Novel North American Hominins, Next Generation Sequencing of Three Whole Genomes and Associated Studies

Melba S. Ketchum[1] (Corresponding author)
Patrick W. Wojtkiewicz[2]
Aliece B. Watts[3]
David W. Spence[4]
Andreas K. Holzenburg[5]
Douglas G. Toler[6]
Thomas M. Prychitko[7]
Fan Zhang[8]
Ray Shoulders[1]
Ryan Smith[1]

[1] DNA Diagnostics, Nacogdoches, TX 75965
[2]North Louisiana Criminalistics Laboratory, Shreveport, LA 71101
[3]Integrated Forensic Laboratories, Inc., Euless, TX 76039
[4]Southwestern Institute of Forensic Sciences, Dallas, TX 75207
[5]Texas A&M University, Microscopy & Imaging Center, Department of Biology and Department of Biochemistry & Biophysics, College Station, TX 77843-2257
[6]Huguley Pathology Consultants, P.A., Ft. Worth, TX 76115
[7]Helix Biological Laboratory, Detroit, Michigan, 48202
[8]UNT Center for Human Identification, University of North Texas Health Science Center, Fort Worth, TX 76107

Dr. Melba S. Ketchum, Director
DNA Diagnostics, Nacogdoches, TX 75965
E-mail: msketchum@dnadiagnostics.com

Abstract

One hundred eleven samples of blood, tissue, hair, and other types of specimens were studied, characterized and hypothesized to be obtained from elusive hominins in North America commonly referred to as Sasquatch. DNA was extracted and purified from a subset of these samples that survived rigorous screening for wildlife species identification. Mitochondrial DNA (mtDNA) sequencing, specific genetic loci sequencing, forensic short tandem repeat (STR) testing, whole genome single nucleotide polymorphism (SNP) bead array analysis, and next generation whole genome sequencing were conducted on purported Sasquatch DNA samples gathered from various locations in North America. Additionally, histopathologic and electron microscopic examination were performed on a large tissue sample.

The mtDNA whole genome haplotypes obtained were uniformly consistent with modern humans. Of the 20 whole and 10 partial mitochondrial genomes sequenced, 16 diverse haplotypes were found suggesting that these hominins did not originate in a single geographic location.

In contrast, consistent, reproducible, novel data were obtained when nuclear DNA was amplified utilizing various platforms. Nuclear DNA obtained from Sasquatch samples produced novel SNPs, off ladder alleles on human STRs, retained human sequence interspersed with novel sequence, and whole genome SNPs that fell outside the human threshold. Three of the Sasquatch samples were subjected to next generation whole genome sequencing, each of which independently yielded high quality complete genomes.
Analysis of preliminary phylogeny trees derived from supercontigs generated from all three samples showed homology to human chromosome 11 reference sequence hg 19, and to primate sequences. The totality of the DNA evidence suggests the Sasquatch nuclear DNA is a mosaic comprising human DNA interspersed with sequence that is novel but primate in origin.

In summary, our data indicates that the Sasquatch has human mitochondrial DNA but possesses nuclear DNA that is a structural mosaic consisting of human and novel non-human DNA.

Keywords
Species Identification
Human DNA Identification
Forensic Hair Analysis
Electron Microscopy
Next Generation Whole Genome Sequencing
Mitochondrial DNA

Introduction

Eyewitness reports of encounters with tall, hairy hominins commonly called Sasquatch have been recorded for thousands of years; from Sumerian stories of wild hairy men (Figure 1) to reports in North America recorded in the early 1800s by Reverend Elkanah Walker and Daniel Boone[1-3].

