

# An investigation of reported Yeti Hair Sequence >MG131873

Marilynn Amborski, Ella Archer, Grace Herr, and Lon Knappenberger  
Westfield Academy and Central School, 203 East Main Street, Westfield, New York, 14787  
and The Western New York Genetics in Research and Health Care Partnership

## Abstract

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 Lindqvist conducted a comprehensive genetic survey of field-collected and museum specimens to explore their identity. The purpose of this project was to use the newly acquired skills of gene annotation to verify the claims of this article. Sequences cited in the article and stored in GenBank were translated into six reading frames using <http://insilico.ehu.es/translate/> and completing a Protein BLAST of the results. Multiple sequence alignments and phylogenetic trees were then constructed comparing the sequence to other primates and Himalayan mammals to determine the source of the sequence. The sequence appears to be from the Brown Bear (*Ursus arctos*).

## Introduction

The region of the Tibetan Plateau and Himalayas is famous for the legend of a large bipedal hairy hominid creature called the 'yeti', 'chemo', 'mheti' 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 µl volume. The resulting DNA was sequenced by Mycroarray (Now Arborbiosciences. <http://www.arborbiosci.com>).

The purpose of this project is to determine if sequence >MG1318\_\_ came from a scientifically unknown relic bipedal Hominidae, 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. Picture of reported Yeti Hair.  
Source: <https://phys.org/news/2017-11-abominable-snowman-nopestudy-ties-dna.html>



## Methods

Websites that were used:

- 1) Proceedings of the Royal Society Biological Sciences <http://rsob.royalsocietypublishing.org/content/284/1868/20171804> 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 hair.

>MG131873.1  
GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTA  
GAACATCCCTCATTATTATCGGACAACATAGCCATCTTTTCT  
ACTTTACAACTCCCTAGTACTATACCCATCGCTGGAATTAT  
TGAAACAACCTCTTAAATGAAGAGCTTTGTAGTATAGCAA  
TTACCTTTGGTCTGTAAGCCAA

## Results

### Protein BLAST of 6 reading frames

>Frame 1  
XQTY\*H\*HELEDNQ\*NIPSLSDN\*PPFTLQSS\*YLPSELLEKTT\*NEESLYSNYLGLVS  
Q  
No significant similarity

>Frame 2  
GRPTNTNMWRTSRSTSHYRTSLHFLYNPSTYTHRWNY\*KQPLMKLSLSIAITLV  
L\*AA  
No significant similarity

>Frame 3  
ADLLTLT\*IGGQVPEHPIIQLASIFYTILLVLIPIAGIENLLK\*RVFV\*QLPWSCKP  
cytochrome b, partial (mitochondrion) [Ursus arctos isabellinus]  
Sequence ID: A1289720.1 Length: 60 Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gaps
90.1 bits(222)	1e-22	Compositional matrix adjust.	47/49(96%)	48/49(97%)	0/49(0%)

Query 1 ADLALTY\*IGGQVPEHPIIQLASIFYTILLVLIPIAGIENLLK 49  
ADLALTY\*IGGQVPEHPIIQLASIFYTILLVLIPIAGIENLLK 49  
ADLALTY\*IGGQVPEHPIIQLASIFYTILLVLIPIAGIENLLK 49

Figure 2. Reading Frame 3 gave a top Hit of Cytochrome b for Himalayan Brown Bear (*Ursus arctos isabellinus*).

>Frame -1  
LAYTKVIAILQRLFILRGCFQ\*FQRWV\*VLGGL\*SRKWRLVVR\*\*\*RDVLLVLFQMLVLVG  
LP  
No significant similarity

>Frame -2  
WLTRPRL\*LYKXSSFEVFNNSDGYK\*EDCKVENG\*LSNNEGMFYWSSNSC\*  
\*VCC  
No significant similarity

>Frame -3  
GLDDGGQYKTTLHFLKSLIPAMGISTRIVK\*KMEASCPIMKGCSTGCPPIHVSVR  
SA  
MULTISPECIES hypothetical protein [Chlamydia] Score 55.5 bits. E-value  
1e-08. Identity 28/50(56%)

### nBlast of gene and Himalayan Mammals *Ursus arctos* (Brown Bear)

Score	Expect	Identities	Gaps	Strand
349 bits(386)	6e-97	193/193(100%)	0/193(0%)	Plus/Plus

