Supplemental Data 11- Unpublished Manuscript from Helix Biological Laboratories Predating this Manuscript

Introduction

Efforts continue to be put forth to prove the existence of a large, hairy bipedal creature known across the world by such names as Bigfoot, Yeti and Yowie. Some individuals believe that this large, hairy bipedal creature is an ape or ape-like primate that has yet to be classified and survives today as part of a relict population. Others believe that the creature is more human-like and may possibly belong to a relict lineage of hominids that have survived into the present day. In either case, individual reports continue to be submitted to local authorities and cryptozoology research organizations describing this elusive creature in detail. These reports document sightings, encounters and tangible pieces of evidence such as footprints, vocalizations captured on audiotape, photographs, videotape footage and hair morphology analysis.

Analysis of footprint samples initially provided the best pieces of evidence for these studies, and in some instances has provided possible evidence of the existence of large bipedal hominid creatures in North America (Meldrum 2007). More recently, with the advent of molecular biology techniques, the analysis of collected hair samples using DNA sequencing techniques is now included as a potential method of obtaining evidence to prove the existence of a living cryptid bipedal primate species. DNA sequence analysis studies involving the extraction of DNA from hair samples, allegedly derived from cryptid bipedal primate species, have in some instances contributed minimal amounts of scientific data that can be applied toward this area of research (Coltman and Davis 2006; Milinkovich et al. 2004). However, other studies tend to be incompletely described and lack proper detail, or they reach extremely vague conclusions. Although the body of work in this area is still lacking, DNA analysis of hair samples has the potential to provide credible evidence for the existence of a still yet to be described bipedal primate or hominid species.

Since the follicle portion of the hair degrades so rapidly after removal from the epidermal layer of the skin, hair samples are limited in that they allow only mitochondrial DNA to be extracted almost exclusively. The only exception is if a DNA extraction procedure can be performed immediately upon the removal of a hair sample from an individual.

The human mitochondrial genome is a closed, circular molecule located within the mitochondrial matrix and present in thousands of copies per cell. It is 16,569 base pairs (bp) in length (Anderson *et al*. 1981). Mitochondrial DNA (mtDNA) has two strands, consisting of a guanine-rich heavy (H) strand and a cytosine-rich light (L) strand. The heavy strand contains 12 of the 13 polypeptide-encoding genes, 14 of the 22 tRNA-encoding genes and both rRNA-encoding genes. Introns are absent in mtDNA, and all of the coding sequences are contiguous (Anderson *et al*. 1981, Wallace *et al*. 1992, Zeviani *et al*. 1998). The only non-coding segment of mtDNA is the displacement loop (D-loop),

a region of 1121 bp that contains the origin of replication of the H-strand (O_H) and the promoters for L and H-strand transcription.

One of the features of mtDNA genes is that they have a much higher mutation rate than the coding regions of nuclear DNA genes. For this reason, mtDNA sequences are frequently used in phylogenetic analyses involving vertebrates and provide substantial resolution in defining evolutionary relationships at both the genus and species level Miyamoto and Boyle 1989; Meyer and Wilson 1990; Irwin et al. 1991). This feature along with a well-studied familiarity of mtDNA sequence and structure, make mtDNA a good candidate for other types of DNA analysis in addition to phylogenetic studies. All of these factors provide a considerable level of confidence for using mtDNA in a study involving the analysis of hair samples.

The goal of this particular study is not to unequivocally prove or disprove at this time the existence of a living cryptid bipedal primate species or to locate the position of such a creature on the evolutionary tree. The goal of this study involves a search for scientific facts through proper experimental methodology and analysis of scientific data. In order to accomplish this goal, mtDNA sequence data obtained from specific hair samples derived from alleged cryptid bipedal primate species will be carefully analyzed and compared using molecular techniques via PCR and DNA sequencing. The hair samples used in this study have been collected during various field research expeditions and also directly from an alleged living cryptid bipedal primate species residing on the Carter farm located in Cookeville Tennessee.

Materials and Methods

Collection of Samples

The hair samples analyzed in this study were collected from several locations including Tennessee, Texas, Michigan and Australia. These hair samples were collected from areas where frequent sightings of or encounters with cryptid bipedal primate species have been reported. In order to minimize possible human contamination, individuals collecting hair samples wore gloves. Hair samples were placed in sealed plastic bags for storage until used in laboratory analysis. In each instance, multiple hair samples were collected from each location. Single individual hair samples were analyzed in separate DNA extraction, PCR and sequencing experiments.

