

**SEPA** SCIENCE EDUCATION  
PARTNERSHIP AWARD  
Supported by the National Institutes of Health



University  
at Buffalo

## Methods

Websites that were used:

- 1) Proceedings of the Royal Society Biological Sciences <http://rspb.royalsocietypublishing.org/content/284/186/8/20171804> to obtain sequence identifiers for GenBank.
- 2) DNA to Protein translation <http://insilico.ehu.es/translate/> to obtain 6 possible reading frame translation for sequences.
- 3) BLAST <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- 4) GenBank: <https://www.ncbi.nlm.nih.gov/nucleotide/> to find test sequences, comparable primate sequences, and Himalayan mammal sequences.
- 5) T-COFFEE <http://www.ebi.ac.uk/Tools/msa/tcoffee/> to construct multiple sequence alignments and Phylogenetic Trees.

Raw Nucleotide Sequence collected from a paw

>MG131871.1  
AACAGTAGAACATCCCTTCATTATTATCGGACAACCTCGC  
CTCCATTCTCTACTTTACAATCCTCCTAGTACTTATACCC  
ATCGCTGGAATTATTGAAAACAACCTCTTAAAGTGGA  
GTCTTTGTAGTATAGCAATTACCTCGGCTCTGTAAGCCA

### Protein BLAST of 6 possible reading frames

>Frame 1  
XTVEHPFIIIGQLASILYFTILLVLIPIAGIIENNLLK\*RVFVV\*QLPRSCP

cytochrome b, partial (mitochondrion) [Ursus arctos pinnosus]  
Sequence ID: [AT769718.1](#) Length: 38 Number of Matches: 1

Range 1: 1 to 37 [GenPept](#) [Graphics](#) [▼](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
70.9 bits (172)	3e-15	Compositional matrix adjust.	36/37(97%)	37/37(100%)	0/37(0%)

Query 2 TVERHPTIIIGGLASIIYPTTILAVPIAGIIGNNLAK 38  
Hit 1 TVERHPTIIIGGLASIIYPTTILAVPIAGIIGNNLAK 38

Figure 2. Results of Blast for Reading Frame 1 showing a cytochrome b gene product in a *Ursus arctos pruinosus* (Tibetan Blue Bear).

>Frame 2  
EQ\*NIPSLSDNSPPFSTLQSS\*YLYPSLELLKTTs\*SEESL\*YSNYLGL  
VSC  
**No significant similarity found**

>Frame 3  
NSRT LHYRTTRLHSLLYNPPSTYTHRWN\*YKAPAL  
KVKSLCSIAITSVL\*A  
**No significant similarity found**

>Frame -1  
WLTRPR\*LLYKDSSL\*EVVFNNSSDGKY\*EDCKVENGGELSDNNE  
GMFYCS  
**No significant similarity found**

>Frame -2  
GLQDRGNCYTTKTLHFKRLFSIIPAMGISTRIRIVK\*RMEASCPIMKGC  
STVC  
MULTISPECIES: hypothetical protein [Chlamydia]  
Score 38.9, Evalue 0.027, 54% identity.

>Frame -3  
AYKTEVIALQRLFTLRGCFQ\*FQRWV\*VLGGL\*SREWRRVVR\*\*\*  
RDVILF

No significant similarity found

**Humans (*Homo sapiens*)**

Homo sapiens isolate Z012 mitochondrion, complete genome

Sequence ID: KP240688.1 Length: 16557 Number of Matches: 1

Rank 1: 15779 to 15915 GenBank Graphics [View Match](#)

Score	Expect	Identities	Gaps	Strand/Plus
51.8 bits(56)	2e-04	95/137(69%)	2/137(1%)	
Query 23	ATTATGAGCAACCTGGCTGCTCTTCTTCTACCTGCTAGTATCTATTCACCT			82
Shj02 15779	ATTATGAGCAACCTGGCTGCTCTTCTTCTACCTGCTAGTATCTATTCACCT			82
Query 83	CTGTGATATTTGAAGAACACTCTTAAATGAGAGCTTTTGATATGACATATCAT			140
Shj03 15839	CTGTGATATTTGAAGAACACTCTTAAATGAGAGCTTTTGATATGACATATCAT			140
Query 141	CTGGGTTCTTTGATGACGA 157			1589
Shj019 15899	CTGGGTTCTTTGATGACGA 15915			1589

Figure 3. Score is low 51.8. Identity is low at 69% for 95 base pairs. 2 Gaps are seen.

