

Supplement:

The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*)

Webb Miller^{1,10}, Daniela I. Drautz¹, Jan E. Janecka², Arthur M. Lesk¹, Aakrosh Ratan¹, Lynn P. Tomsho¹, Mike Packard¹, Yeting Zhang¹, Lindsay R. McClellan¹, Ji Qi¹, Fangqing Zhao¹, M. Thomas P. Gilbert³, Love Dalén⁴, Juan Luis Arsuaga⁵, Per G. P. Ericson⁶, Daniel H. Huson⁷, Kristofer M. Helgen⁸, William J. Murphy², Anders Götherström⁹, Stephan C. Schuster^{1,10}

¹Pennsylvania State University, Center for Comparative Genomics and Bioinformatics, 310 Wartik Building, University Park, PA 16802, USA.

²Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843

³Centre for Ancient Genetics, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark

⁴School of Biological Sciences, Royal Holloway, University of London, Egham TW20 0EX. United Kingdom

⁵Centro Mixto UCM-ISCIH de Evolución y Comportamiento Humanos, c/Sinesio Delgado 4 Pabellon 14, 28029 d, Spain

⁶Department of Vertebrate Zoology, Swedish Museum of Natural History, P.O. Box 50007, S-10405, Stockholm, Sweden

⁷Center for Bioinformatics Tübingen, Tübingen University, Tübingen, Germany

⁸Division of Mammals, National Museum of Natural History, Smithsonian Institution, MRC 108, P.O. Box 37012, Washington, D.C. 20013-7012, USA

⁹Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyv. 18D, S-752 36 Uppsala, Sweden

¹⁰Corresponding authors. Email: webb@bx.psu.edu, scs@bx.psu.edu; fax (814) 863-6699.

Contents

Plans to clone the Tasmanian tiger	3
Thylacine specimens	3
Fig. S1. Painting of living thylacine 2	5
Fig. S2. Photograph of living thylacine 2	5
Numbat samples	6
Assembling the thylacine and numbat mitochondrial sequences	6
Measuring human contamination	6
Ratio of nuclear to mitochondrial DNA	6
Notes on the cytochrome b amino-acid sequences	7
Table S1. Differences between our new assembly of cytochrome b and the one from GenBank	8
Damage	9
Table S2. Observed rate of cytosine hydrolytic deamination in the modern numbat, ethanol-preserved thylacine, and mammoth samples of comparable thermal ages	10
Fraction mtDNA / Human contamination	10
Fig. S3. Length distributions of various classes of sequencer reads from the ethanol-stored thylacine specimen 2	11
Table S3. Percentages of endogenous mtDNA in three samples, and percentages of human contamination in thylacine sample 2	11
Interspersed repeats in the thylacine	11
Table S4. Summary of the repeat content in the metagenome of Thylacine 2	12
Metagenomics	12
Fig. S4.A. MEGAN taxa tree of thylacine specimen 1	14
Fig. S4.B. MEGAN taxa tree of thylacine specimen 2	15
Table S5. Top 10 abundant bacterial species in two thylacine metagenomic samples	16
Fig. S5. (A) Microbial attributes of the assigned bacterial reads for the numbat hair and (B) the thylacine 2 hair biomes	17
Fig. S6 A. Viromes of thylacine 2 and numbat liver	18
Supplemental References (other references are in the main paper):	19

Plans to clone the Tasmanian tiger. Movies of thylacines in the Hobart and London zoos can be seen at:

http://www.naturalworlds.org/thylacine/films/motion_film_footage.htm

There are still occasional unconfirmed reports of thylacine sighting in the Tasmanian forests:

http://www.livescience.com/animals/070703_tasmanian_tiger.html

However, few if any scientists believe there is any chance that the species survives. A 2006 story reports the existence of 714 known (dead) thylacine specimens:

http://www.naturalworlds.org/thylacine/morefeatures/itsd/itsd_1.htm

Plans to clone the thylacine were widely discussed in the media, starting around 1999:

<http://news.bbc.co.uk/2/hi/science/nature/343702.stm>

http://www.naturalworlds.org/thylacine/additional/cloning/cloning_1.htm

<http://www.wired.com/science/discoveries/news/2002/06/52959>

Attempts by the Australian museum to create a thylacine genomic library were unsuccessful:

<http://www.amonline.net.au/thylacine/summary.htm>

<http://www.abc.net.au/science/news/stories/s1302459.htm>

A news story in May 2005 reported that the project had been picked up by another group:

<http://www.smh.com.au/news/Science/Clone-again/2005/05/14/1116024405941.html>

Thylacine specimens. At the close of the nineteenth century, no American zoo had found itself able to display the largest marsupial predator of modern times, the thylacine (*Thylacinus cynocephalus*), and concern about the animal's increasing rarity in its native Tasmania was mounting (Guiler 1985). At the urging of the National Zoo in Washington D.C., an American official stationed in Australia finally arranged for a wild-captured thylacine to be shipped to the United States in early 1902 (Wemmer 2002). This female thylacine was pregnant when captured, and soon after, in April, gave birth to three tiny, naked pouch young. After months of crated transport, first by boat across the Pacific and then by train across the American continent (Wemmer 2002), the thylacine mother arrived at the zoo on September 3 with her three offspring, by now furred and the size of rabbits (Fig. S1). Ill from the long journey, one of the cubs died nine days later. The mother lived for two years at the zoo, and her two remaining offspring, male and female, survived into adulthood.

This particular thylacine family has loomed large in the study of the species, both in life and death. The longer surviving offspring, subjects of one of the most famous photographs of the species (Fig. S2), provided a very rare example of thylacine pouch

young surviving to maturity in a zoo (Guiler 1985). When the zoo obtained another, adult male thylacine from Tasmania in 1904, the surviving daughter was the subject of one of few attempts at captive breeding, ultimately unsuccessful (Paddle 2000). All of these thylacines are today preserved as specimens in the United States National Museum of Natural History (USNM), Smithsonian Institution, where they have been studied by several generations of comparative morphologists. The mother, on display at USNM as a taxidermy mount since 1904, can still be viewed in the museum's public galleries, where it has been seen by tens of millions of visitors. Notably, the accompanying skeleton of this specimen has been the subject of previous efforts to obtain thylacine molecular sequence data using aDNA protocols (Krajewski et al. 1992, 1997, 2000) and the USNM thylacines provide a partial focus for our study as well.