DNA extraction

Hair samples used for DNA extractions were acquired from several sources. DNA was extracted using the Promega DNA IQ kit (Madison,WI) according to manufacturer's instructions. One hair strand was used for each DNA extraction in order to ensure that the DNA extracted was obtained from a single individual. Hair samples were carefully removed from sealed containers using sterile forceps. Each hair sample was cut into

several smaller pieces using sterile scissors in order to improve the overall quality of the DNA extraction.

In addition to the collected hair samples, DNA extractions using positive and negative control samples were also performed. The positive control samples consisted of hair samples from a domestic dog (*Canis familiaris*) and a domestic cat (*Felis catus*). The positive control hair samples were treated and prepared identically to the hair samples from the alleged cryptid primate species. The negative control sample did not contain a hair sample and consisted only of the DNA extraction reagents from the DNA IQ kit.

Primer design, PCR and DNA sequencing

With regard to the mtDNA genome, we focused on amplifying and sequencing a region that was part of the mitochondrial cytochrome b (cyt b) gene. This target region within cyt b is approximately 308 bp in length and spans nucleotide positions 14841-15149 in the human mtDNA genome (Genbank Accession Number EU571946).

The selected 308 bp target region of the mitochondrial cyt b gene was amplified using primers that were previously described by Kocher et al. (1989). The sequences of these primers are as follows:
L14841 (5' AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA 3') and H15149 (5' AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA 3'). Each of these primers was synthesized by Sigma Genosys (The Woodlands, TX).

We attempted to amplify longer fragments of the mitochondrial cyt b gene using additional primers described by Kocher et al (1989), with the objective of obtaining more extensive mtDNA sequence data. However, amplifying longer fragments (in excess of 800 nucleotides) proved at this point to be difficult, as we were not able to generate a large enough quantity of PCR product of this length for effective DNA sequencing. Most likely this is due that the fact that the mtDNA from the hair samples has undergone a degree of degradation from the time they were collected until the time that mtDNA was extracted from them. This makes it more difficult to obtain lengthier sized amplicons using PCR.

PCR reactions were run in a 50ul volume using the Promega Master Mix according to the manufacturer's instructions. Each PCR reaction contained ~ 100 ng of purified total DNA extracted from hair samples. The thermal cycler conditions for PCR were as follows: initial denaturation: 94°C for 2 min; Touchdown PCR (4 cycles for each C) 94°C for 45 s; 60°-53°C for 45 s; 68°C for 2 min; final extension: 68°C for 10 min. To protect against possible contamination, a negative control sample for which no DNA template was added was included in each PCR.

In order to determine the quality of PCR amplification, amplified PCR products were electrophoresed on a 1.5% agarose gel at 90V for 45 minutes. Following electrophoresis, the agarose gel was stained with ethidium bromide, viewed under UV light and

photographed. Amplified mtDNA fragments were compared against a 100 base pair standard ladder (Promega) to ensure that amplicons were the correct size. In all PCR experiments, the negative control reactions did not show amplification, indicating that no contamination had taken place during PCR experimentation.

Unincorporated primers and other reagents were removed using the Qiagen QIAquick PCR Purification kit (Valencia, CA). DNA sequencing reactions were done using the Applied Biosystems Big Dye primer cycle sequencing kit (Foster City, CA) according to the manufacturer's specifications. Sequencing of the reactions was performed on the ABI 3700 (Applied Biosystems). Primer L14841 and primer H15149 were both used for DNA sequencing of each sample.

Results

Excellent resolution of mtDNA sequence was obtained for at least ten hair samples originating from different collection locations. Each of the individual mitochondrial cyt b sequences obtained was analyzed for clarity and resolution. After passing visual inspection of this criteria, each sequence was subjected to a BLAST search using the National Center for Biotechnology Information (NCBI) website http://www.ncbi.nlm.nih.gov/. In the BLAST search, each of the individual mitochondrial cyt b sequences is submitted to a database containing all published genomic sequences on record. The submitted sequence is compared to all sequences in the NCBI database with the objective of matching the submitted sequence with like or identical sequences contained in the database.

In regard to mtDNA sequences obtained from hair samples, we did not necessarily expect to find a mtDNA sequence that was unique when compared to other sequences in the database. Instead, we were generally searching for actual "nucleotide anomalies" present in the sequence that would indicate a possible primate or hominid association, necessitating further examination at the molecular level.

There were four instances when the BLAST search identified the mitochondrial cyt b sequence obtained from an individual hair sample as being derived either from bison, wild boar or bobcat. However, in all other instances, the BLAST search identified the mitochondrial cyt b sequence obtained from an individual hair sample as being identical to the human mitochondrial cyt b sequence for the 308 bp region analyzed. These results are shown in Figure 2.