**Chimpanzee (*Pan troglodytes*)**

Pan troglodytes chromosome 22 BAC RP43-00219, complete sequence  
Sequence ID: [AL954227.3](#) GenBank: 192219 Number of Matches: 1

Rank 1: 36974 to 37067 Length: 92 bases [Next Match](#)

Score	Expect	Identities	Gaps	Strand
46.4 bits(50)	0.001	68/94(72%)	3/94(3%)	Plus/Plu
Query 63	TCTTCAATGATGATCCACCTGCTGAAATTTGAAACACCTCTTAAAGTAGG-AGTC			121
Subject 36974	TCTTCTTTTATCATCTGACTGAAATTTGAAACAGCTCGTCAATATGATGCTG			370333
Query 122	TCTTGATGATTA--GCAGTACTCTGGGTCTGTAA	153		
Subject 37034	ATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	37067		

Figure 4. Score is low 46.4. E-value is barely significant 0.001, Identity is low at 72% with 68 base pairs. The alignment includes 3 gaps.

**Orangutan (*Pongo pygmaeus*)**

Pongo pygmaeus genome assembly, chromosome: XII  
Sequence ID: [LT571452.1](#) Length: 136302887 Number of Matches: 247

Range 1: 48847897 to 48847915 [GenBank](#) [Graphics](#)

Score	Expect	Identities	Gaps
35.6 bits(38)	0.009	19/19(100%)	0/19(0%)

Query 99 ACAACCTCTCTTAAAGTGAAG 117  
Sbjct 48847915 ACAACCTCTTAAAGTGAAG 48847897

Figure 5. Score is low 35.6. E-value is 0.009. Identity is 100% but for only 19 base pairs starting at position 99.

**Himalayan Black Bear (*Ursus thibetanus mupinensis*)**

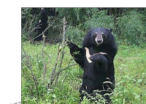


Figure 6. Himalayan Black Bear Source:  
<https://www.coniferousforest.com/wp-content/uploads/2016/05/Asian-Black-Bear-Habitat.jpg>

Ursus thibetanus mupinensis mitochonchion, complete genome  
 Accession ID: [DQ402478.1](#) Length: 16868 Number of Matches: 4

Range 1: 16163 to 16317 [GenBank](#) [Graphics](#) ▼ Next Match    Ph

Score	Expect	Identities	Gaps	Plus/Minus
199 bits(220)	2e-54	137/155(88%)	0/155(0%)	
Query 3	CAGTGAAGCATCTCCCTTCATATCTAGGACCAACCTCCCTTCATCTCTGATTTAAATACCT	62		
Subject 16163	CAGTGAAGCATCTCCCTACATATCTAGGACCAACCTCCCTTCATCTCTGATTTAAATACCT	16222		
Subject 16223	TCCTGATCTATACCACTCACTCGAATATCTTAAGAACCACTCTTAAGTAAAGAGCTCT	122		
Subject 16223	TCCTGATCTATACCACTCACTCGAATATCTTAAGAACCACTCTTAAGTAAAGAGCTCT	16282		
Subject 123	TTGTGATATAGCAATCACTACCTGGCTCTGTAGGACA	157		
Subject 16282	TTGTGATATAGTAACTACCTGGCTCTGTAGGACA	16317		

Figure 7. Score is high 199. E-value is highly significant. Identity is 88% for 137 base pairs. The source of >MG131871 is something bear-like.