The first Smithsonian specimen we investigated was the mother, USNM 124662 (skin) / 49723 (skull), adult female, which died at the National Zoo in 1904. We sampled approximately < 0.1 grams of hair from the left side and the hair tufts between several toes, without appreciable damage to the skin. The right side of the skin was left fully intact. This is the specimen that had been sequenced by Krajewski et al. (1997), but this time DNA recovery was unsuccessful.

The second specimen was the offspring (son) of the first specimen. The same mitochondrial sequence is of course to be expected. This specimen was USNM number 125345 (skin) / 49724 (skull), adult male, which died at the National Zoo in 1905. The skin of this specimen is prepared as a scientific study skin (Fig. 1 on the main paper); it is housed in the USNM scientific collections and has not been on display since its death. We sampled approximately 0.01 gram of hair from the right side of the skin and from the hair tufts between the toes on the right hindfoot, without appreciable damage to the skin. The left side of the skin was left fully intact.

The Swedish thylacine specimen in ethanol is NRM 566599 in the mammal collection of the Swedish Museum of Natural History (Stockholm), adult female, dated 30 April 1893 and received from the London Zoo. This may be the adult female that died in the London Zoo in April 1893 that was discussed by Flower (1931). We also made a first attempt to obtain sequence from the mounted skin of an adult male thylacine specimen in the Swedish Museum that was collected in 1870 (NRM 592206), but were unsuccessful.



Fig. S1. A 1902 painting by Joseph Gleeson depicts a thylacine mother with three cubs (including one in the pouch) as they appeared on arrival at the National Zoo (courtesy Smithsonian Archives). Our study sampled hairs from the mother and one of the cubs.

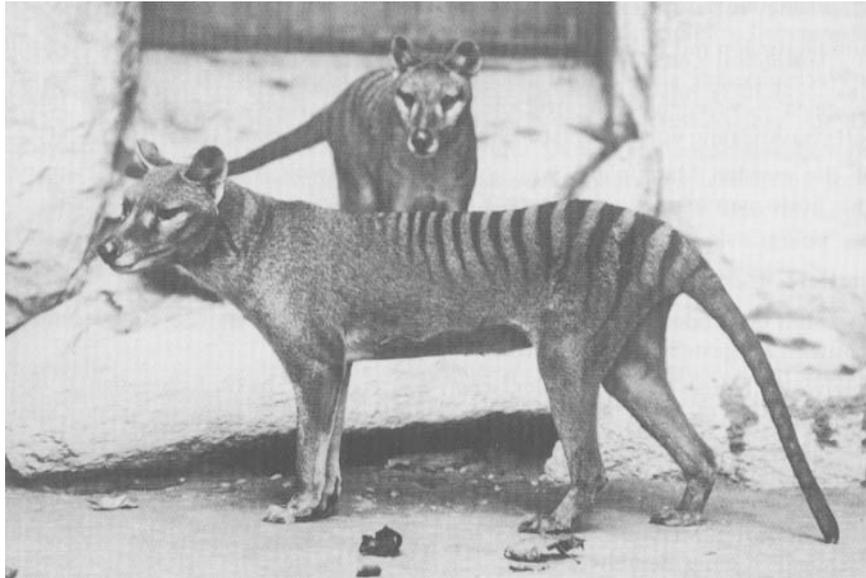


Fig. S2. Among the highest quality of known thylacine photographs, this portrait apparently depicts the two surviving offspring (see Fig. S1) two to three years after their arrival at the National Zoo. One of these animals (more likely in front) is USNM 125345, a specimen sequenced for this study (courtesy Smithsonian Archives).

Numbat samples. The numbat was DEC Donnelly District specimen number M108. It was collected on November 28, 2004 from 700 meters north of Eclipse Road on Corbalup Road, within the Corbal forest block in the Greater Kingston National Park (approx. coordinates: 34° 08' 15" S 116° 27' 30 E). This is approximately 30km northeast of Manjimup in Western Australia. The specimen has been stored at -20 °C since its collection. The sex of the animal was not recorded. Samples were put in 70% ethanol for transit.

Assembling the thylacine and numbat mitochondrial sequences. Sequence reads aligning to the mitochondrial genome of the northern quoll (*Dasyurus hallucatus*; GenBank Accession AY795973.1) were filtered and trimmed to require at least 70% alignment coverage and 60% identity, then assembled. Subsequently, all reads were aligned to the assembled contigs at higher stringency (80% coverage and at least 95% identity) and the hits were reassembled. The sequence coverage at each position averaged 49.7-fold for thylacine, 41.2-fold for the numbat liver sample, and 206.5-fold for numbat hair-shafts (using only one-quarter of a FLX run). Several segments of the thylacine mtDNA genome, including the entire cytochrome b gene, were resequenced by PCR amplification and Sanger sequencing.

Sequences generated in this study have been deposited in GenBank. The thylacine mitochondrial genomes have Accession Numbers 000000000 and 000000000, while the numbat mitochondrial genome is 000000000.

Measuring human contamination. We aligned the reads to the March 2006 assembly of the human genome (chromosomes 1-22, X and Y) and to the human mitochondrial genome (GenBank entry NC_001807), and required that they align with at least 97% identity and over at least 90% of the length of the read.

Ratio of nuclear to mitochondrial DNA. To estimate the nuDNA:mtDNA ratio of human contamination in the thylacine samples, we simply divided the number of putative human nuDNA reads by the number of putative human mtDNA reads. For the endogenous ratio nuDNA:mtDNA in each of the numbat samples, we subtracted the number of identified mtDNA reads from the total number of reads, then divided that difference by the number of identified mtDNA reads; this ratio will be an overestimation if many of the reads are non-numbat DNA. For the thylacine sample 2, we divided one third of the total number of reads from the sample by the number of putative thylacine mtDNA reads, based on our estimation that roughly one third of the reads are from the thylacine nuclear genome.