The mitochondrial cyt b sequences that were obtained from the individual dog and cat hair samples were also subjected to a BLAST search. The results of the BLAST search indicated that the target mitochondrial cyt b sequence obtained from the domestic dog hair sample and the domestic cat hair sample each matched with 100% identity the respective domestic dog and domestic cat mitochondrial cyt b sequences in the NCBI database, thus confirming that these hair samples were derived from dog and cat.

The hair samples collected in Tennessee at the Carter farm are alleged to be directly from a cryptid bipedal primate creature that has lived in cohabitation with humans at that site for several decades. Two individual hair samples obtained from different areas of the body of this alleged cryptid primate creature were both subjected to mtDNA analysis. The results of the BLAST search indicated that the 308 bp mitochondrial cyt b sequence obtained from the two individual hair samples from this alleged cryptid hominid creature matched with 100% identity the human mitochondrial cyt b sequence for the same region of cyt b. In order to see if these results could be replicated, DNA isolation, PCR amplification and DNA sequencing were repeated two additional times for each sample The identical result was achieved in each replicated experiment.

Discussion

In phylogenetic and sequence comparison studies, the examination of sequence from the mitochondrial cyt b gene provides various degrees of information, depending upon the divergence times of the organisms under analysis. This needs to be taken into strong consideration when comparing the sequences of orthologous cyt b genes.

By definition, orthologs are genes that are related by vertical descent from a common ancestor and encode proteins with the same function in different organisms. Thus, two genes are defined as orthologous if they diverged after a speciation event, such as the divergence that produced the lineage of modern day humans (*Homo sapiens*) separate from chimpanzees (*Pan troglodytes*).

The fact that several of the individual hair samples analyzed produced mitochondrial cyt b sequences that matched human cyt b sequence for the 308 bp region is a very curious find. The problem of contamination in PCRs or misidentification of hair samples collected in the field cannot be discounted. However, the fact that the individual hair samples from domestic dog and domestic cat produced cyt b sequences for the 308 bp region that identically matched their respective cyt b sequences in the NCBI database, is an indication that the methodology used during collection of hair samples, DNA extraction and PCR is not introducing contamination.

More importantly, in instances where the analyzed hair samples were verified as being derived from bison, wild boar and bobcat, the correct identity of these cyt b sequences were eventually demonstrated through the BLAST search.

Although the authenticity of the cyt b sequences that matched identically with human cyt b sequence for the 308 bp region examined can be debated, there may be a plausible explanation as to why this result was seen. As mentioned, the amount of information yielded by mitochondrial cyt b sequence is dependent upon the degree of divergence between the organisms under analysis. Organisms separated by many millions of years of evolutionary time will have undergone greater divergence. The amount of divergence between organisms will be reflected in the number of nucleotide substitutions (i.e. polymorphic sites) in the comparison of orthologous genes. When comparing orthologous

gene sequences of organisms that are separated by a greater amount of evolutionary time, more divergence between the sequences will be seen.

Primate divergence times continue to be debated and there has long been speculation as to the correct divergence time between chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens*). Some place the divergence between *Pan* and *Homo* at 5 million years ago (MYA) while others place the *Pan* and *Homo* split at closer to 10 MYA. There has been considerable progress in primate palaeontology since the publication of Sarich and Wilson's (1967) paper. The oldest fossil, which unquestionably is on the *Homo* lineage is *Orrorin tugenensis* (Senut et al. 2001) and its age, more than 6 MYA, palaeontologically, refutes any estimates that have placed the divergence between *Pan* and *Homo* at any younger date. *Orrorin* had upright gait and considering its advanced morphological distinction it is likely that the divergence between *Pan* and *Homo* took place much earlier than the age of the *Orrorin* fossil itself.

Based on analyses of mtDNA the oldest divergences of modern humans have been estimated to be about 1/30 of the divergence time between *Pan* and ancestral *Homo* (Vigilant et al. 1991). Using a date of 5 MYA for the *Pan-Homo* split hence gives an estimate of about 170,000 years ago (Vigilant et al. 1991; Ingman et al. 2000) as the coalescence time of modern humans (mitochondrial Eve). Providing the method used to calculate the 170,000 years ago estimates (Vigilant et al 1991; Ingman et al. 2000) was correct, the proposed dating of about 10 MYA for the *Pan-Homo* divergence gives an estimate of about 350,000 years ago for the origin of modern humans. That estimate would be consistent with the existence of human populations in Palestine some 200,000 years ago, whereas the traditional dating, 170,000 years ago or less, for the origin of modern humans is inconsistent with these archaeological finds.