### Snow Leopard (*Panthera uncia*)

Range 1: 15894 to 16046		GenBank	Genbank	Next Match	Previous Match
Score	Expect	Identities	Gaps	Strand	
141 bits(156)	5e-36	123/153(80%)	0/153(0%)	Plus/Plus	
Query 5	GTAGACATCCCTTCATTATATCGGACATCGCTCCATCTCTCTCTTACATCTC	64			
Subject 15894	GTAGATATCTTCTTATGACATCGGACATAGCTCTATCTCTCTCTCTCTCTCT	15953			
Query 6	CTAGTACTTATCCATCCATCGCGAATATGAAACACCTCTTAAGTGAAGAGCTTT	124			
Subject 15954	CTATGATATATACCACTTCAGGAGATATGAAACCGCCTCTCTAATGAGAGCTT	16013			
Query 125	GTAGTATAGCAATTACCTGGGTCTGTGAACGA	157			
Subject 16014	GTAGTATAGCAATTACCTGGGTCTGTGAACGA	16046			

Figure 8. Score is medium 141. E-value is 5e-36. Identity is 80% but for only 123 base pairs starting at position 5.

The Top BLAST Hit for the Nucleotide sequence was a Himalayan Black Bear (*Ursus arctos pruinosus*). First, we used protein BLAST to identify 6 possible reading frames where the only significant result was *Ursus arctos pruinosus* (Tibetan Blue Bear) with a 97% identity for Reading Frame 1. Reading Frame -2 BLAST resulted in a multispecies: hypothetical protein [Chlamydia] which ended up to be unlikely due to 57% identity. We found that the possible *Yeti* gene taken from a paw is not related to a *homo sapien*, *pan troglodytes*, or *Pongo pygmaeus* by BLASTing the nucleotide sequence of the suspected paw to these animals. After this we BLASTed the nucleotide sequence against another type of bear found in the Himalayas (*Ursus thibetanus*), which had an high score of 199 and a high identity of 88% with 137 base pairs. E-value 2e-54. Therefore, this paw is most likely a type of bear and not a yeti.

1 Lan T, Gill S, Bellemain E, Bischof R, Nawaz MA, Lindqvist C. 2017 Evolutionary history of enigmatic bears in the Tibetan Plateau-Himalayan region and the identity of the yeti. *Proc. R. Soc. B* 284:20171804.

<http://dx.doi.org/10.1093/csb/2017.1804>

## Acknowledgments

Supported by an NIH Science Education Partnership Award  
R25OD010536.  
Thank you to Dr. Stephen Koury, University at Buffalo

[www.buffalo.edu](http://www.buffalo.edu)

The region of the Tibetan Plateau and Himalayas is famous for the legend of a large bipedal hairy humanoid creature called the 'yeti', 'chemo', 'methi' or 'bharmando'. If such a creature exists, there must be feces, bones and fur in the area that contain unique DNA sequences that would align with other known primates such as humans, chimpanzees, gorillas, and orangutans.

In the December 2017 issue of the *Journal The Proceedings of the Royal Society: Biological Science* Volume 284, Issue 1868, Lan, Gill, Bellemain, Bischof, Nawaz, and Charlotte Lindqvist (Department of Biological Sciences, University at Buffalo) conducted a comprehensive genetic survey of field-collected and museum specimens to explore their identity. The resulting journal article was entitled "Evolutionary history of enigmatic bears in the Tibetan Plateau–Himalaya region and the identity of the yeti". A total of 24 field and museum collected samples, including hair, tissue, bone and feces, were analyzed in their study. The samples came from Tibet, India, Pakistan, and Nepal.

DNA was extracted from the samples using the Qiagen DNA Stool Kit and DNeasy Blood & Tissue DNA Kit . Extracted DNA was amplified using Polymerase Chain reactions in a 25 ul volume. The resulting DNA was sequenced by Microarray (Now Arbor Biosciences, <http://www.arborbiosci.com>).

The purpose of this project is to determine if sequence >MG1318\_\_ came from a scientifically unknown relic bipedal Hominiidae, commonly called the Yeti or another known primate or another Himalayan Mammal. Critical thinking skills and gene annotation skills were used to scientifically investigate a popular cryptid. The second purpose is to independently verify the work of Dr. Charlotte Lindqvist and the other scientists.



Figure 1. Photo of a reported Yeti Paw.  
Source:  
<https://www.atlasobscura.com/articles/saga-of-the-yeti-hand>