTAATCCCAATGGTTACAATTAT	760			
Frag_4F	CCAATGCCCTTCACCACTAC	100	105	ND5-mtDNA	Mplx-B
Frag_4R	GGCTGTCTCGATGATTAGGTCTT	100			
Frag_4F2	AATGCCCTTCACCACTACCT	760	83	ND5-mtDNA	Splx-Fr4
Frag_4R3	[Btn]-CTTTGGAATAAAAACCTGTTAGGA	760			
Frag_5F3	CTCATCAACTCCATTAAACGTCTC	175	60	ND5-mtDNA	Mplx-A
Frag_5R3	AGAAATTAGATATCCTGCAAAGATACT	175			
Frag_5F2	CATCAACTCCATTAAACGTCTCT	760	53	ND5-mtDNA	Splx-Fr5
Frag_5R1	[Btn]-TTAGATATCCTGCAAAGATACTTC	760			
Frag_6F	CGAAAATCACACCCCCTTATCA	175	103	cytB-mtDNA	Mplx-A
Frag_6R	CTCCTAGAAGGGAGCCGAAGT	175			
Frag_6F1	ACACCCCCTTATCAAAATTA	760	83	cytB-mtDNA	Splx-Fr6
Frag_6R1	[Btn]-AGCCGAAGTTTCATCATG	760			
Frag_7F2	CTTTATCCCCGTAGCGCTTTT	75	87	ND5-mtDNA	Mplx-B
Frag_7R2	AATCGGTTGATGTATGGGTCTG	75			
Frag_7F3	TCGTCACATGGTCCATCAT	760	63	ND5-mtDNA	Splx-Fr7
Frag_7R7	[Btn]-GGTTGATGTATGGGTCTGAGT	760			
Frag_8F1g	CACCCAGCACGAGAACCTAAA	200	111	ND5-mtDNA	Mplx-B
Frag_8R1	GGGCCTTCTATGGCTGATG	200			
Frag_8.1F2	[Btn]-ACCCAACACGAGAACCTAAATA	760	72	ND5-mtDNA	Splx-Fr8.1
Frag_8.1R1	GAATTGGGCGGATTTACCT	760			
Frag_8.2F2	AAATATTCCATTACTAGGGCTTC	760	88	ND5-mtDNA	Splx-Fr8.2
Frag_8.2R2	[Btn]-TTCTATGGCTGATGGCAGT	760			
Frag_9 F	TCTCGCGGCTAAAAACTTAAAA	300	101	ND5-mtDNA	Mplx-B
Frag_9 R	TTGATGGGAGGCGGTGTATT	300			
Frag_9F1	CGCGGCTAAAAACTTAAAAT	760	78	ND5-mtDNA	Splx-Fr9
Frag_9R1	[Btn]-ACAATTGGAAAGTACCCTAAGA	760			
T139_58F	CACCGGCCGCGTGAACAT	150	91	Taqpep-chrA1	Mplx-C
T139_148R	CCTCGGCGCTCTCGCAGTC	150			
D228_324F	TCAGGGAGGGGCTCTTCTT	150	64	Taqpep-chrA1	Mplx-C
D228_387R	CCGGCCTGTCCTTACCTG	150			
W841_80655 F	GCGTTGCCTTCGGAAGTG	150	101	Taqpep-chrA1	Mplx-C
W841_80755 R	GCTCATCGCATAAGTGAGTTGAAT	150			

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table S2. Pyrosequencing primers (mtDNA)

Primer	Primer seq (5'-3')
Frag_1S1	GGACAACCACGATTCA
Frag_2S1	AATCTATTTCAACTATTCAT
Frag_3S1	GCTCTCACACAAAATGATAT
Frag_4S1	CCTTCACCACTACCTCC
Frag_5S1	ACTCCATTAAACGTCTCTTA
Frag_6S1	CCCCCTTATCAAAATTAT
Frag_7S1	GGTCCATCATAGAATTCTC
Frag_8AS1	TTACCTGTGGCTGCT
Frag_8BS2	TCCATTACTAGGGCTTCT
Frag_9S1	CGGCTAAAAACTTAAAATT