At first glance the relative number of nuDNA to mtDNA contaminant sequences in the thylacine dataset seems rather high. However, this relative ratio is likely a result of the innate ratio of total mtDNA to nuDNA sequences in the living tissue that was the source of contamination. (For example, previous studies on bone have demonstrated a ratio of approx 650 nuDNA sequences to every mtDNA sequence, while studies on mammoth hair yield ratios as low as 50 nuDNA sequences to every mtDNA sequences

(Poinar et al. 2006, Gilbert et al. 2007a, Gilbert et al. 2007b), and the numbat hair sample has a nuDNA:mtDNA ratio of roughly 2.3:1.

Notes on the cytochrome b amino-acid sequences. The cytochrome b amino acid sequences from our assembly and from GenBank entry THNMTCYTB are both 380 residues long. They differ at 33 positions (91% identical). Almost all of the differences are conservative, i.e., consist of a pair of amino acids that frequently replace one another in functional orthologous proteins. We modeled the three-dimensional structure of both proteins using SWISS-MODEL, and observed that the changes at buried positions are conservative. Even though it appears to be a functional molecule, our best guess is that it has a numt origin.

Almost all of the amino acid sequence changes are what would be regarded as conservative mutations. These comprise the following pairs:

L ↔ M (6 positions)

I ↔ L (5 positions)

T ↔ A (3 positions)

V ↔ A (2 positions)

I ↔ V (4 positions)

S ↔ A

I ↔ T

N ↔ D

F ↔ L

S ↔ I

T ↔ L

A few are not conservative:

K ↔ I (charge change)

L ↔ R (charge change)

I ↔ A (moderate size change)

F ↔ I (moderate size change)

F ↔ S (large size change)

I ↔ P

S ↔ P

Note that this suggests that we are looking at two related real sequences, rather than a single sequence corrupted by random noise (which, despite the internal structure of the genetic code, would be expected to have more non-conservative mutations).

Table S1. Differences between our new assembly of cytochrome b and the one from GenBank.

Pos	New	GB	majority	comment
2	K	I		I is not the majority residue but it appears in two other sequences, and the majority residue is similar in size and charge to I; R (as in cytbNew) is a very unusual residue at this position
4	L	M	L	similar sidechains, M appears elsewhere
10	I	L	L	similar sidechains, I appears elsewhere
24	T	A	A	similar sidechains, T appears elsewhere
94	M	L	M	similar sidechains, but L absolutely conserved except for cytbNew
161	V	A	V	similar sidechains, A appears elsewhere
164	I	V	I	similar sidechains, but I absolutely conserved except for cytbGB
169	S	A	S	similar sidechains, A appears elsewhere
187	F	S	F	large difference in size between F/S, F absolutely conserved except for cytbGB
189	I	V	I/V	similar sidechains
192	L	R	L	charge difference in sidechains, R does not appear elsewhere
193	V	A	V	similar sidechains, A appears elsewhere
194	I	T	I	similar sidechains, T only in cytbGB
198	I	L	L	similar sidechains I only in cytbNew
214	N	D	N/D	similar sidechains
228	I	A	I/A	similar sidechains
234	I	L	L	similar sidechains, I appears elsewhere
235	M	L	M	similar sidechains, L appears elsewhere
237	F	L	L	similar sidechains, F appears elsewhere
238	I	P	VILMTA	position variable, P can cause problems, but P appears elsewhere
240	I	L	L	similar sidechains, but I only in cytbNew
241	S	P	L	P can cause problems, P only in cytbGB
249	M	L	M	similar sidechains, L appears elsewhere
295	I	L	L	similar sidechains, L only in cytbNew
296	T	A	A	similar sidechains, T only in cytbNew but position can take dissimilar M
306	M	L	L	similar sidechains, M only in cytbNew but position can take F
320	F	I	I	some side difference, F only in cytbNew
329	T	A	T	similar sidechains, A appears elsewhere
355	S	I	S	similar sidechains, but S absolutely conserved, I only in cytbGB
357	M	L	L	similar sidechains, M appears elsewhere

It was possible to build reasonable models of both sequences, using SWISS-MODEL. The models are virtually identical. Also, calculating accessibility of residues from the model gave the following: (Note that the model did not include residues 1-3; S = accessible, B = buried).

Of the positions that differ between the new and GenBank assemblies, only the following were buried (implying a higher degree of constraint):

24 T/A
 94 M/L
 161 V/A

189 I/V
 235 M/L
 355 S/I
 361 T/L

All changes at buried positions are conservative.

```

      13      23      33      43      53      63
LRKTHPILKTINHSFIDLPTPSNISAWWNFGSLLGICLVIQILTGLFLAMHYTSDTSTAF
SSSSSSSSSSSSBSBSSSSSSBSSSSSSSSBSSSSBSSSSBSSSSSSSSSSSSSSSSS
      73      83      93      103     113     123
SSVAHICRDVNYGWLIRNLHANGASMFFMCMFLHVGRGIYYGSYLYKETWNIGVILLTV
SBSSSBSSSSSSBSSSSSSSSBBSBBSBSSBSSSSSSBSSSSBSSSSSSSSSSBSSBSSB
      133     143     153     163     173     183
MATAFVGYVLPWGQMSFWGATVITNLLSAIPYIGTTLVEWIWGGFSVDKATLTRFFAFHF
SSBSSBSSBSSSSBBBBBSSBSSSBSSSSBSSSSBSSSSBSSSSSSSSBSSSBSSBSSSSS
      193     203     213     223     233     243
ILPFIITALVIVHLIFLHETGSNNPSGINPNSDKIPFHPYYTIKDILGLMIMLFILISLA
SSSSSBSSSSSSSSSSSSBBSBSSSSSSSSSSSSSSSSBSSSSSSBSSSBSSBSSSSB
      253     263     273     283     293     303
LFSPDMLGDPDNFSPANPLNTPPHIKPEWYFLFAYAILRSIPNKLGGVLALITSILILLI
SSSSSSSSSSSSSSSSSSSSBSSSSSSSSBBSBSSBSSSSSSBSSSSSSBSSSSBBSBSS
      313     323     333     343     353     363
IPMLHTSNQRSMFRPFSQTLFWILTANLLTLTWIGGQPVEQPFIIIGQLASIMYFLTII
BSSSSSBSSSBSSBSSBSSBSSBSSBSSBSSBSSBSSBSSSSSSSSBSSBSSBSSBSS
      373
ILMPLAGLLENYMLE
SSSSSBSSBSSSSS

```

There is a helical structure from residues 222–246, a region that contains the proline residues at 238 and 241 of cygGB. This raises eyebrows (prolines tend to destabilize helices) but it is not absolutely impossible. The two prolines do not automatically make this a pseudogene.