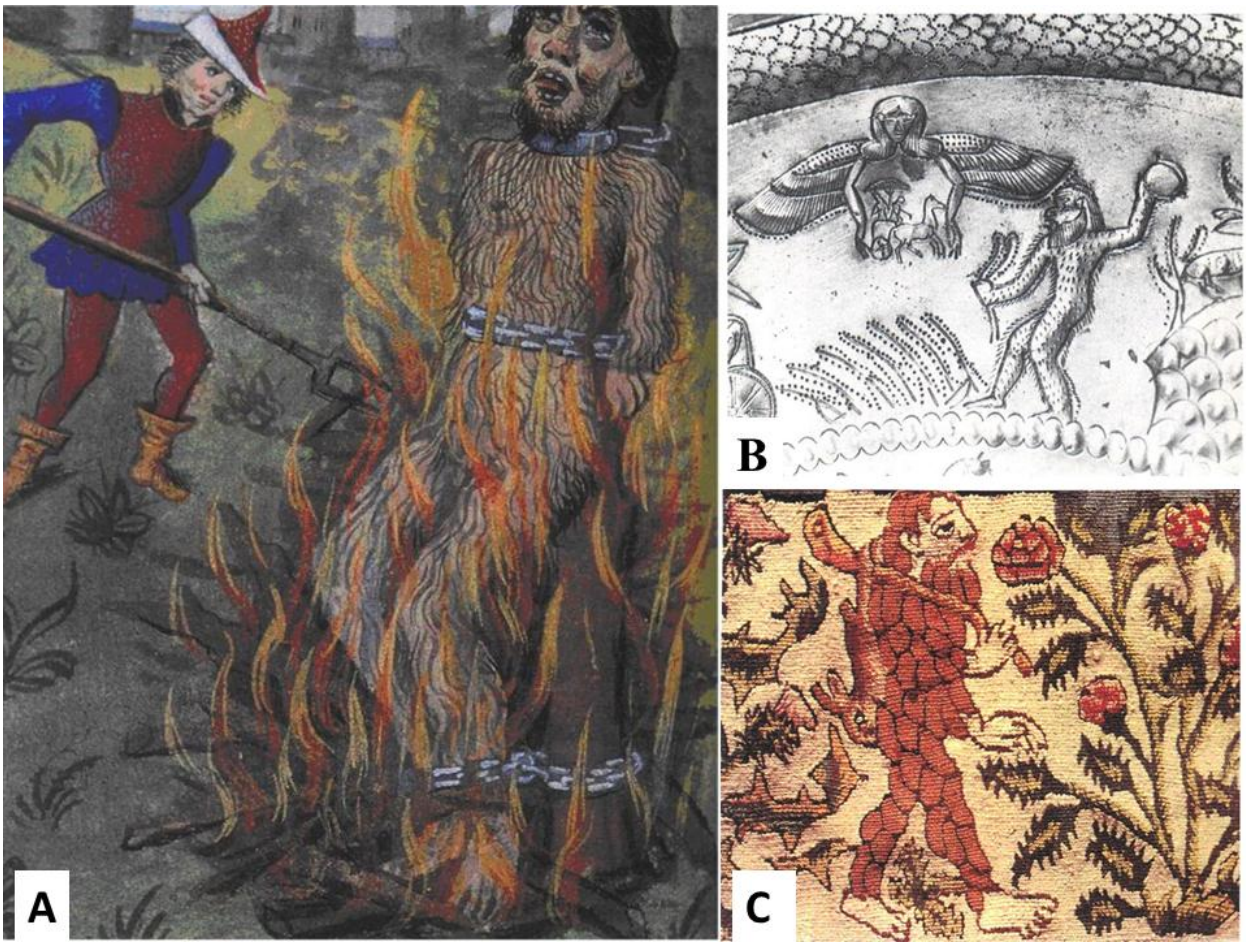

Figure 1 Sasquatch recorded through history in art. (A) A French illustration circa 1100 from *Roman d'Alexandre* (B) Phoenician plate circa 700 BC showing a Wildman attacking (C) Tapestry in a museum in Saint-Jean-Les-Saverne.

