MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 42–55 doi: 10.1111/1755-0998.12420

Impact of enrichment conditions on cross-species capture of fresh and degraded DNA

JOHANNA L. A. PAIJMANS,*†§ JOERNS FICKEL,†‡ ALEXANDRE COURTIOL,‡ MICHAEL HOFREITER*† and DANIEL W. FÖRSTER‡

*Department of Biology, University of York, York, YO10 5DD, UK, †Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str 24-25, 14476 Potsdam, Germany, ‡Evolutionary Genetics Department, Leibniz-Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str. 17, 10315 Berlin, Germany

Abstract

By combining high-throughput sequencing with target enrichment ('hybridization capture'), researchers are able to obtain molecular data from genomic regions of interest for projects that are otherwise constrained by sample quality (e.g. degraded and contamination-rich samples) or a lack of *a priori* sequence information (e.g. studies on nonmodel species). Despite the use of hybridization capture in various fields of research for many years, the impact of enrichment conditions on capture success is not yet thoroughly understood. We evaluated the impact of a key parameter – hybridization temperature – on the capture success of mitochondrial genomes across the carnivoran family Felidae. Capture was carried out for a range of sample types (fresh, archival, ancient) with varying levels of sequence divergence between bait and target (i.e. across a range of species) using pools of individually indexed libraries on Agilent SureSelect™ arrays. Our results suggest that hybridization capture protocols require specific optimization for the sample type that is being investigated. Hybridization temperature affected the proportion of on-target sequences following capture: for degraded samples, we obtained the best results with a hybridization temperature of 65 °C, while a touchdown approach (65 °C down to 50 °C) yielded the best results for fresh samples. Evaluation of capture performance at a regional scale (sliding window approach) revealed no significant improvement in the recovery of DNA fragments with high sequence divergence from the bait at any of the tested hybridization temperatures, suggesting that hybridization temperature may not be the critical parameter for the enrichment of divergent fragments.

Keywords: degraded DNA, Felidae, hybridization capture, mitogenomes, next-generation sequencing, sequence enrichment

Received 3 December 2014; revision received 21 April 2015; accepted 23 April 2015

Introduction

The combination of high-throughput sequencing with target-sequence enrichment methods plays an important role in many disciplines, ranging from human health to ecology and genomics. While it is now possible to obtain complete genomes, sequencing costs for population-scale genomic studies are not insignificant, and for some research questions it is more appropriate to obtain a relatively limited number of loci (100s–1000s) for a large number of individuals than a large number of loci (i.e. complete genomes) from a limited number of individuals (Lemmon & Lemmon 2013). Enrichment utilizes DNA or RNA baits with sequence complementarity to the genomic region(s) of interest that can be immobilized on arrays (Hodges *et al.* 2009) or beads (Gnirke *et al.* 2009). The baits hybridize to the desired DNA fragments in a DNA pool, and nontarget DNA is washed away. Target regions are then released from the baits and sequenced on high-throughput sequencing platforms (Mamanova *et al.* 2010). Thus, enrichment is a proportional increase in target sequences, achieved by a reduction in nontarget sequences. Commercial target enrichment assays are readily available from multiple suppliers, allowing researchers to address a wide scope of questions when studying humans and model organisms (e.g. Bodi *et al.* 2013; Elhaik *et al.* 2013). In addition to such predesigned commercial enrichment assays, custom-designed protocols are available for use in research on nonmodel organisms (e.g. Maricic *et al.* 2010; Bi *et al.* 2012; Peñalba *et al.* 2014).

Correspondence: Johanna L. A. Paijmans and Daniel W. Förster, Fax: +49 (0)331 977-6389; E-mails: paijmans.jla@gmail.com, dwgfoerster@gmail.com

§Present address: Institute for Biochemistry and Biology University of Potsdam, Karl-Liebknecht-Str 24-25, 14476 Potsdam, Germany

Whether the aim is to subsample the genome to focus on a specific region (e.g. exome), or to enrich highly contaminated or degraded samples for endogenous DNA, hybridization capture has proved to be a versatile approach. Furthermore, the flexibility of hybridization capture allows for the investigation of species for which no *a priori* sequence information is available using baits designed from the sequence(s) of the target region(s) of a close relative. Cross-species capture thus mitigates problems typically associated with the investigation of non-model organisms and extinct lineages, as it allows for the capture of previously unknown sequences from a variety of species (e.g. Mason *et al.* 2011; Bi *et al.* 2012; Hancock-Hanser *et al.* 2013; Mitchell *et al.* 2014; Peñalba *et al.* 2014). The approach has also proved to be highly suitable for the analysis of degraded samples, such as environmental samples or ancient and archival samples, by targeting endogenous DNA even in the presence of a large amount of contaminant DNA (Rizzi *et al.* 2012; Hofreiter *et al.* 2015). This has resulted in the recovery of ancient genetic information from mitochondrial genomes (mitogenomes), exomes and even whole nuclear genomes (e.g. Carpenter *et al.* 2013; Castellano *et al.* 2014; Devault *et al.* 2014; Enk *et al.* 2014).

In addition to the sequence similarity between bait and target, other factors that may impact the efficiency of hybridization capture are hybridization temperature, bait tiling and posthybridization washing temperatures (Ávila-Arcos *et al.* 2011; Bodi *et al.* 2013; Li *et al.* 2013). To our knowledge, only one study has explicitly tested the impact of hybridization temperature (Li *et al.* 2013); the authors found that for MYBaits™ in-solution capture, a touchdown approach (i.e. a stepwise decrease in temperature during hybridization) significantly improved capture efficiency compared to using the 65 °C suggested by standard protocols (Gnirke *et al.* 2009). Recent studies have used lower hybridization temperatures (60, 50, 48 and 45 °C or a touchdown approach) aiming to increase capture efficiency of short fragments generally found in ancient extracts (Mason *et al.* 2011; Dabney *et al.* 2013; Enk *et al.* 2013, 2014; Meyer *et al.* 2013; Templeton *et al.* 2013; Mitchell *et al.* 2014). However, it has not yet been investigated how these protocols compare to the standard hybridization temperature of 65 °C or the touchdown approach.

In this study, we have explicitly evaluated the impact of the hybridization temperature, a key experimental parameter, on the specificity of cross-species capture using a solid-state enrichment method [Agilent Microarrays (Hodges *et al.* 2007, 2009)]. Whole mitogenomes were targeted for 21 species within the Felidae, including at least one representative of each of the eight major lineages (based on Johnson 2006), as well as the extinct scimitar cat *Homotherium serum* (basal to all extant felids; Barnett *et al.* 2005). We chose to focus on the mitogenome because reference sequences for this marker are relatively abundant in public databases, and molecular studies of extinct populations or species (i.e. archival and subfossil material) have often focused on this marker (reviewed by Paijmans *et al.* 2013). Our analysis focuses on two main topics: (i) the impact of hybridization temperature on the proportion of on-target sequences and target coverage for different sample types and (ii) the impact of regional bait–target sequence similarity on capture success.

Materials and methods

Samples

Our cross-species capture experiment was carried out on 25 samples representing 21 felid species, including 19 fresh samples (tissue or blood), four ancient bone samples (~30 000–15 000 years old) and two archival bone samples (up to 125 years old; Fig. 1a; Table 1). Fresh samples were extracted using the GEN-IAL all-tissue DNA-extraction-kit (GEN-IAL GmbH, Troisdorf, Germany). Archival and ancient samples were extracted in dedicated ancient DNA facilities at the University of York, with appropriate contamination prevention in place (Knapp *et al.* 2012). DNA extraction of the ancient samples (bone) was performed using a silica-based extraction protocol (Rohland *et al.* 2010), including a total of six extraction blanks as contamination controls. Archival samples (including two blanks) were extracted using the extraction buffer from Rohland *et al.* (2004), while the remainder of the procedure was performed using spin columns (Rohland *et al.* 2010). For additional details regarding the extraction procedure, see Appendix S1 (Supporting information).