Comparing the number of differences in the DNA sequences:

first position of codon 39
 second position of codon 23
 third position of codon 119

Conclusions: Both predicted amino-acid sequences for cytochrome b look reasonable. There are deviations from conservation patterns in both sequences, but modeling of structure does not rule out either one.

Damage. DNA damage levels (measured as cytosine to uracil deaminations, leading to observed C→T miscoding lesions) are shown in Table S2. Similar to previous observations on DNA amplified from mammoth hair (Gilbert et al. 2007a), and despite the thermal ages (Smith et al. 2003) of the thylacine samples being comparable to the mammoth samples due to their curation for over 100 years at room temperature, the

observed cytosine deamination rate is lower than that observed in comparable bone samples. The observation of 0.15% deaminated cytosines in the ‘modern’ numbat hair also provides evidence that a background level of deamination can be expected even within modern hair, presumably derived from a combination of the damage undergone by the mtDNA during the notoriously harsh hair shaft cell keratinization process and damage that happens subsequently during the animal’s life (Linch 2001). A previous analysis of modern cpDNA extracted from fresh yellow poplar (*Liriodendron tulipifera*) tissue has indicated a background sequencing C→T error rate of a maximum of 0.04% (Gilbert et al. 2007b), which is also suggested here by the numbat liver sample, thus indicating a very preliminary damage rate of 0.11% in freshly sampled hair. Clearly however, among other things, this rate will depend upon the age of the hair prior to the animal’s death.

Table S2. Observed rate of cytosine hydrolytic deamination in the modern numbat, ethanol-preserved thylacine, and mammoth samples of comparable thermal ages.

	C	C→T	%C: C→T	Approx. Thermal Age
Numbat liver	172,110	70	0.04	1/a
Numbat hair	154,450	1260	0.15	1/a
Thylacine 2	199,828	1107	0.55	300-1700 ^a
Mammoth bone ‘Poinar’			1.70	173-411 ^b
Mammoth hair ‘M18’			0.39	305-638 ^b
Mammoth hair ‘M22’			0.56	297-609 ^b
Mammoth hair ‘Adams’			0.71	516-910 ^b

^a Estimates conservatively assuming damage commenced at death in 1895, calculated for storage at constant room temperatures of 15°C and 25°C respectively

^b From Gilbert et al. 2007a.

Fraction mtDNA / Human contamination. In addition to the samples from thylacine hair shafts, we sequenced numbat samples from two tissues, liver and hair shafts. The yield of mtDNA from fresh numbat hair was extremely high: about 30% of our reads were mtDNA. From liver, the yield was lower than for thylacine hair, which was roughly comparable to that seen in woolly mammoth hair.

Intriguingly, the data from the thylacine samples indicates that the sequence lengths of the identifiable human contaminant DNA are of similar length to the thylacine (Figure S3). As modern DNA is expected to be longer, this observation suggests that the contaminants amplified in the thylacine dataset are old (that is, do not derive from the DNA extraction or sequencing processes), and are more likely from the sample’s original preparation, or from the ethanol that is known to have been replaced on a number of occasions. Overall, the level of contaminant DNA in the sample is surprisingly high (Table S3), considering that hair shafts have been demonstrated to be easy to decontaminate on several occasions (e.g. Gilbert et al. 2004, 2006, 2007a). In light of the apparent age of the contaminants, one explanation is that over the 100+ years the hair shafts have been soaking in ethanol, the ethanol has helped the contaminant DNA to penetrate much deeper into the shaft than conventional sources of contamination, and was therefore much harder to remove with conventional bleach treatment.

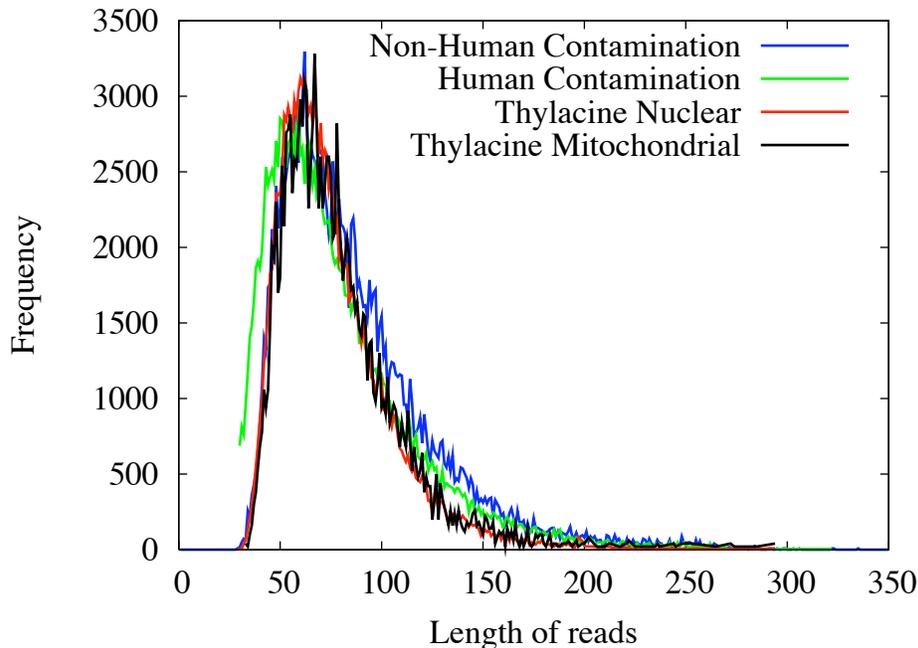


Fig. S3. Length distributions of various classes of sequencer reads from the ethanol-stored thylacine specimen 2.