One North American investigative group has publicly stated that they have documented 30,000 reports of sightings and other evidence for the existence of these creatures. These putative hominins are reported to walk bipedally in a linear manner that is distinct from humans (Figure 2).

Figure 2 Linear trackway.

Eyewitnesses report that the Sasquatch are 8 to 12 feet in height, are covered in hair with arms longer than those of modern humans and possess sloped anterior craniums with short, thick cervical regions reminiscent of Neanderthals. Eyewitness accounts also describe an opposable thumb with hands and feet that are large in comparison to body size.

The above commonly reported traits, as well as other scientific evidence lending credence to the existence of Sasquatch, have been thoroughly researched and documented in both books and in peer reviewed manuscripts.[4-13] Although there have been thousands of sightings, footprint casts and other circumstantial evidence of Sasquatch across the modern world for centuries, incontrovertible proof of their existence has been elusive. The failure to present a deceased individual or skeletal remains has exacerbated scientific skepticism towards the circumstantial evidence.

There are reports from witnesses who allegedly live in close proximity to, and interact with, Sasquatch family units[14]; however, in the absence of additional evidence, most of these reports have been dismissed. These witnesses and other eyewitnesses further report that the Sasquatch also build stick structures in the forests, some of which are thought to be shelters while others have unknown meanings. Notably, some of the samples in the present study were submitted by individuals claiming to have obtained them from the Sasquatch with which they have had long term interaction. Sample 168 in this study was retrieved from a stick shelter as seen in Figure 3.