Library preparation

Illumina libraries for paired-end sequencing were prepared following Fortes & Paijmans (2015). Library preparation for archival and ancient samples was performed in dedicated ancient DNA facilities. In addition to the regular index for multiplex sequencing (sometimes referred to as barcode or ID-tag) nested within the Illumina P7 adapter, we also utilized an in-line index at the 3′ end of the P5 Illumina adapter (5′ end of the template) that is incorporated during the primary adapter ligation. This second index is incorporated to detect potential PCR artefacts or chimeras between library templates (Kircher *et al.* 2011; Fortes & Paijmans 2015).

Following primary library amplification, each library was split into four aliquots, each of which was indexed with a unique P7-indexing primer in an independent

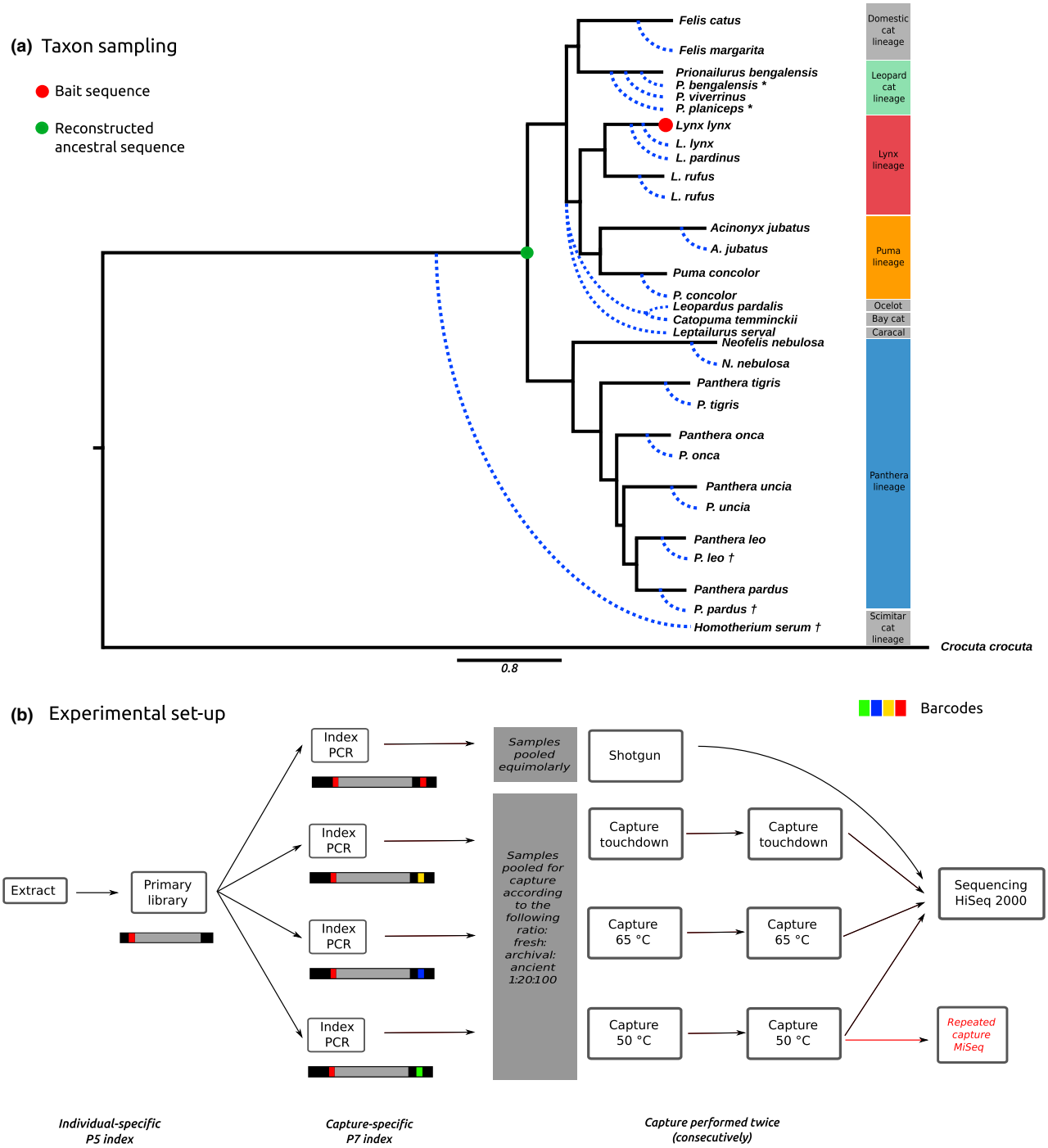

Fig. 1 (a) Phylogeny of the extant Felid species that have a complete mitogenome currently available on GenBank (for phylogenetic reconstruction methods, see Appendix S1), in approximate phylogenetic relation to the species included in this study (displayed by blue dotted branches; Table 1). The capture array was designed based on *Lynx lynx* (indicated by a red dot). The reconstructed ancestral sequence that was used as reference for mapping in some analyses is indicated by the green dot. Archival samples are indicated by an asterisk (*) and ancient samples by a dagger (†). (b) Flow chart of the experimental design for library preparation and enrichment.

PCR (Fig. 1b; Appendix S1, Supporting information). In this manner, we obtained 100 (25 × 4) indexed libraries, with the P7-index being unique to each library (total 100) and the P5-index unique to each individual sample (total 25). Four library pools were created, each containing the full set of 25 samples: one set for shotgun sequencing (i.e. no capture), one set for capture at a hybridization temperature of 65 °C, one set for capture at a hybridization

Table 1 Samples included in this study and the closest available reference mitogenome available on GenBank

Species code	Species name	Common name	Lineage (according to Johnson 2006)	Sample numbers		Closest reference	Reference Genbank Acc Nr.
				Fresh	Ancient/Archival		
AJU	*Acinonyx jubatus*	Cheetah	Puma lineage	AJU_001		*Acinonyx jubatus*	AY463959
CTE	*Catopuma temminckii*	Asian golden cat	Bay cat lineage	CTE_28		*Prionailurus bengalensis*	HM185183
FMA	*Felis margarita*	Desert cat	Domestic cat lineage	FMA_313		*Felis catus*	U20753
HSE	*Homotherium serum*	Homotherium	Scimitar cat lineage		HSE_1714†; HSE_1007†	*Lynx lynx*	KM982549
LLY	*Lynx lynx*	Eurasian lynx	Lynx lineage	LLY_2P01		*Lynx lynx*	KM982549
LP	*Leopardus pardalis*	Ocelot	Ocelot lineage	LP_534		*Lynx lynx*	KM982549
LPA	*Lynx pardinus*	Iberian lynx	Lynx lineage	LPA_placenta		*Lynx lynx*	KM982549
LRU	*Lynx rufus*	Bobcat	Lynx lineage	LRU_1303		*Lynx rufus*	GQ979707
LSE	*Leptailurus serval*	Serval	Caracal lineage	LSE_Togo		*Lynx rufus*	GQ979707
NNE	*Neofelis nebulosa*	Clouded leopard	Panthera lineage	NNE_Fell		*Neofelis nebulosa*	DQ257669
OMA	*Otocolobus manul*	Pallas cat	Domestic cat lineage	OMA_483		*Prionailurus bengalensis*	HM185183
PBE	*Prionailurus bengalensis*	Asian leopard cat	Leopard cat lineage	PBE_83	PBE_55*	*Prionailurus bengalensis*	HM185183
PCO	*Puma concolor*	Puma	Puma lineage	PCO_526		*Puma concolor*	JN999997
PLE	*Panthera leo*	Lion	Panthera lineage	PLE_001	PLE_1684†	*Panthera leo persica*	KC834784
PO	*Panthera onca*	Jaguar	Panthera lineage	PO_1		*Panthera onca*	KF483864
PPA	*Panthera pardus*	Leopard	Panthera lineage	PPA_032	PPA_2177†	*Panthera pardus*	EF551002
PPL	*Prionailurus planiceps*	Flat-headed cat	Leopard cat lineage		PPL_001*	*Prionailurus bengalensis*	HM185183
PRU	*Prionailurus rubiginosis*	Rusty-spotted cat	Leopard cat lineage	PRU_402		*Prionailurus bengalensis*	HM185183
PTS	*Panthera tigris*	Tiger	Panthera lineage	PTS_001		*Panthera tigris*	EF551003
PU	*Panthera uncia*	Snow leopard	Panthera lineage	PU_001		*Panthera uncia*	EF551004
PVI	*Prionailurus viverrinus*	Fishing cat	Leopard cat lineage	PVI_62		*Prionailurus bengalensis*	HM185183