Table S3. Percentages of endogenous mtDNA in three samples, and percentages of human contamination in thylacine sample 2.

Sample	DNA	Reads	Total Dataset	%
Numbat liver	Numbat mtDNA	2,809	537,424	0.52
Numbat hair	Numbat mtDNA	38,773	129,584	29.9
Thylacine 2	Thylacine mtDNA	11,984	1,037,369	1.15
Thylacine 2	Human mtDNA	136	1,037,369	0.01
Thylacine 2	Human nuDNA	44,493	1,037,369	4.29

Interspersed repeats in the thylacine. We scanned the 69 Mb of sequence from the two thylacine samples for transposable elements (TEs) using RepeatMasker (<http://www.repeatmasker.org>), with the latest version of Repbase 13 (Jurka et al. 2005). LINES are the most abundant element, contributing 9% of the total metagenomic sequence, while SINES, LTR retroposons, and DNA transposons comprise 3.71%, 5.20%, and 2.37% of the sequence, respectively (Table S4). However, these proportions may well be underestimated because of the short read-length of the thylacine data sets and the lack of thylacine-specific repeats in the current Repbase. The composition and relative frequency for the thylacines' TEs are quite similar to those of *Monodelphis domestica*.

Table S4. Summary of the repeat content in the metagenome of Thylacine 2.

Repeat type	Repeat family	Total length (bp)	percentage of sequence (%)
SINEs		2,577,383	3.71
	MIR	1,152,261	1.66
LINEs		6,251,813	9.00
	LINE1	4,630,479	6.66
	LINE2	566,260	0.81
	RTE	427,946	0.62
	CR1	295,538	0.43
LTR		3,616,183	5.20
	MaLRs	209,463	0.30
	ERVL	158,674	0.23
	Gypsy	1,101,555	1.59
	Copia	725,014	1.04
	ERV1	792,027	1.14
DNA		1,644,058	2.37
	MER1_type	114,444	0.16
	MER2_type	57,878	0.08
Total		14,089,437	20.27

Metagenomics. We compared 67,753 reads from thylacine sample 1 and 1,037,369 reads from thylacine sample 2 against the GenBank non-redundant protein database (version November 25 2007) using BlastX. Metagenomic species profiles based on a binning approach using the software MEGAN (version 2beta14) is depicted in Fig. S4 A and B, (Huson et al. 2007, <http://www-ab.informatik.uni-tuebingen.de/software/megan/welcome.html>). Respectively, about 41% and 20% of all reads could be assigned to taxa (27,970 and 202,675 reads) using the parameters MinScore=30, TopPercent=10, MinSupport=40 for thylacine sample 1, and Min Support=100 for sample 2. Similar computations were carried out on 86,812 reads from the numbat hair (38,485 reads assigned) and 102,512 reads from numbat liver metagenome (29,804 reads assigned).

Analysis of the bacterial component of the hair biomes of both thylacine and numbat revealed a much larger number of identified taxa from the thylacine specimen after normalizing for the total number of reads (Fig. S5 A,B; top 10 in Table S5). For this analysis, a value of MinSupport = 3 was chosen for the numbat data set and MinSupport = 36 for the 12-fold larger thylacine 2 data set. Despite the large diversity in the thylacine

microbial biome, 8 bacterial species are present with a significantly higher frequency (Table S5): *Comamonas testosteroni* (5535 reads), *Corynebacterium efficiens*: (1015 reads), *Delftia (Comamonas) acidovorans*: (900 reads), *Flavobacterium johnsoniae*: (1248 reads), *Oceanobacillus iheyensis* (1050 reads), *Propionibacterium acnes* (2261 reads), *Psychrobacter cryohalolentis* K5 (1308 reads) and *Psychrobacter* sp. PRwf-1 (2480 reads). *Comamonas testosteroni*, *Corynebacterium efficiens*, *Delftia (Comamonas) acidovorans*, *Corynebacterium efficiens*, *Flavobacterium johnsoniae*, and *Propionibacterium acnes* are likely to have constituted part of the thylacine's commensal microflora and may have greatly increased their numbers during the putrefaction process and the storage. *Oceanobacillus iheyensis*, *Psychrobacter cryohalolentis* K5 and *Psychrobacter* sp. PRwf-1, however, have been shown to be psychrotolerant, marine organisms. While both genera are also present in the numbat sample at low frequency, their species frequency is 5 – 7 fold higher in thylacine hair. This observation is paralleled by the presence of other cold-adapted marine microorganisms such as *Shewanella frigidimarina* (77 reads), *Shewanella baltica* (99 reads), *Pseudoalteromonas haloplanktis* (85 reads), *Pseudoalteromonas atlantica* (244 reads), *Marinobacter* sp. ELB17 (59 reads), *Oceanospirillales* (94 reads), *Oceanicola batsensis* (38 reads), *Psychroflexus torques* (192 reads), *Croceibacter atlanticus* (106 reads), and *Herpetosiphon aurantiacus* (38 reads). The hair biome of the thylacine sample might therefore still contain environmental microorganisms from its native Tasmania.