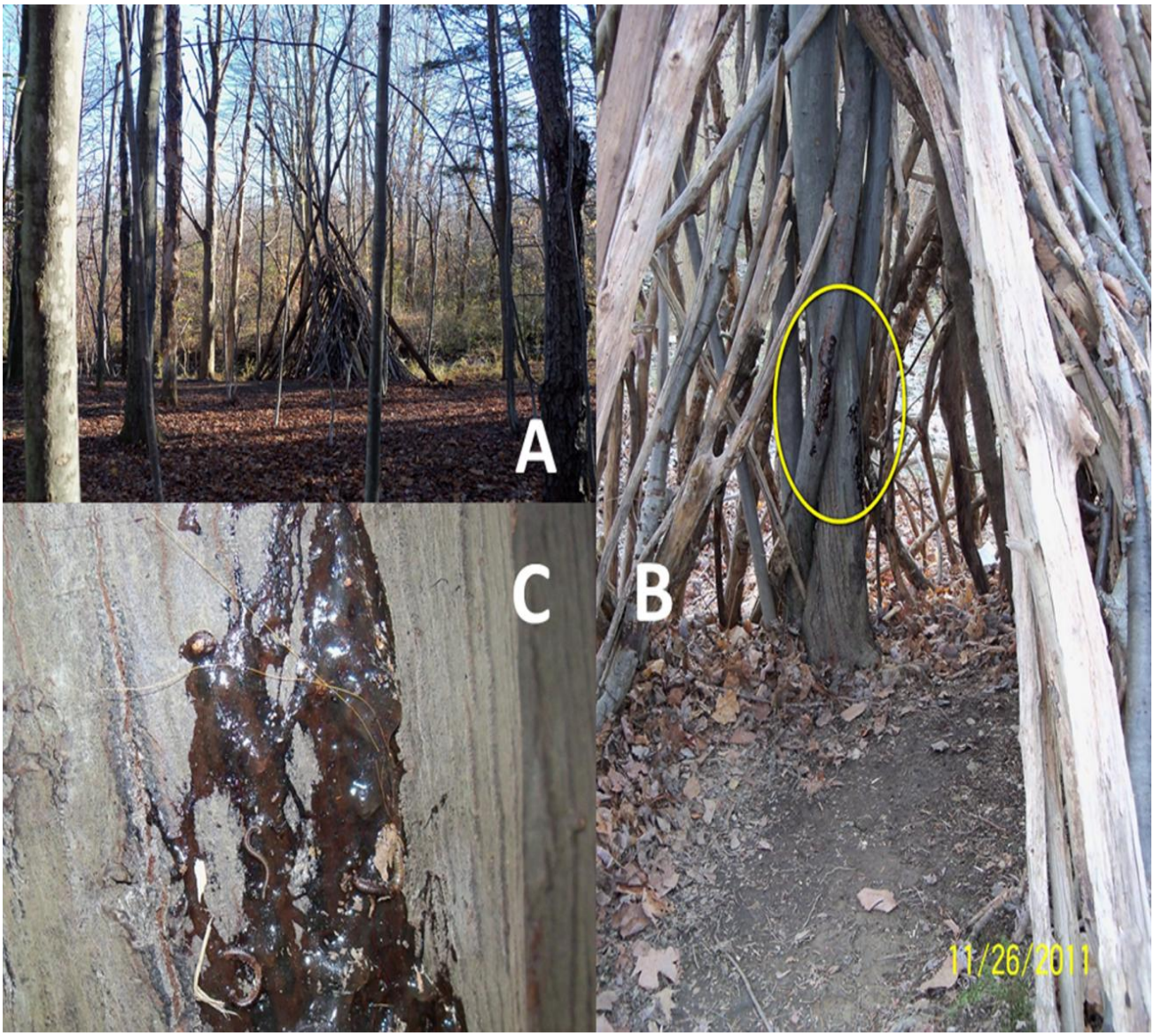

Figure 3 (A) Stick shelter with hair trap (B) Hair trap on center pole (C) Close-up of hair Sample 168 stuck to the glue of the hair trap prior to collection.

DNA analysis of hair and tissue samples, modern video image capture and audio recording technology, morphological examination of hair and tissue could cumulatively strengthen support for the existence of these unknown hominins. Some photographic evidence also exists such as Figure 4 is a reddish brown Sasquatch sleeping in the forest and is Sample 37 in the study. Video of the same Sasquatch is seen in Supplementary Movie 1 where her respirations are counted at only 6 per minute.

Figure 4 Reddish brown Sasquatch juvenile sleeping in the forest, Sample 37.

During the present five year study, approximately one hundred and thirteen separate samples of hair, blood, mucus, toenail, bark scrapings, saliva and skin with hair and subcutaneous tissue attached were submitted by dozens of individuals and groups from thirty four separate hominin collection sites around North America (Table 1).

Here we report the morphological and histopathologic examination, whole mitochondrial and nuclear DNA sequencing and analysis, and electron microscopic studies of DNA extracted from fresh hominin tissue.

Overall the data are suggestive of an unusual contemporary hominin with mitochondrial DNA consistent with modern humans but showing marked anomalies in the nuclear DNA. These findings suggest the existence of a novel contemporary indigenous North American hominin.

Materials and Methods

The materials and methods are found as Supplementary Data, Materials and Methods S1 in the Supporting Information due to the length of the manuscript.

Supplementary Sequence Information

The raw sequence data utilized in this manuscript from the Sasquatch samples has been provided in the Supplementary Data S3, S4, S5 and S6 in FASTA format. These files were added as supplemental because the sequences were not able to be uploaded to a GenBank® because of their lack of taxon according to GenBank personnel.

Examination of Novel Hair Samples

A wide variety of hair shaft profiles were observed using stereomicroscopy and light microscopy (Figure 5) with most exhibiting a stiff, wiry texture and wavy, curved appearance with rounded tapered tips, consistent with aged hairs that had not been cut. The full length of the hairs was approximately 15 cm (Figure 5A) and diameters ranged from 80 to 110 µm. (Human head hairs typically range from 55 to 100 µm in diameter) [15-19].