Query 1 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 60  
Sbjct 15673 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 15732  
Query 61 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTGGAATTAT 120  
Sbjct 15733 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTGGAATTAT 15792  
Query 121 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 180  
Sbjct 15793 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 15852  
Query 181 TCTTTTAAACCA 193  
Sbjct 15853 TCTTTTAAACCA 15865

### *Ailurus Fulgens* (Red Panda)

Score	Expect	Identities	Gaps	Strand
232 bits(256)	Se-63	167/193(87%)	0/193(0%)	Plus/Plus

Query 1 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 60  
Sbjct 15169 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 15228  
Query 61 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTGGAATTAT 120  
Sbjct 15229 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTGGAATTAT 15279  
Query 121 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 180  
Sbjct 15280 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 15339  
Query 181 TCTTTTAAACCA 193  
Sbjct 15340 TCTTTTAAACCA 15352

### *Ursus tibetanus* (Himalayan Black Bear)

Score	Expect	Identities	Gaps	Strand
289 bits(320)	Se-80	180/193(93%)	0/193(0%)	Plus/Plus

Query 1 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 60  
Sbjct 16747 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 16806  
Query 61 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTGGAATTAT 120  
Sbjct 16807 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTGGAATTAT 16866  
Query 121 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 180  
Sbjct 16867 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 16926  
Query 181 TCTTTTAAACCA 193  
Sbjct 16927 TCTTTTAAACCA 16939

### nBlast of gene and Primates. *Pan Troglodytes* (Chimpanzee)

Score	Expect	Identities	Gaps	Strand
105 bits(116)	2e-21	140/192(73%)	2/192(1%)	Plus/Plus

Query 2 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 60  
Sbjct 15159 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 15218  
Query 62 TCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTG 121  
Sbjct 15219 TCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTG 15278  
Query 122 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 180  
Sbjct 15279 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 15338  
Query 180 TCTTTTAAACCA 193  
Sbjct 15339 TCTTTTAAACCA 15350

### *Gorilla gorilla* (Gorilla)

Score	Expect	Identities	Gaps	Strand
94.2 bits(92)	1e-15	136/193(70%)	2/193(1%)	Plus/Minus

Query 1 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 60  
Sbjct 1399 GCAGACCTACTAACCTAACATGAATTGGAGGACAACACAGTAACATCCCTCATTATT 1340  
Query 61 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTG 120  
Sbjct 1399 ATCGGACAAGTACCTCCATTTCTTCTTACAACTCCCTAGTACTATACCCATCGCTG 1280  
Query 121 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 180  
Sbjct 1279 GGAATATGAAACACCTCTTAAATGAAGAGCTTTGTAGTATAGCAATCACTCTGG 178  
Query 179 GCTCTTTAAACCA 191  
Sbjct 1219 GCTCTTTAAACCA 1207



Figure 3: Himalayan brown bear (*Ursus arctos isabellinus*) the top hit for Protein and Nucleotide BLASTs.  
Source: <https://www.zoochat.com/community/media/himalayan-brown-bear-ursus-arctos-isabellinus.237040/>

### Phylogenetic Trees of >MG131873 and Cytochrome b sequences from other Mammals

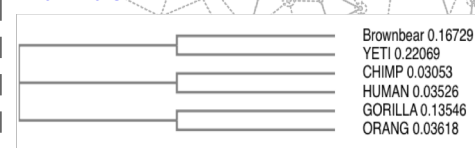


Figure 4: Phylogenetic Tree for >MG131873 and Cytochrome b (shown as "YETI") and Brown Bear (*Ursus arctos*) and Possible Primates. The "Yeti" sequence is on the same branch as the Brown Bear as expected.

## Conclusion

The sequence >MG131873.1 was acquired from GenBank and transcribed and translated into six different reading frames. The different protein sequences were blasted, however they only produced one significant result (Cytochrome b in an Asian Brown Bear). Next, the raw nucleotide sequence was BLASTed against primates and Himalayan mammals. The results showed that it was closest to the Himalayan mammals, more specifically, the Himalayan Brown Bear *Ursus arctos*.



## References

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

<http://dx.doi.org/10.1098/rspb.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)