Archival samples are indicated by an asterisk (*) and ancient samples by a dagger (†).

temperature of 50 °C and one set for capture using a touchdown approach. For the three capture experiments, samples were pooled to account for the difference in expected endogenous DNA content in the libraries, a ratio of 1:20:100 for fresh vs. archival vs. ancient samples [expectation of 100%, 5% and 1% endogenous DNA, respectively (M. Hofreiter, pers. obs.); Fig. 1b]. This corresponds to 15 ng of library for each fresh sample, 75 ng of library for each archival sample and 1500 ng of library for each ancient sample. Pooling multiple samples on an array reduces the cost of array capture, and in this study, it allowed us to control conditions between the different hybridization experiments. Shotgun sequencing was performed to establish the original endogenous content in libraries and therefore did not require such a pooling strategy.

Enrichment

Agilent DNA SureSelect™ 244k microarrays (Hodges *et al.* 2009) were used for the enrichment of libraries. Microarrays were designed based on the mitogenome sequence of the Eurasian lynx *Lynx lynx* (GenBank Accession no. KM982549; Appendix S1, Supporting information). Capture was performed according to the protocol in Fortes & Paijmans (2015), using three different hybridization temperatures: 65 °C, 50 °C and 'touchdown' [starting at 65 °C, lowered in 5 °C increments every 16¼ h to a final temperature of 50 °C, following recommendations by Li *et al.* (2013)]. Library pools were serially captured two consecutive times under the same conditions, as a second capture round has been shown to improve the enrichment rate (Li *et al.* 2013; Templeton *et al.* 2013). Sequencing was carried out on one lane of an Illumina HiSeq 2000 (100-bp paired end, plus 8-bp index read) by BGI Beijing (China).

In the 50 °C capture experiment, most reads from fresh samples did not pass the index-matching filter (>99%; Table S1, Supporting information), while the ancient and archival samples in the same capture reaction were not affected by this problem (Table S1, Supporting information). Therefore, capture of fresh samples at 50 °C was repeated without inclusion of archival and ancient samples.

NGS data processing

Sequences were trimmed using CUTADAPT v1.2.1 (Martin 2011). Forward and corresponding reverse reads were merged using FLASH v1.2.7 (Magoč & Salzberg 2011). All sequences for which the P5- and P7-indices did not match (allowing for one error in the P5-index) were discarded using CUTADAPT. Mapping of reads involved the BURROWS-WHEELER ALIGNER v6.2 (BWA; Li & Durbin 2009) with seeding disabled (Schubert *et al.* 2012), SAMTOOLS v0.1.18 (Li *et al.* 2009) and GATK v.2.8 (http://www.broadinstitute.org/gatk/).

For some analyses, we chose to deduplicate sequences before mapping to a reference, as we wanted to avoid the effect of clonality (PCR duplicates) when comparing the number of on-target sequences between hybridization temperatures. TALLY v0.13-231 (http://www.ebi.ac.uk/~stijn/reaper/tally.html) was used to deduplicate sequences; then, an equal subset of sequences (10 000) was used per sample for mapping. The '10k subsample' per sample was generated using SEQTK v.1.0-r57 (https://github.com/lh3/seqtk), to equalize the sample sets for comparison. Two of 25 samples (ancient lion and leopard samples) did not yield any sequences on target in the 10k subsamples (Table 2) and were therefore excluded from further analysis. For additional information on NGS data processing, see Appendix S1.

DNASP v5.10 (Librado & Rozas 2009) was used to retrieve the pairwise identity between the target species and the bait sequence in 60-bp sliding windows, with a 10-bp offset, in order to investigate the regional effect of bait–target sequence divergence. BEDTOOLS v.2.17.0 (Quinlan & Hall 2010) was used to retrieve the read depth in sliding window intervals.

Ancestral mapping strategy

As interspecific comparisons featured heavily in our analyses, it was important to choose a reference sequence that would limit the introduction of a bias during the mapping of sequences ('mapping bias'). Mapping algorithms become less effective when the reference is very dissimilar (Prüfer *et al.* 2010; Schubert *et al.* 2012) impacting the number of sequences identified as 'on-target'. Although *de novo* assembly is not subject to such a mapping bias, it is not suitable for short sequences, such as those recovered from archival and ancient samples (Schubert *et al.* 2012). As some of the species in this study currently do not have a reference sequence available from GenBank (Table 1), we reconstructed an ancestral mitogenome of all Felidae (Appendix S1; Table S2, Supporting information) to be used as reference. This strategy has several benefits: while substitution rates may differ between lineages, the ancestral sequence has on average equal evolutionary distances to all target sequences (Fig. 1a). This means that the mapping bias between species will be less severe than in alternative mapping strategies. Furthermore, in some cases, the (expected) divergence between the closest available GenBank reference and the target species may be greater than the divergence to the ancestral sequence – for example, the mitogenome of any extant felid is more divergent from *Homotherium serum* than the ancestral sequence

Table 2 Percentage of sequences that map to the ancestral mitogenome sequence; from raw sequences and deduplicated 10k subsample, as well as the portion of the mitogenome covered with ≥1× and ≥15× sequence depth. The data from the repeated experiment are indicated with grey shading