Table S3. Sequences to analyse in pyrosequencing reactions (mtDNA)

Primer	Primer seq (5'-3')
Frag_1 #C	AYWCCTTRAATCTAATCAATGAAAATCAAYACCC
Frag_2	CGGYTGAGAGGGAGTRGGAATYATATCYTTYCT
Frag_3	YAAAAAAATTGTTGCCTTTTCAACCTCMAGCCAACTGGGYCTRAT
Frag_4	YTAATCATTGGAAGCCTCGCRCTYACRGGTA
Frag_5	ATYGGAA
Frag_6	TAAYCACTCATTCATCGAYCTACCYRCCCCATCTAACATCTCARC
Frag _7	ARTRTGRTAYATRYAC
Frag_8A	ARTAGRAGYCCTAGTAATGGARTA
Frag_8B	AYTAGCAGCCACAGGYAAATCCGCCCAATTCGGCCTRCATCCRTG
Frag_9	YAYATACCCYTCAARYCTCTTTAAGTTTTCYAA

Table S4. Details of Ion Torrent sequencing runs

ION	Barcoded ancient samples (N)	Polyclonal	Total reads	Mean read length (bp)	Aligned reads
31	88	43%	265,745	78	136,382
32	88	41%	377,657	56	28,041
33	88	34%	261,117	72	148,205
34	88	39%	254,715	58	24,252
40	80	44%	515,694	79	343,414
41	88	44%	536,969	85	414,676
42	88	40%	566,386	92	325,774
44	88	31%	658,721	92	494,968
46	80	32%	629,746	86	376,467

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table S5. Approximate geographic locations of the sites from which the samples are derived. Key numbers are reported as in Fig. 1.

Key	Country	Location
1	Germany	Berlin, Ralswiek.
2	Belgium	Trou Chaleux.
3	France	Entzheim-Geispolsheim, Entzheim-Lidl, Illfurth Buergelen.
4	Italy	Galgenbühel/Dos de la Forca.
5	Spain	Cova Fosca, Tabernas, Valencina.
6	Greece	Kastanas, Kassope, Tiryns.
7	Romania	Cheia, Hârşova, Icoana, Pietrele, Vităneşti.
8	Bulgaria	Arbanas, Budjaka, Dolnoslav, Durankulak, Garvan, Hotnitsa, Jambol, Kabile, Koprivec, Sliven.
9	Turkey	Didyma, Kadıkalesi.
10	Turkey	Demircihüyük, Yenikapı.
11	Turkey	Sagalassos, Bademağacı.
12	Turkey	Aşıklı Höyük.
13	Turkey	Norşun Tepe, Hassek Höyük, Lidar Höyük, Anzaf.
14	Armenia	Keti, Dvin.
15	Syria Lebanon	Tell Guftan, Shheil 1, Qasr al-Hayr al-Sharqi. Sidon.
16	Jordan	Aqaba, Tall al-'Umay'ri, Hesban, Khirbet es-Samra.
17	Iran	Siraf.
18	Saudi Arabia	Al-Yamama.
19	Oman	Qalhât.
20	Egypt	mummies from Natural History Museum and British Museum (London, UK).
.21	Egypt	Oxyrhynchus.
22	Egypt	Shenhur.
23	Egypt	Berenike.
24	Tunisia	Bigua.
25	Morocco	Mogador.
26	Senegal	Gorée.
27	Congo	Modern wildcats from museum collections.
28	Burundi-Rwanda	Modern wildcats from museum collections.
29	Kenya	Modern wildcats from museum collections.
30	Tanzania	Songo Mnara, Unguja Ukuu, Makangale (Mapangani) Cave.
31	Angola	Modern wildcats.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.