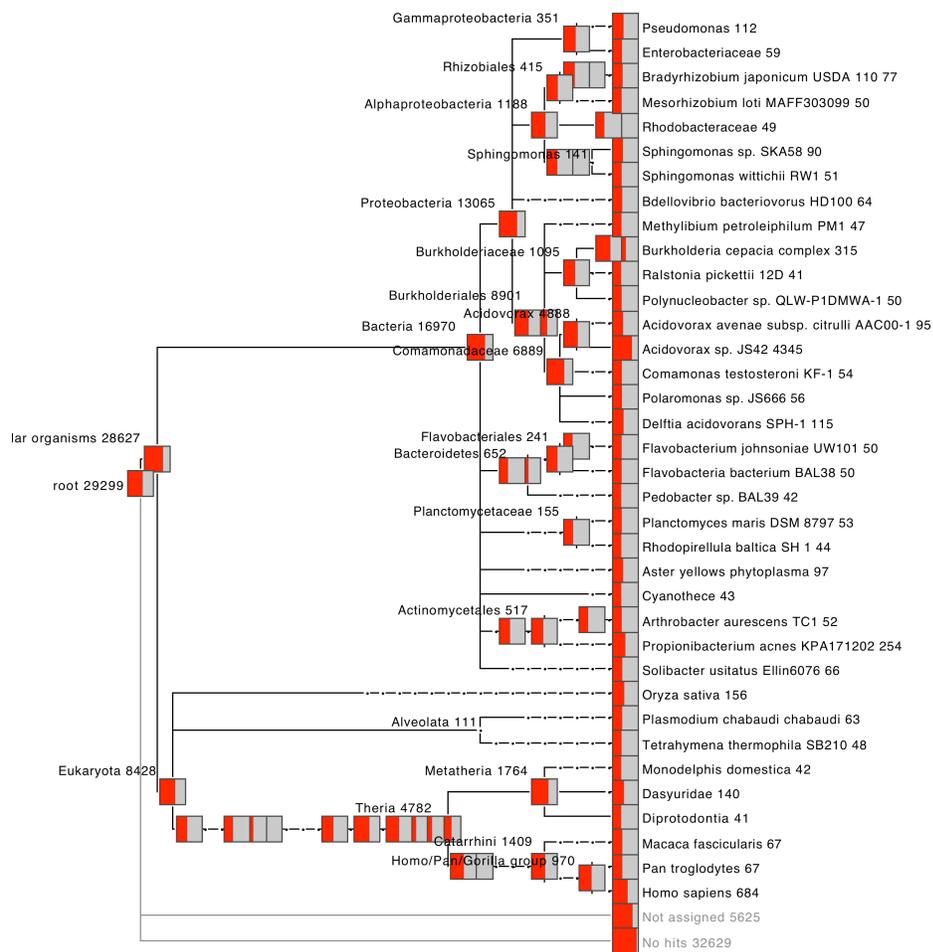


Fig. S4.A. MEGAN taxa tree of thylacine specimen 1

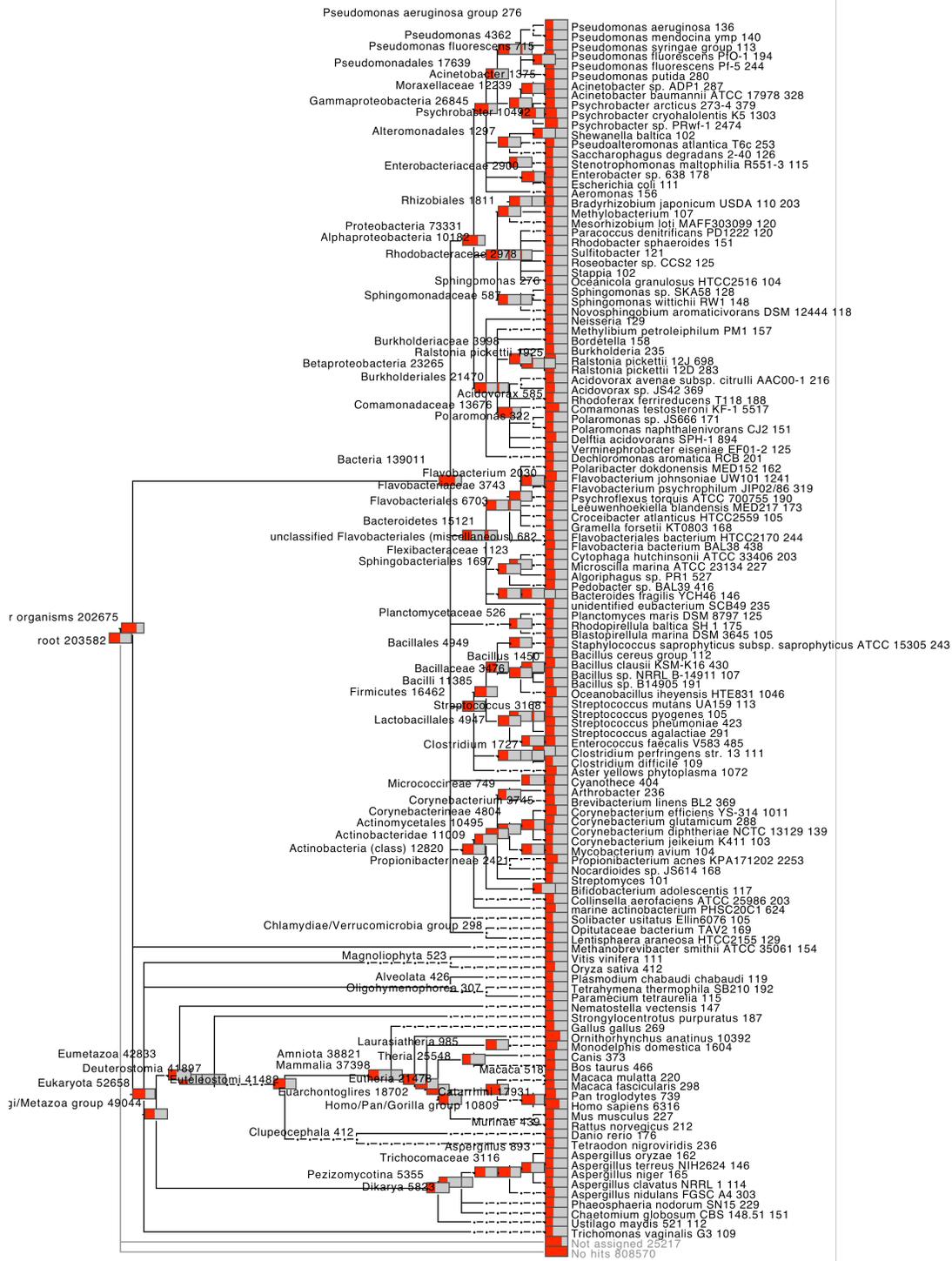


Fig. S4.B. MEGAN tax tree of thylacine specimen 2

Table S5. Top 10 abundant bacterial species in two thylacine metagenomic samples.