	% On-target reads in raw reads				% On-target reads in Tally'd 10k reads				% Mitogenome ≥1× from 10k subsample				% Mitogenome ≥15× from 10k subsample			
Sample	Shotgun	Touchdown	65 °C	50 °C	Shotgun	Touchdown	65 °C	50 °C	Shotgun	Touchdown	65 °C	50 °C	Shotgun	Touchdown	65 °C	50 °C
AJU_001	0.04	58.45	50.40	*61.80*	0.05	41.61	21.52	*51.98*	2.02	45.13	48.61	*68.70*	0.0	1.1	4.3	*29.8*
CTE_28	0.41	86.88	80.86	*70.23*	0.18	74.37	52.90	*43.15*	4.19	67.75	69.54	*73.13*	0.0	28.9	27.2	*32.6*
FMA_313	0.48	85.38	81.80	*76.44*	0.24	71.59	50.32	*60.26*	5.77	69.09	67.05	*73.70*	0.0	26.2	21.7	*30.0*
HSE_1007†	0.00	0.12	0.20	0.06	0.00	0.15	0.28	0.05	0.00	2.38	2.42	0.70	0.0	0.0	0.0	0.0
HSE_1714†	0.00	3.31	8.55	0.86	0.01	2.27	6.23	0.53	0.27	10.29	11.29	5.38	0.0	0.0	0.0	0.0
LLY_2P01	0.11	86.56	86.69	*73.94*	0.07	64.56	56.65	*51.17*	1.94	77.27	79.21	*84.64*	0.0	30.2	31.8	*38.6*
LP_534	1.12	86.90	86.11	*81.83*	0.54	71.00	59.43	*57.13*	10.10	64.22	66.27	*72.63*	0.0	28.3	25.7	*34.1*
LPA_placenta	0.13	79.59	45.14	*40.29*	0.03	60.73	55.22	*17.07*	0.66	70.56	69.14	*61.34*	0.0	23.2	25.8	*8.1*
LRU_1303	3.38	92.57	91.62	*82.77*	2.09	81.71	73.55	*58.11*	32.42	80.23	77.57	*84.35*	0.0	40.0	36.1	*44.2*
LSE_Togo	0.01	16.25	14.21	*20.74*	0.01	6.55	2.71	*14.49*	0.24	19.50	18.44	*36.02*	0.0	0.8	0.6	*6.4*
NNE_Fell	0.01	41.16	35.82	*31.44*	0.01	23.15	9.71	*21.67*	0.32	29.82	29.71	*40.63*	0.0	2.8	2.2	*8.8*
OMA_483	0.59	86.08	87.70	*89.53*	0.28	77.46	63.16	*83.65*	6.67	63.49	71.96	*73.74*	0.0	13.5	27.0	*35.1*
PBE_55*	0.16	70.71	76.57	57.01	0.10	51.03	69.53	24.92	2.75	52.65	50.30	79.17	0.0	17.1	18.4	33.9
PBE_83	3.58	91.62	90.32	*83.60*	1.80	81.62	72.77	*54.16*	29.57	74.76	71.24	*69.78*	0.0	33.5	31.5	*31.2*
PCO_526	0.08	69.90	69.38	*73.90*	0.11	55.17	35.48	*61.58*	2.46	53.14	61.48	*44.60*	0.0	1.7	13.5	*10.1*
PLE_001	0.32	82.83	82.74	*79.43*	0.16	66.86	54.28	*62.80*	3.68	64.08	61.79	*69.24*	0.0	20.7	21.1	*29.1*
PLE_1684†	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0
PO_001	0.13	78.30	74.36	*73.77*	0.13	57.00	38.49	*62.73*	2.85	61.80	61.31	*70.34*	0.0	16.0	17.1	*26.5*
PPA_2177†	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0
PPA_32	0.02	69.51	71.51	*66.53*	0.00	47.24	36.61	*52.33*	0.00	46.66	50.52	*56.47*	0.0	4.2	6.2	*17.7*
PPL_1*	0.07	54.56	62.72	37.04	0.05	33.94	53.74	12.34	1.06	46.42	46.49	37.96	0.0	10.5	12.4	4.2
PRU_402	0.06	67.98	48.89	*57.60*	0.06	48.52	16.29	*48.31*	1.70	56.79	52.06	*70.40*	0.0	4.8	6.6	*25.1*
PTS_001	0.27	80.11	76.78	*78.52*	0.19	62.89	43.16	*60.14*	4.64	60.66	59.34	*66.72*	0.0	19.7	17.1	*24.0*
PU_001	0.10	70.31	67.78	*51.08*	0.07	51.42	32.96	*37.21*	1.28	51.38	50.77	*59.54*	0.0	10.4	13.5	*17.8*
PVI_62	0.37	83.51	84.18	*78.65*	0.25	69.48	54.93	*63.90*	5.38	65.08	62.92	*72.07*	0.0	25.1	22.3	*32.7*
Average	0.46	61.70	58.97	54.68	0.26	48.01	38.40	39.99	4.80	49.33	49.58	54.85	0.0	14.3	15.3	20.8

*Archival samples.
†Ancient samples.
Repeated 50 °C capture indicated in italics.

(Fig. 1a). Where comparisons were conducted among multiple species, we used the ancestral mitogenome sequence as reference to equalize mapping bias for each species. In all other cases, the intraspecific or most closely related reference sequence was used (Table 1). Reconstruction of the ancestral sequence was performed using ANCESTORS v1.1 (Diallo *et al.* 2007, 2010; see Appendix S1, Supporting information). The resulting ancestral mitogenome has been made available on Dryad (http://dx.doi.org/10.5061/dryad.cd711). To further maximize identification of on-target sequences, we performed iterative mapping, which is efficient in recovering divergent sequences (Tsai *et al.* 2010; Hahn *et al.* 2013).

Effect of temperature on capture performance

We assess capture success of the three hybridization temperatures in terms of *enrichment specificity* and *target coverage*. We define enrichment specificity as the percentage of on-target sequences following capture. The target coverage is defined as the percentage of the target – the mitogenome – covered at a particular sequencing depth (≥1× and ≥15×).

We also evaluate the impact of sequence divergence between bait and target on capture, which we term *mismatch tolerance.* The mismatch tolerance was determined by correcting read depth (following capture) at a given level of divergence by the total number of sequences per sample. The analysis was carried out at a regional scale using sliding windows (60 bp long, 10-bp offset); this window size reflects both the length of baits printed on the Agilent SureSelect™ arrays and is a relatively close approximation of average library insert sizes. Too large or too small windows relative to median fragment length would result in scoring too many or too few mismatches per sequence-identity bin/category. We performed the sliding window analysis only for fresh samples of species for which a mitogenome reference sequence was available on GenBank (10 species in total), as this enabled us to calculate the pairwise similarity between bait and target even in the absence of coverage (for detailed methods, see Appendix S1, Supporting information). We constrained the analysis to a bait–target sequence divergence range of 0–25% (totalling 99.6% of data), as all species had at least five 60-bp windows within this range (i.e. a minimum of five windows at 25% sequence divergence for each species).

Statistical analysis

Statistical analyses were carried out in the statistical programming environment R (http://www.cran.r-project.org/). The influence of temperature and sample type upon enrichment specificity and target coverage was tested using linear mixed-effects models fitted and assessed with the package SPAMM v1.3.0 (Rousset & Ferdy 2014). A Gaussian error structure for fixed effects (temperature treatment, sample type and their interaction, as categorical parameters) was considered; counts defining the dependent variable were large enough for this approximation to be correct. Dependency between the different temperature treatments applied on the same samples was modelled by a Gaussian random effect. The overall effect of the temperature treatment was tested by comparing the full model described above with one considering only the sample type as fixed effect. When the interaction between temperature treatment and sample type was significant, models were rerun separately on each sample type considering only the temperature treatment as a fixed effect. For the study of mismatch tolerance, we also ran a linear mixed effect model considering temperature treatment (categorical), level of divergence (continuous) and their interaction as fixed effects. A Gaussian error was also considered after log-transforming the dependent variable to ensure normality. Dependency between the different measurements applied on the same samples was again modelled by a Gaussian random effect. All linear models were adjusted by maximum likelihood using the function *HLfit* (Hlmethod = ML) and fixed effects were tested using the function *fixedLRT* which computed the Bartlett correction of the likelihood ratio statistic from 1000 bootstrap replicates.

Results and discussion

The flexibility of hybridization capture – that is its tolerance for varying levels of sequence similarity between bait and target – allows the implementation of cross-species capture. The method we applied here has resulted in successful enrichment of the mitochondrial genome of species with a divergence time from the bait species of up to 20 million years (Johnson 2006; Barnett *et al.* 2005; Table 2; Table S1, Supporting information).

The original on-target content of precapture libraries, represented by the shotgun data, was found to significantly (positively) impact read depth following enrichment (rho = 0.29, $P < 0.001$). Portions of the target that were absent in the shotgun library were also not present in the captured library (Fig. S1, Supporting information), confirming that the enrichment data provided a good representation of the template pool available in the original library.