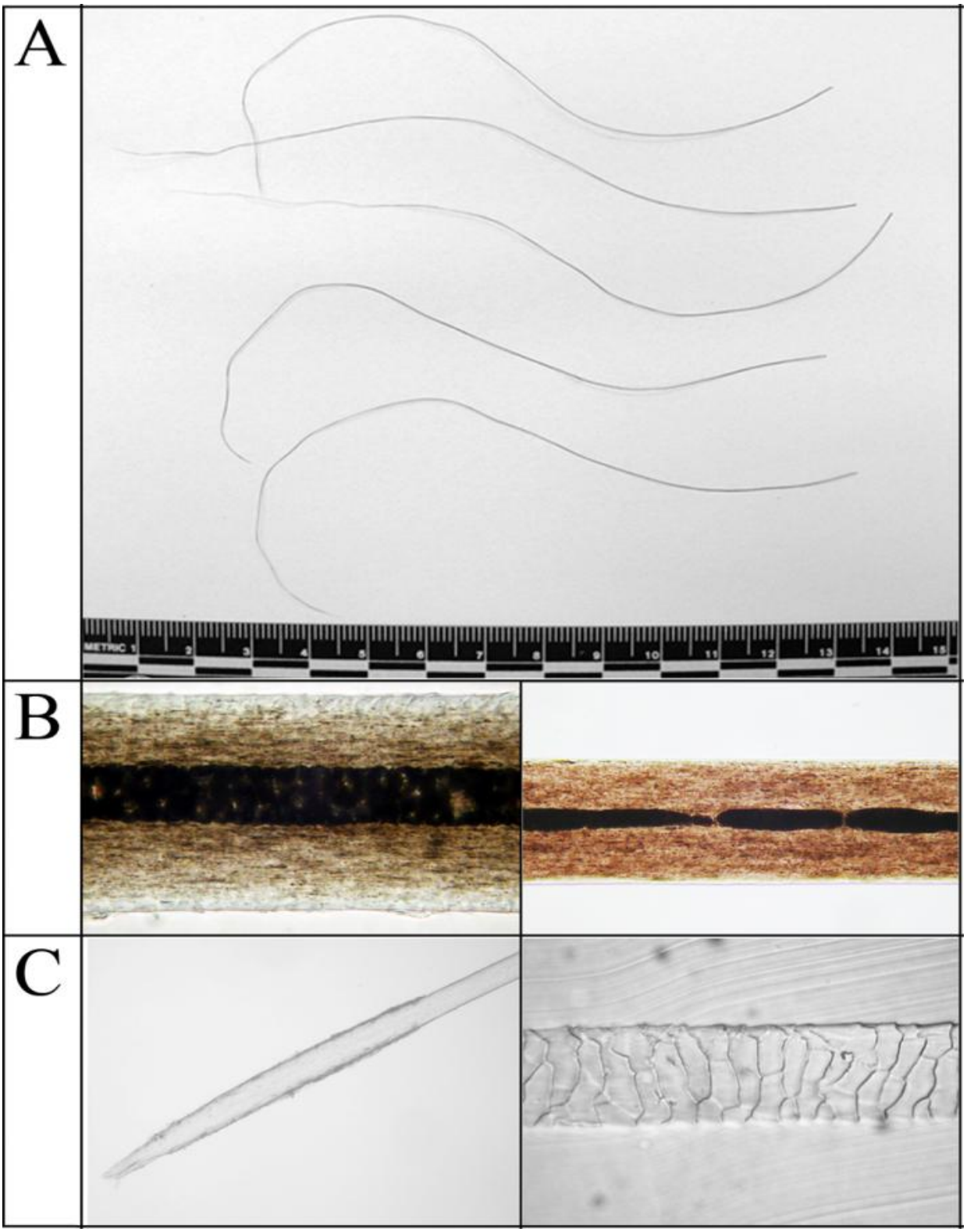

Figure 5 Results from microscopic hair examination. (A) Gross morphology and length (Sample 33). (B) Light microscopy showing medulla and internal hair structure: Left- Unknown hair (Sample 26) from unknown species, Right- human hair, both at 400X with a

camera digital zoom factor of 5.7. (C) Left- Hair with spade shaped root, (Sample 18),; Right, (Sample 26), Imbricate Cuticle.