Evolutionary biologists have applied a "molecular clock" approach in order to estimate dates of divergence from DNA sequence distance data. The molecular clock approach has been applied to both nuclear as well as mitochondrial DNA sequences. The mitochondrial DNA molecular clock (mtDNA clock) was initially calibrated as a 2% per million years mtDNA clock (Brown et al. 1979). A 2% per million years mtDNA clock predicts that nucleotide substitutions will occur at a rate of 0.01 substitutions per nucleotide site per lineage per million years. Primates became the group on which the original 2% per MYA mtDNA clock was based. It was discovered however that in order to more accurately estimate dates of divergence from DNA sequence distance data under the assumption of a molecular clock, the number of substitutions that have occurred since two sequences diverged must be estimated under an appropriate model of nucleotide substitution.

Arbogast and Slowinski (1998) re-evaluated the calibration of the primate mtDNA clock using a best-fit model termed the gamma-HKY85 model. They determined that for primates, the rate of substitution estimated for the cytochrome *b* gene under the best-fit model (i.e. the gamma-HKY85 model) was 0.0278 and 0.0252 substitutions per site per lineage per million years, respectively, or more than 2.5 times faster than the rate predicted by a 2% per million years mtDNA clock.

On the scale of evolutionary time, the divergence of *Pan* and *Homo* and the time of the origin of modern humans are very recent. If one examines the rate of evolution of the cyt b gene in primates, it is logical to conclude that one would see only a very small number of nucleotide differences when comparing the cyt b sequences of recently diverged primates. Obviously, the more recent the divergence (example *Pan* and *Homo*) the more scarce the nucleotide differences become. If one considers the time of origin of modern humans and the time of possible divergence of multiple *Homo* lineages, the number of nucleotide differences becomes even more scarce simply because not enough time has elapsed for nucleotide substitutions to accrue.

When comparing the 308 bp region of cyt b, the number of nucleotide difference between human cyt b and chimpanzee cyt b is five nucleotides. When one considers that the divergence of *Pan* and *Homo* was likely as far back as 10 MYA, a difference of five nucleotides over this region of cyt b is relatively small. If a cryptid bipedal primate species does exist today that is closely related to modern day human, it is very likely to have shared common ancestry with modern humans much more recently than 10 MYA. An estimated time of shared ancestry between the lineage that became modern humans and other *Homo* lineages is probably as recent as 2 MYA (citation).

Evidence exists that as recently as 30,000 years ago, multiple species of *Homo* (i.e. *Homo sapiens* and *Homo neanderthalensis*) inhabited the earth simultaneously (Krings et al. 1997; Green et al. 2008). It has long been the belief that other species of *Homo* became extinct over time leaving *Homo sapiens* as the sole species of *Homo* inhabiting the earth. If another species of *Homo* still exists today, it would be more closely related to modern humans (i.e. *Homo* sapiens) than chimpanzees (*Pan troglodytes*) are related to modern humans. With that in mind, one would expect this degree of relatedness to be directly reflected at the genotypic level. One would therefore expect to find fewer nucleotide differences in the comparison of orthologous genes of two species of *Homo* than in the comparison of orthologous genes of *Homo* and *Pan*.

In this analysis, we did not see any nucleotide differences between human cyt b sequence and the cyt b sequence obtained from hair samples derived from alleged cryptid primate species, for the 308 bp region of mitochondrial cyt b examined. If another species of *Homo* still exists alongside modern humans, based on the evolutionary time line and nucleotide substitution rate (according to the mtDNA molecular clock) described earlier, it would not be surprising to find that the genomic DNAs of the two are nearly 100% identical.

Recently, mtDNA extracted from the remains of a tissue sample that was allegedly derived from a creature fitting the description of a cryptid bipedal primate species, produced DNA sequence data that when analyzed showed a single nucleotide difference in a region of sequence of a mitochondrial gene that was not seen in the corresponding human mitochondrial gene sequence (Curt Nelson personal communication). A single nucleotide substitution such as this in a similar comparative mitochondrial gene sequence analysis is consistent with our finding.

As mentioned, we cannot totally rule out contamination or misidentification of hair samples collected in the field. However, the number of samples analyzed along with the fact that carefully run positive and negative control samples were incorporated into the experimentation substantiates that the DNA sequence results that were obtained are credible and authentic.

Whereas the phenotype of another *Homo* species may be different from that of *Homo sapiens*, there may be only a miniscule difference between the genotypes of the two. In this case, the number of nucleotide substitutions seen may not be as critical as the location of these substitutions in the genome or the particular genes where the substitutions are found. For example, a change in the third position in a codon of a gene carries potentially much less impact on the expression of that gene than a change occurring in the first or second position.