Mismatch tolerance

Previous studies have examined the impact of bait–target similarity at a per-target scale (e.g. per exon;

Vallender 2011; Bi *et al.* 2012; Peñalba *et al.* 2014) and reported that target coverage markedly decreases when divergence between bait and target reaches above 4–10%. For the standard hybridization temperature of 65 °C (Hodges *et al.* 2009), our regional scale analysis using sliding windows shows that read depth following enrichment gradually decreases with increasing sequence divergence from the bait (Spearman's correlation test, rho = −0.82, $P < 0.001$, Fig. 2a). Despite the reduced enrichment in regions of higher divergence, windows at the most extreme value considered (25% divergence) still displayed a considerable increase in on-target content in comparison with the shotgun libraries (mean 599-fold increase for capture at 65 °C; ranging from 20- to 2488-fold). Successful enrichment of regions of high divergence (up to 40% divergence between bait and target) has also been reported for nuclear DNA (Li *et al.* 2013). In more divergent species, low divergence sequences (windows with ~0–3% divergence) are preferentially recovered compared to high divergence windows sequences (windows with ~10–25% divergence; Fig. 2a). This effect is exemplified

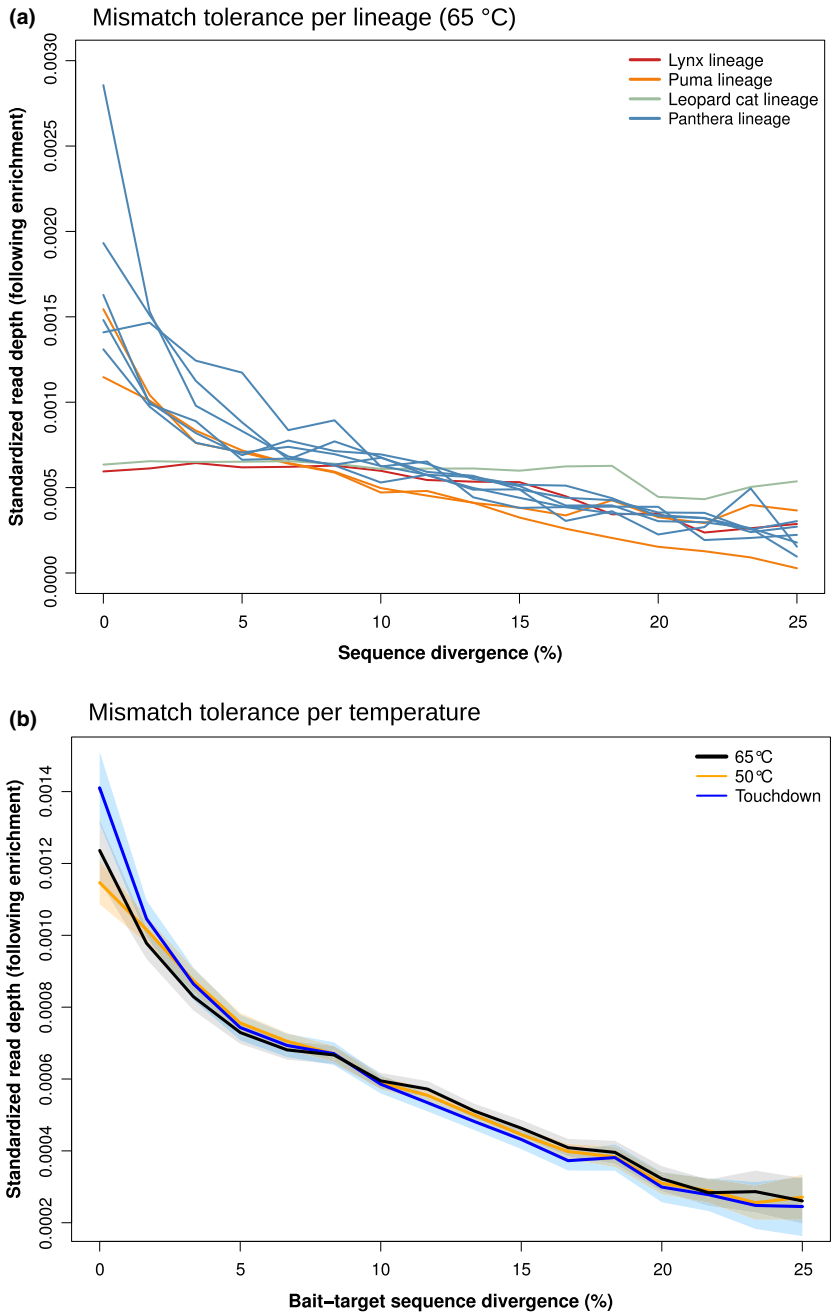

Fig. 2 Impact of bait–target sequence divergence on enrichment, for 10 felid species [phylogenetic lineages according to Johnson (2006)] are indicated by colour in (a). Mean read depth for 60-bp sliding windows at a given level of divergence was determined per sample and then standardized to account for differences in total number of sequences per sample. Data are presented for serial cross-species capture for each sample at the standard 65 °C hybridization temperature (a) and for each hybridization temperature averaged across all samples (b). Shaded areas represent the 95% confidence interval for each temperature.

in more detail for two species: a close relative of the bait species (the bobcat *Lynx rufus*) and a more distantly related species (the snow leopard *Panthera uncia*, Fig. S1, Supporting information).

Modifying the stringency of the hybridization process in a way that more mismatches are permitted between bait and target would benefit enrichment of more divergent fragments and thus be of particular interest for cross-species capture. Therefore, we examined whether hybridization temperature impacted the retrieval of more divergent sequences by investigating the mismatch tolerance between bait and target at the three different hybridization temperatures (Fig. 2b). Our analysis did not provide evidence that a lower hybridization temperature significantly improves the retrieval of highly divergent fragments (test of the interaction between temperature and divergence: Bartlett-corrected LR statistic = 1.85, d.f. = 2, $P = 0.40$). Moreover, there was no significant effect of the temperature treatment, irrespectively of the level of divergence (test of the temperature effect: Bartlett-corrected LR statistic = 0.28, d.f. = 2, $P = 0.87$), but as expected the recovered read depth dropped with divergence (test of the divergence effect: Bartlett-corrected LR statistic = 550, d.f. = 1, $P < 0.0001$). To investigate whether mismatch tolerance was affected by sequences in flanking windows, we performed the analysis on a subset of the data incorporating every 20th window along the mitogenome (windows spaced at least 140 bp apart). The observed relationship between the sequencing depth and sequence divergence remains the same (Fig. S2, Supporting information).

There are several possible explanations for why we find no impact of hybridization temperature on mismatch tolerance: (i) the range of sequence divergence in our study was not large enough; (ii) the range of hybridization temperatures applied (50–65 °C) may not have been broad enough; or (iii) other parameters are more crucial for the recovery of divergent fragments than hybridization temperatures (Dabney *et al.* 2013; Li *et al.* 2013; Meyer *et al.* 2013). Postcapture washing temperature is of particular interest for this aspect, as it is decisive for the stringency of washing away nontarget DNA, and may thus also impact the release of high divergence fragments from the baits.

Hybridization temperature: enrichment specificity

The hybridization temperatures tested in this study had a significant effect on enrichment specificity (i.e. the percentage of on-target sequences; Fig. 3). Importantly, the effect of hybridization temperature varied with sample type (i.e. fresh, archival, ancient; test of the interaction between temperature and sample type: all sequences, Bartlett-corrected LR statistic = 10.2, d.f. = 4, $P = 0.037$; 10k subsample, Bartlett-corrected LR statistic = 22.2, d.f. = 4, $P = 0.0002$). While for fresh samples, a higher proportion of on-target sequences was recovered using a touchdown capture approach, capture of degraded samples performed better at 65 °C. We investigated the effect of hybridization temperature for both the raw data (Fig. 3a–c) and the 10k subsample that was deduplicated prior to mapping (Fig. 3d–f). The raw data (i.e. all trimmed and merged sequences) reflect the overall success of the experiment as it is affected by many factors, including library preparation procedures and amplification, as well as capture performance. The data that have been deduplicated prior to mapping enable a comparison of unique on- and off-target sequences and so compensate for differential amplification between samples or treatments. This will provide a more accurate indicator of the impact of hybridization temperature on the capture of unique sequences.