	Species	Number of hits
Thylacine 1 (67,753 reads)	<i>Acidovorax</i> sp. JS42	4345
	<i>Burkholderia cepacia</i>	315
	<i>Propionibacterium acnes</i>	254
	<i>Delftia acidovorans</i> SPH-1	115
	<i>Acidovorax avenae</i> subsp. <i>Citrulli</i> AAC00-1	95
	<i>Sphingomonas</i> sp. <i>SKA58q</i>	90
	<i>Bradyrhizobium japonicum</i> USDA110	77
	<i>Solibacter usitatus</i> Ellin6076	66
	<i>Bdellovibrio bacteriovorus</i> HD100	64
	<i>Polaromonas</i> sp. JS666	56
Thylacine 2 (1,037,369 reads)	<i>Comamonas testosteroni</i>	5517
	<i>Psychrobacter</i> sp. PRwf-1qq	2474
	<i>Propionibacterium acnes</i>	2253
	<i>Ralstonia pickettii</i>	1642
	<i>Psychrobacter cryohalolentis</i>	1303
	<i>Flavobacterium johnsoniae</i>	1241
	<i>Oceanobacillus iheyensis</i>	1046
	<i>Clostridium perfringens</i>	1012
	<i>Corynebacterium efficiens</i>	1011
	marine actinobacterium PHSC201C1	624

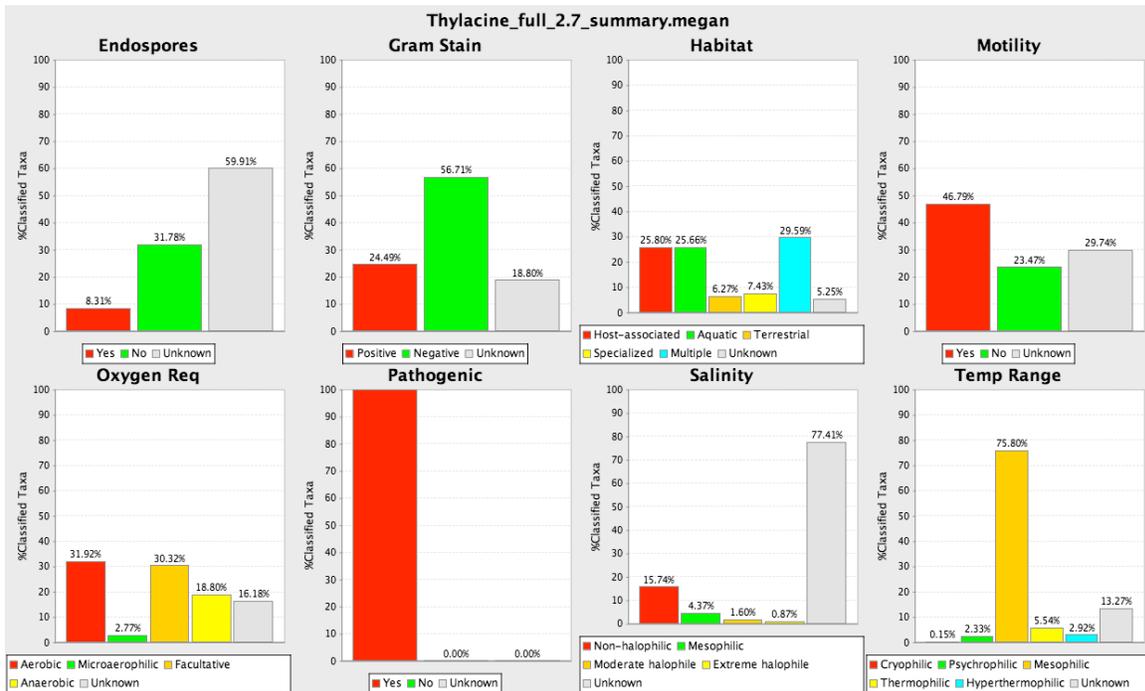
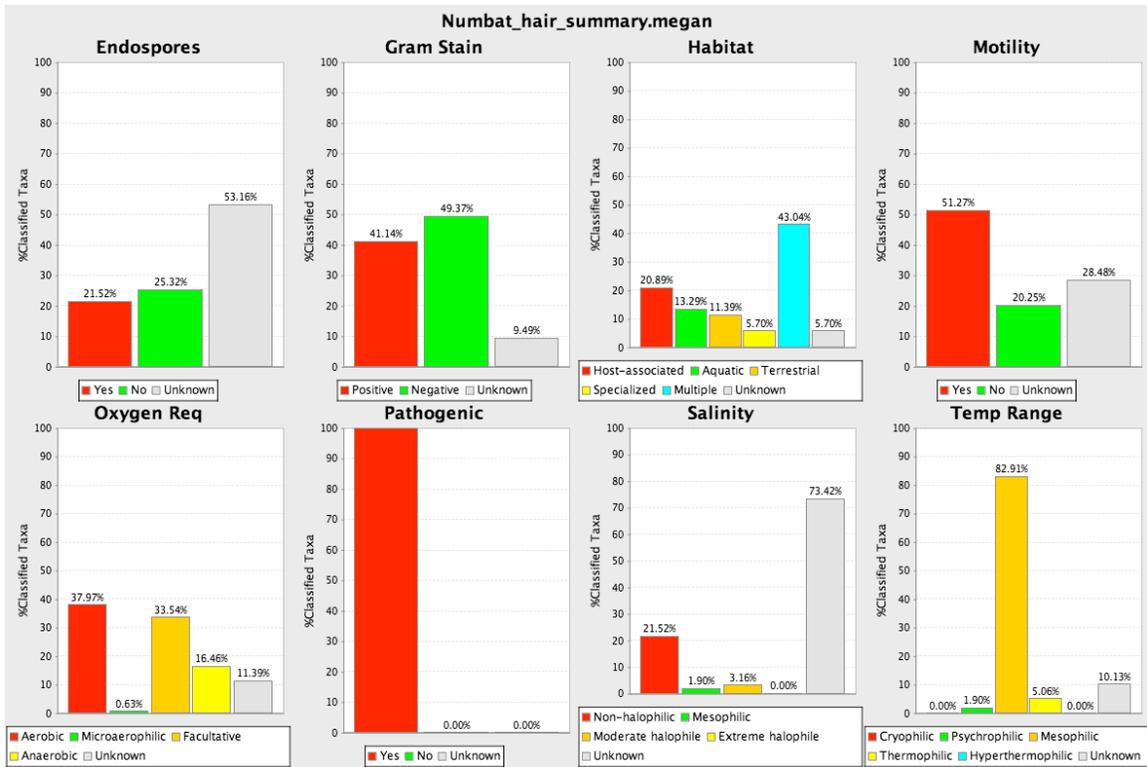