The 308 bp region of mitochondrial cyt b is but a small "snapshot" of the entire human genome. The data obtained from this analysis provides a limited amount of information as to whether the hair sample is derived from a cryptid primate species that is a species of *Homo*. However, it raises legitimate questions that warrant further exploration of this subject at the molecular level. In order to determine if these hair samples or other tissue samples that may come along are from a cryptid primate species that is a species of *Homo,* a much larger portion of the genome must be analyzed. This would involve the sequencing of additional genes and the collection of a great deal more DNA sequence data either from the mitochondrial genome or the nuclear genome. Of particular interest would be the collection of sequence data from the noncoding regions of the nuclear genome such as introns, pseudogenes, untranslated regions and 5' and 3' flanking regions of genes. These regions of the genome are much less constrained by selective pressure and incur mutations leading to nucleotide substitutions at a much greater rate than the coding regions of genes. These regions of the genome potentially contain a plethora of information that could be used to verify the existence of individuals representing a lineage of Hominids.

Conclusions

The community of scientists and other researchers who continue to participate in research efforts toward the discovery and identification of a living cryptid primate species continues to grow and become more scientifically and technologically savvy. However, it appears that many of the research teams and programs remain too fragmented or isolated within their realms. In order to augment their research endeavors, more collaborative efforts need to be established in order to put in place more well defined methods of collection and handling of samples and to share data, ideas and technologies more efficiently.

The research involved with the search for a living cryptid primate species is intriguing, but presents obstacles that are challenging. Regardless of ones own personal beliefs, one

must always keep an objective opinion as a researcher and follow proper scientific methodologies and protocols in order to obtain samples and data that have authenticity and that are respected.

If it is discovered one day that there indeed exists a living cryptid primate species, it will likely change many beliefs extending from science to religion. It will alter many of the current beliefs regarding primate evolution, human evolution and possibly the origin of mankind itself. This is a scenario that we must be prepared to possibly face.

Bibliography

Anderson, S., A.T. Bankier, B.G. Barrel, M.H.L. de Bruijn, A.R. Coulson and J. Drouin et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-74.

Arbogast, B.S. and J.B. Slowinski. 1998. Pleistocene speciation and the mitochondrial DNA clock. *Science* 282: 1955.

Coltman, D. and C. Davis. 2006. Molecular cryptozoology meets the sasquatch. *Trends in Ecology and Evolution* 21: 60-61.

Green R.E., *et al*. 2008. A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. *Cell* **134** (3): 416-426.

Ingman, M., H. Kaessmann, S. Paabo and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408: 708-713.

Irwin, D.M., T.D. Kocher and A.C. Wilson. 1991. Evolution of the cytochrome b gene in mammals. *Journal of Molecular Evolution*. 32: 128-144.

Kocher, T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Paabo, F.X. Villablanca and A.C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceeding of the National Academy of Sciences USA* 86: 6196-6200.

Krings M., Stone A, Schmitz R.W., Krainitzki H., Stoneking M. and S. Pääbo. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* **90** (1): 19–30.

Meldrum, D.J. 2007. Ichnotaxonomy of giant hominid tracks in North America. *Cenozoic Vertebrate Tracks and Traces: New Mexico Museum of Natural History and Science Bulletin 42*. 225-231.

Meyer, A. and A.C. Wilson. 1990. Origin of tetrapods inferred from their mitochondrial DNA affiliations to lungfish. *Journal of Molecular Evolution* 31: 359-364.

Milinkovich, M.C., A. Caccone and G. Amato. 2004. Molecular phylogenetic analyses indicate extensive morphological convergence between the “yeti” and “primates”. *Molecular Phylogenetics and Evolution* 31: 1-3.

Miyamoto, M.M. and S.M. Boyle. 1989. The potential importance of mitochondrial dNA sequence data to eutherian mammal phylogeny. In: Fernholm B., Bremer K., Jornvall H. (eds) The hierarchy of life. Elsevier, Amsterdam, pp 437-450.

Sarich, V.M. and A.C. Wilson. 1967. Immunological time scale for hominid evolution. *Science* 158: 1200-1203.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A.C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503-1507.

Wallace, D.C., M.T. Lott, J.M. Shoffner and M.D. Brown. 1992. Diseases resulting from mitochondrial DNA point mutations. *Journal of inherited metabolic disease* 15: 472-479.

Zeviani, M., V. Tiranti , and C. Piantadosi. 1998. *Medicine* (Baltimore). Jan;77(1):59-72. Review.