Enrichment specificity is significantly higher for fresh samples at touchdown capture than at 65 °C, as well as the repeated 50 °C capture (raw sequences, Bartlett-corrected LR statistic = 11.9, d.f. = 2, $P = 0.0026$; 10k subsample, Bartlett-corrected LR statistic = 16, d.f. = 2, $P = 0.0003$). Our results are in agreement with previous findings of an improved performance found for in-solution nuclear capture of fresh samples when using a touchdown approach (Li *et al.* 2013). This suggests that the effect of hybridization temperature is consistent across different capture protocols and substrates (in-solution and on-array).

Contrary to the fresh samples, hybridization at a constant temperature of 65 °C appears to perform best for degraded samples (Fig. 3b,e); data for raw and deduplicated (10k) sequences are in agreement. However, due to the large variance among ancient samples as well as the small sample size, the difference between temperature treatments was only significant for archival samples (raw sequences, Bartlett-corrected LR statistic = 7.17, d.f. = 2, $P = 0.028$; 10k subsample, Bartlett-corrected LR statistic = 9.98, d.f. = 2, $P = 0.0068$). When looking at the samples individually (Fig. S3, Supporting information), it is clear that consistently more sequences were retrieved for archival as well as ancient samples at 65 °C. Our data demonstrate that hybridization temperature affects fresh samples differently than degraded, contamination-rich samples. It is conceivable that for samples with a high proportion of contaminant (bacterial) DNA, a lower hybridization temperature may lead to the inadvertent capture of exogenous DNA, even when there is only limited sequence similarity between the bait and target. Thus, hybridization at lower temperatures could potentially be counterproductive for contamination-rich sources such as DNA isolated from ancient and archival samples.

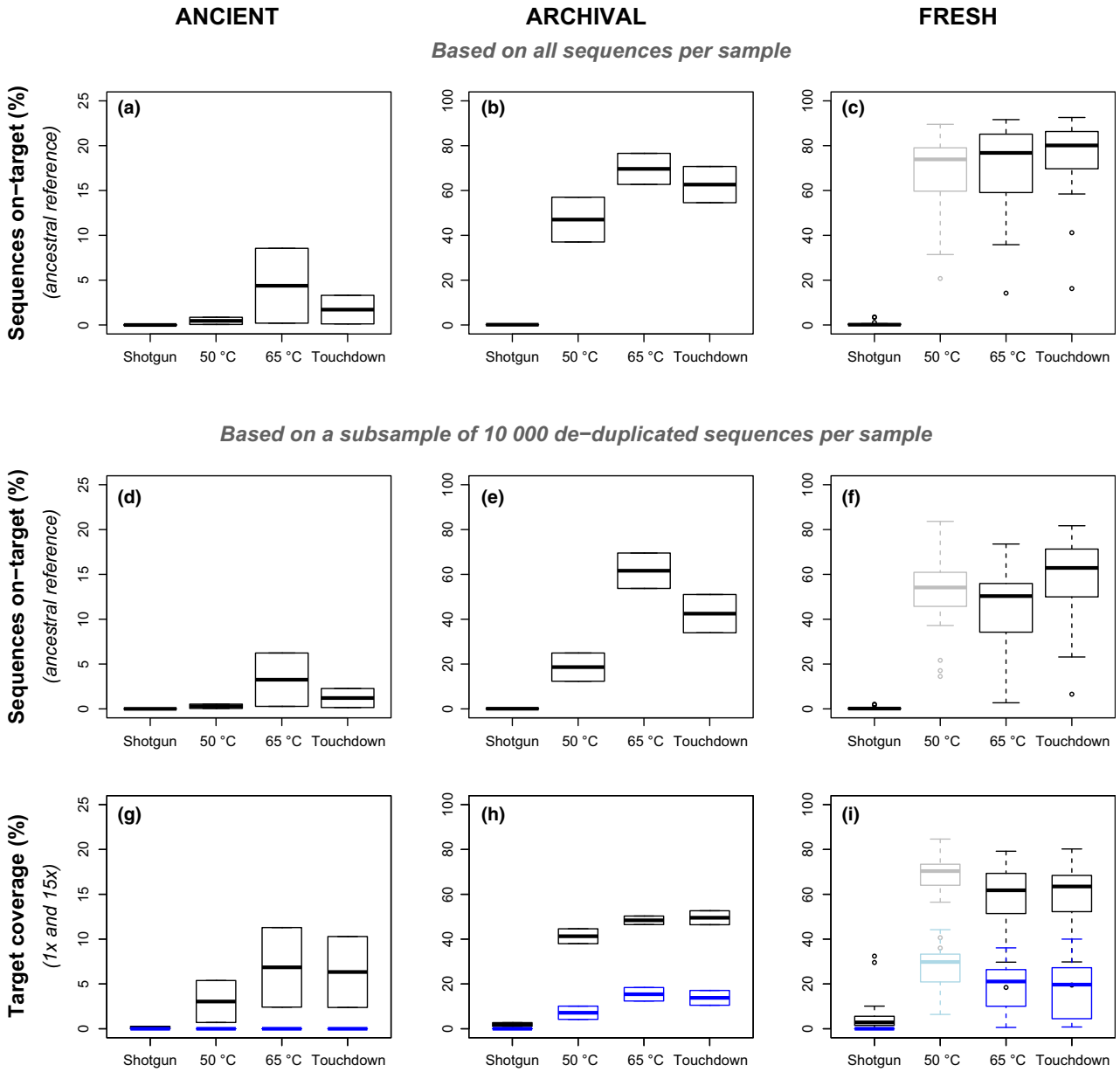

Fig. 3 Percentage of on-target sequences for the raw data (a, b, c) and deduplicated data (d, e, f), as well as the percentage of mitogenome covered (g, h, i) at different sequence depths (≥1× in black, ≥15× in blue) for the three different sample types (ancient: a, d, g; archival: b, e, h; fresh: c, f, i). Data from the repeated experiment (50 °C capture of fresh samples) are indicated with a light grey and light blue border (see Results and discussion). The boxes represent the range within which 75% of the variation falls, the bold bar the median value and whiskers extend to the furthest data point within a 1.5× interquartile range. Outliers are indicated with an open circle. Graphs for individual samples can be found in Figs S3 and S4 (Supporting information).

As hybridization temperature affects enrichment specificity differently for fresh and degraded (archival, ancient) samples, this should be taken into account during experimental design of studies involving hybridization capture. Choosing the most appropriate hybridization temperature to increase enrichment specificity is important, as it permits researchers to increase the number of samples on a sequencing run, while still achieving the same number of on-target sequences per individual.

Hybridization temperature: target coverage

Target coverage was measured at two sequence depths, ≥1× and ≥15×, as different NGS studies often have specific requirements regarding sequence depths. For

example, for the reconstruction of mitogenomes from degraded samples, relatively shallow coverage may be sufficient, while other applications such as calling heterozygote positions in diploid organisms generally require deeper coverage. Therefore, we measured the mitogenome coverage at shallow as well as deeper sequence coverage for each temperature.