The medulla and root were found to be the two most discriminating characteristics of the microscopic examination. Most of the novel hairs had medullary structures and diameter ratios that were clearly distinct from human hairs. Even though a variety of medullary structures were observed, the micrographs in Figure 5B depict those most commonly encountered. Most of the novel hairs had elongated roots with a somewhat "spade" shape, which is a feature of some animal hairs but is typically not seen in human hairs (Figure 5C, left). Human hairs exhibit characteristic uniform imbricate scale patterns of the cuticle. Several different cuticle patterns were observed on the submitted samples. The hairs exhibited wide imbricate scale patterns proximally that transformed to close imbricate patterns distally. These patterns are distinctly non-human in appearance (Figure 5C, right). Most of the submitted hairs were not microscopically consistent with any of the hairs from the reference collection of common animal hairs that included human, cat, dog, cow, horse, deer, elk, antelope, moose, sheep, fox, bear, coyote, wolf, rat, mouse, monkey, beaver, squirrel, llama and others.

The hairs were evaluated for DNA testing by observing the presence or absence of hair roots and adherent tissue material. Hairs with apparent translucent tissue material and/or anagen or catagen phase roots were considered as suitable candidates for nuclear DNA (nuDNA) testing. Hairs with telogen phase roots without tissue or hairs lacking roots or tissue were considered suitable for mitochondrial DNA (mtDNA) testing. Hairs that exhibited non-human microscopic characteristics and that did not match any known animal species were recommended for DNA testing.

Collection and Classification of Hominin Samples

DNA Diagnostics of Nacogdoches, Texas, received samples of hair, toenail, tissue, blood, mucus, scratched tree bark and saliva claimed by submitters to be from an unknown and previously undescribed hominin. The samples were collected from research and sighting locations in 14 states and two Canadian provinces. They were treated as forensic samples and catalogued to maintain an appropriate chain of custody. Table 1 and Figure 6 summarize the locations, the sample collectors and their methodology.