Fig. S5. (A) Microbial attributes of the assigned bacterial reads for the numbat hair and **(B)** the thylacine 2 hair biomes.

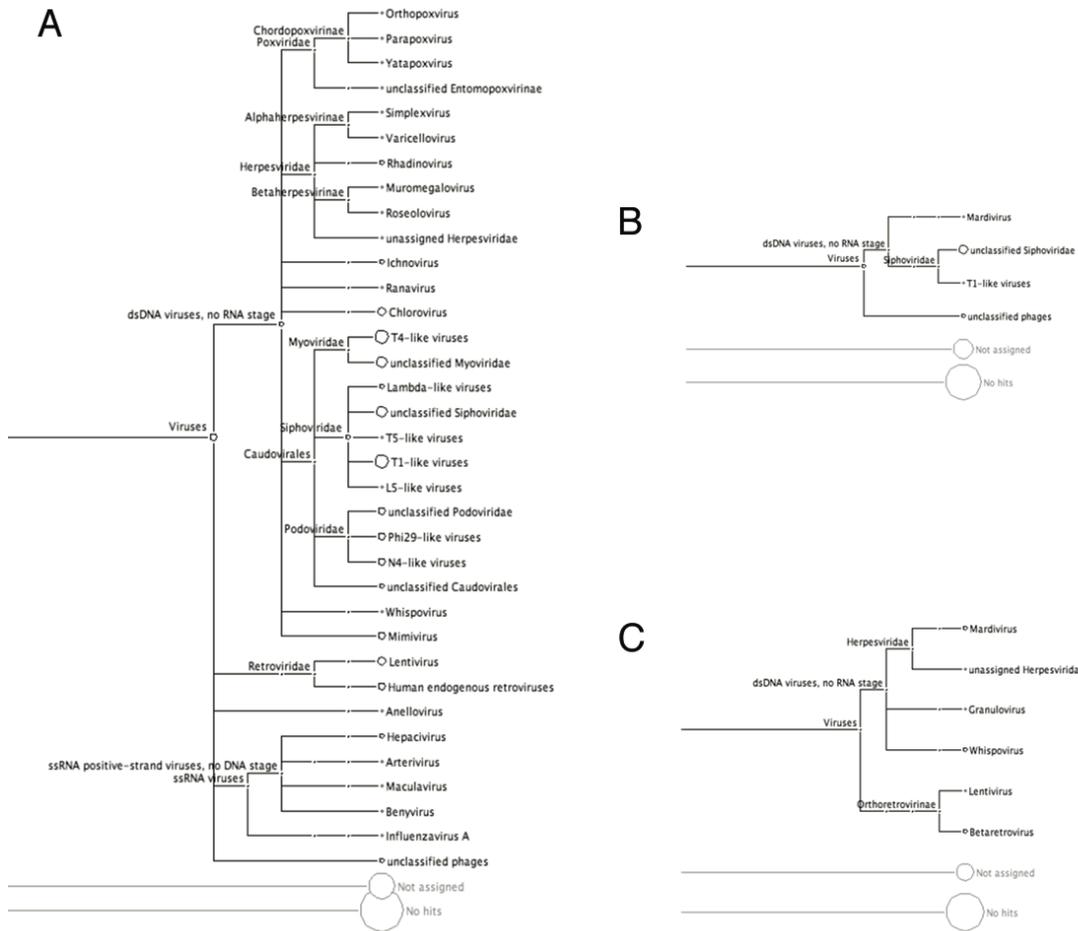


Fig. S6 A. Virome of thylacine 2's hair sample summarized at the genus level. **B.** Virome of the numbat hair sample summarized at the genus level. **C.** Virome of the numbat liver sample summarized at the genus level. Assignments for viral reads were done using MEGAN 2.0 beta14 with the following settings: MinScore = 30.0, TopPercent = 10.0, MinSupport = 2.

Supplemental References (other references are in the main paper):

- Flower, S.S. 1931. Contributions to our knowledge of the duration of life in vertebrate animals. V. Mammals. *Proceedings of the Zoological Society of London* **1931**: 145-234.
- Gilbert, M.T.P., Wilson, A.S., Bunce, M., Hansen, A.J., Willerslev, E., Shapiro, B., Higham, T.F., Richards, M.P., O'Connell, T.C., Tobin, D.J., Janaway, R.C., and Cooper, A. 2004. Ancient mitochondrial DNA from hair. *Curr. Biol.* **14**: R463-464.
- Gilbert, M.T.P., Menez, L., Janaway, R.C., Tobin, D.J., Cooper, A., and Wilson, A.S. 2006. Resistance of degraded hair shafts to contaminant DNA. *Forensic Sci. Int.* **156**: 208-212.
- Guiler, E.R. 1985. Thylacine: Tragedy of the Tasmanian Tiger. *Oxford University Press, Melbourne*.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* **110**: 462-467.
- Wemmer, C. 2002. Opportunities lost: Zoos and the marsupial that tried to be a wolf. *Zoo Biology* **21**: 1-4.