For fresh samples, we find no significant difference in target coverage between 65 °C and touchdown (Fig. 3g–i; ≥1× coverage, Bartlett-corrected LR statistic = 0.21, d.f. = 1, $P = 0.65$; ≥15× coverage, Bartlett-corrected LR statistic = 0.86, d.f. = 1, $P = 0.35$; Supporting Fig. 4). The repeated 50 °C capture shows a significantly higher target coverage, both at ≥1× and at ≥15× (≥1× coverage, Bartlett-corrected LR statistic = 34.4, d.f. = 2, $P < 0.0001$; ≥15× coverage, Bartlett-corrected LR statistic = 20.9, d.f. = 2, $P < 0.0001$). This may be due to either the increase in starting template in this experiment (10× more template), or the lack of competing sample material (ancient and archival samples made up almost 96% of the library pool prior to enrichment in the original experiments; Fig. 1b). For the degraded samples, capture at 50 °C results in lower coverage than capture at the other two treatments. Due to the small ancient sample size and large variance in their performance, the differences between temperature treatments was only significant for archival samples (≥1× coverage, Bartlett-corrected LR statistic = 6.25, d.f. = 2, $P = 0.044$; ≥15× coverage, Bartlett-corrected LR statistic = 10.4, d.f. = 2, $P = 0.0055$). As for the fresh samples, there is no difference in coverage between 65 °C and touchdown temperatures (≥1× coverage, Bartlett-corrected LR statistic: 0.42, d.f. = 1, $P = 0.52$; ≥15× coverage, Bartlett-corrected LR statistic: 2.42, d.f. = 1, $P = 0.12$).

Target coverage is not only dependent on the number of retrieved sequences, but also on their average length. In our data, we observe a difference in mapped sequence length distribution following capture at different hybridization temperatures (Fig. 4). This difference in mapped sequence lengths likely affected target coverage. It should be noted that our experimental design was not aimed at addressing this particular question: (i) duplication removal tools introduce a bias towards longer fragments (Fig. 4a vs. b), and (ii) our libraries were amplified into the plateau with a polymerase that is known to bias towards shorter fragments (Dabney & Meyer 2012). The differences observed in mapped sequence length between hybridization temperatures

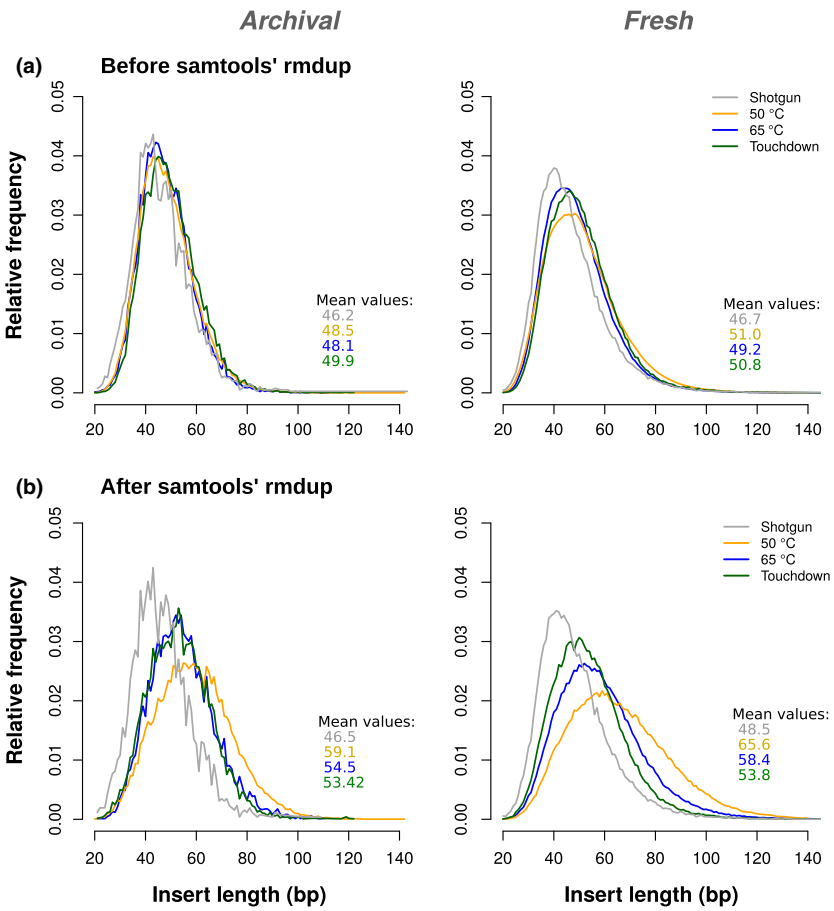

Fig. 4 Library insert length distributions following capture at the different hybridization temperatures and from shotgun sequencing for archival and fresh samples, based on the data before samtools' rmdup (a) and after (b). The ancient samples were excluded from this analysis, as there was not sufficient data to recover a reliable fragment length distribution.

may explain why there is a significant effect of temperature on enrichment specificity, but not on target coverage. The distribution of shotgun sequence length is consistently lower than for the enriched libraries. This is in agreement with previous studies comparing read length of pre- and postcapture libraries, indicating that while the enrichment of short (<40 bp) fragments is possible, longer fragments are preferentially recovered (Dabney *et al.* 2013). Our results emphasize the need for additional, structural studies into the effect of hybridization temperature on the length bias of hybridization capture.

We have further investigated target coverage by investigating the effect of hybridization temperatures on downstream analysis. For each temperature, phylogenetic inference was performed on the retrieved consensus sequences, and a phylogeny was generated for each hybridization temperature (Appendix S1, Supporting information). Due to their limited coverage of the mitogenome, we did not include the ancient samples in this analysis. The phylogenies were highly congruent; only the position of a single species (the serval: *Leptailurus serval*) differed between the phylogenies, but its position had low bootstrap support (<40%; Fig. S5, Supporting information). For this particular sample, we retrieved only a limited portion of the mitogenome (45%, 26% and 11% for 50 °C, 65 °C and touchdown, respectively; Table S3, Supporting information), which is likely to have caused the discrepancies between the 'temperature trees'. All other species had identical positions in all phylogenetic trees.

We found that libraries with low endogenous content (determined by shotgun sequencing) generally also displayed low on-target content and low mitogenome coverage in the enriched libraries (Table 2; Table S3, Supporting information). While this is not unexpected, it highlights the importance of selecting high-quality samples (i.e. high endogenous content) whenever possible. For ancient substrates, there have been a number of studies into endogenous DNA survival (Hofreiter *et al.* 2015). For archival and noninvasive sampling, similar research into sample quality – or the prediction thereof (Wales *et al.* 2012; Enk *et al.* 2013) – for high-throughput sequencing approaches is needed, as these sample types represent valuable sources of data in studies of extinct populations or elusive wildlife.

In conclusion, we successfully performed on-array cross-species capture across the carnivoran family Felidae using baits designed from a single felid species. We showed that hybridization temperature affects the number of sequences that are recovered. We provided evidence that the hybridization temperature at which capture will perform best is likely determined by sample type and quality, rather than by divergence between bait and target. We found that target enrichment from samples with large amounts of contamination will perform better at a higher hybridization temperature. Unexpectedly, our results show no significant improvement in the recovery of more divergent target sequences at any of the hybridization temperatures tested, suggesting that other factors (e.g. the posthybridization washing temperature) may be more critical for increasing the capture success of regions with higher sequence divergence.

Acknowledgements

This research was funded by the National Science Foundation (DEB award 1132229) and by the Leibniz Association (SAW-2013-IZW-2). We would like to acknowledge Kanchon Dasmahapatra (University of York, UK) for contributing scripts, Jelle Reumer, Jennifer Leonard, Hans-Peter Uerpmann, Robert Sommer, Marc Nussbaumer, Andreas Wilting, the American Museum of Natural History New York and the State Museum of Zoology Dresden for providing samples and Riddhi Patel (IZW Berlin, Germany) for providing sequences for *Prionailurus viverinus*, *Prionailurus planiceps*, *Catopuma temmickii* and *Catopuma badia* for the reconstruction of the ancestral sequence. We would like to thank Andreas Wilting and Axel Barlow, and four anonymous reviewers for helpful comments and discussion on the draft manuscript.

References

Ávila-Arcos MC, Cappellini E, Romero-Navarro JA *et al.* (2011) Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports*, **1**, 74.

Barnett R, Barnes I, Phillips MJ *et al.* (2005) Evolution of the extinct Sabretooths and the American cheetah-like cat. *Current Biology*, **15**, R589–R590.

Bi K, Vanderpool D, Singhal S *et al.* (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, **13**, 403.

Bodi K, Perera AG, Adams PS *et al.* (2013) Comparison of commercially available target enrichment methods for next-generation sequencing. *Journal of Biomolecular Techniques*, **24**, 73–86.

Carpenter ML, Buenrostro JD, Valdiosera C *et al.* (2013) Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *The American Journal of Human Genetics*, **93**, 852–864.

Castellano S, Parra G, Sánchez-Quinto FA *et al.* (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 6666–6671.

Dabney J, Meyer M (2012) Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, **52**, 87–94.

Dabney J, Knapp M, Glocke I *et al.* (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15758–15763.

Devault AM, Golding GB, Waglechner N *et al.* (2014) Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *New England Journal of Medicine*, **370**, 334–340.

Diallo AB, Makarenkov V, Blanchette M (2007) Exact and heuristic algorithms for the indel maximum likelihood problem. *Journal of Computational Biology*, **14**, 446–461.

Diallo AB, Makarenkov V, Blanchette M (2010) Ancestors 1.0: a web server for ancestral sequence reconstruction. *Bioinformatics*, **26**, 130–131.

Elhaik E, Greenspan E, Staats S *et al.* (2013) The GenoChip: a new tool for genetic anthropology. *Genome Biology and Evolution*, **5**, 1021–1031.

Enk JM, Rouillard J-M, Poinar H (2013) Quantitative PCR as a predictor of aligned ancient DNA read counts following targeted enrichment. *BioTechniques*, **55**, 300–309.

Enk JM, Devault AM, Kuch M *et al.* (2014) Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, **31**, 1292–1294.

Fortes GG, Paijmans JLA (2015) Analysis of whole mitogenomes from ancient samples. In: *Whole genome amplification. Methods in Molecular Biology* (ed. Kroneis T). Humana Press, New York, USA. Available as ArXiv pre-print: http://arxiv.org/abs/1503.05074

Gnirke A, Melnikov A, Maguire J *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, **27**, 182–189.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Research*, **41**, e129.

Hancock-Hanser BL, Frey A, Leslie MS *et al.* (2013) Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources*, **13**, 254–268.

Hodges E, Rooks M, Xuan Z *et al.* (2009) Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protocols*, **4**, 960–974.

Hodges E, Xuan Z, Balija V *et al.* (2007) Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, **39**, 1522–1527.

Hofreiter M, Paijmans JLA, Goodchild H *et al.* (2015) The future of ancient DNA: technical advances and conceptual shifts. *BioEssays*, **37**, 284–293.

Johnson WE (2006) The late Miocene radiation of modern Felidae: a genetic assessment. *Science*, **311**, 73–77.

Kircher M, Sawyer S, Meyer M (2011) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, **40**, e3.

Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA (2012) Setting the stage – building and working in an ancient DNA laboratory. *Annals of Anatomy - Anatomischer Anzeiger*, **194**, 3–6.

Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJ (2013) Capturing protein-coding genes across highly divergent species. *BioTechniques*, **54**, 321–326.

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)*, **27**, 2957–2963.

Mamanova L, Coffey AJ, Scott CE *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7**, 111–118.

Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, **5**, e14004.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, **17**, 10–12.

Mason VC, Li G, Helgen KM, Murphy WJ (2011) Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research*, **21**, 1695–1704.

Meyer M, Fu Q, Aximu-Petri A *et al.* (2013) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*, **505**, 403–406.

Mitchell KJ, Llamas B, Soubrier J *et al.* (2014) Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science*, **344**, 898–900.

Paijmans JLA, Gilbert MTP, Hofreiter M (2013) Mitogenomic analyses from ancient DNA. *Molecular Phylogenetics and Evolution*, **69**, 404–416.

Peñalba JV, Smith LL, Tonione MA *et al.* (2014) Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources*, **14**, 1000–1010.

Prüfer K, Stenzel U, Hofreiter M *et al.* (2010) Computational challenges in the analysis of ancient DNA. *Genome Biology*, **11**, R47.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Rizzi E, Lari M, Gigli E, Bellis GD, Caramelli D (2012) Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution*, **44**, 1–19.

Rohland N, Siedel H, Hofreiter M (2004) Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *BioTechniques*, **36**, 814–820.

Rohland N, Siedel H, Hofreiter M (2010) A rapid column-based ancient DNA extraction method for increased sample throughput. *Molecular Ecology Resources*, **10**, 677–683.

Rousset F, Ferdy J-B (2014) Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*, **37**, 781–790.

Schubert M, Ginolhac A, Lindgreen S *et al.* (2012) Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, **13**, 178.

Templeton JEL, Brotherton PM, Llamas B *et al.* (2013) DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investigative Genetics*, **4**, 1–13.

Tsai IJ, Otto TD, Berriman M (2010) Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biology*, **11**, R41.

Vallender EJ (2011) Expanding whole exome resequencing into non-human primates. *Genome Biology*, **12**, R87.

Wales N, Romero-Navarro JA, Cappellini E, Gilbert MTP (2012) Choosing the best plant for the job: a cost-effective assay to prescreen ancient plant remains destined for shotgun sequencing (C Lalueza-Fox, Ed,). *PLoS One*, **7**, e45644.

J.L.A.P., D.W.F., J.F. and M.H. conceived the project. J.L.A.P. and D.W.F. designed the experiment. J.L.A.P. and D.W.F. performed the laboratory work. J.L.A.P., D.W.F. and A.C. performed the analyses. J.L.A.P. and D.W.F. wrote the manuscript with input from all authors.

Data accessibility

Sequences are deposited in GenBank under the following Accession nos.: *Lynx lynx* mitogenome (KM982549), consensus sequences for each sample (KR132579–KR132597). Raw Illumina sequence reads are deposited in the NCBI SRA under Accession no. SRP057047. The reconstructed ancestral mitogenome sequence, tree files for the phylogenies displayed in Fig. S5 and the Perl

scripts are deposited on Dryad (http://dx.doi.org/10.5061/dryad.cd711).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Impact of original on-target content and bait-target sequence divergence on enrichment visualized along the mitogenome for two species: a close relative of the bait species (the bobcat *Lynx rufus*, A) and a more distantly related species (the snow leopard *Panthera uncia*, B).

Fig. S2 Impact of bait-target sequence divergence on enrichment of every 20th window, spacing the windows approximately 140 bp apart.

Fig. S3 Histograms displaying the percentage of on-target sequences recovered from the de-duplicated 10k subsample for each individual.

Fig. S4 Histograms displaying the percentage of mitogenome retrieved with ≥1× coverage (A) and with ≥15x (B), based on the de-duplicated 10k dataset.

Fig. S5 Maximum likelihood trees for all temperatures; sequences from GenBank are labeled in grey, sequences new from this study are labeled in black.

Table S1 Number of on-target sequences for each individual, sorted according to hybridization experiment.

Table S2 Species included in the reconstruction of the ancestral sequence, with GenBank accession numbers.

Table S3 Mitogenome coverage (≥3x) in the final consensus sequences used to generate the maximum likelihood trees in Fig. S5.

Table S4 Primer sequences for obtaining the Eurasian lynx (*Lynx lynx*) mitogenome sequence (long-range PCR and sequencing primers).

Appendix S1 Detailed methods and protocols.