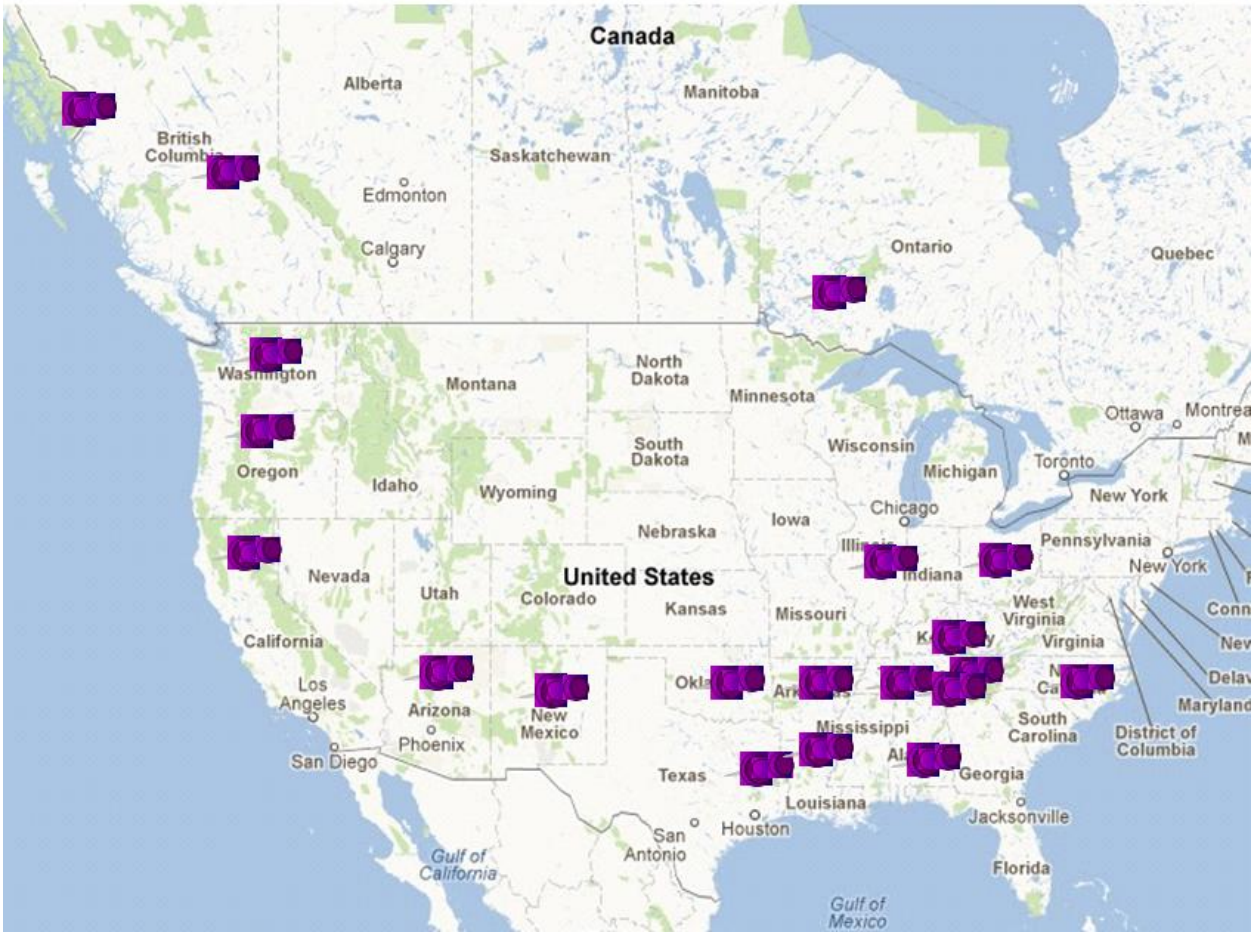

Figure 6 Geographic locations where samples were collected.

Samples were subjected to a preliminary screen by utilizing eyewitness interview information, visual and histological examination, and DNA testing. Those samples that failed to present novel visual structure or that matched a documented species were removed from the study.

Prevention of DNA Contamination by Forensic Methodologies

Throughout this project exhaustive precautions were taken to minimize or eliminate contamination. The samples submitted in this study were either dried or fresh without degradation. The DNA was extracted in a clean room using forensic science procedures that minimized contaminant DNA in the samples while maximizing DNA recovery. Samples were also split and DNA was isolated in two separate laboratories to ensure that sample integrity was maintained.

To further prevent contamination, hair samples were thoroughly vortexed using ethyl alcohol and double distilled DNAse free water in order to wash away any DNA not associated with the sample. The samples were extracted in small batches, prior to any handling of control DNA. The human control DNA was extracted in a separate location specifically utilized for the extraction of high yield DNA samples. Blank extraction controls were tested along with the hominin samples

to assure that there was no contamination during the extraction. Half of the samples were extracted robotically thus minimizing handling by laboratory personnel. All specimens were subjected to testing designed to show a secondary contributor should any contamination occur. One evidentiary specimen (purported dried urine) did yield a mixed profile and was removed from the study.

Control DNA was obtained from the majority of the submitters and was profiled using Promega PowerPlex® 16 [20]. All submitters tested yielded complete profiles. In contrast, when the hominin samples were tested using Promega PowerPlex® 16, partial profiles were evident in almost all cases. Though not all samples were accompanied by DNA from the submitter, these data showed that those samples that were provided were not contaminated by their submitters since the submitters' profiles excluded as being a contributor to the hominin profiles/partial profiles. Additionally, the hominin profiles/partial profiles did not have any extra STR alleles that would indicate a secondary contributor or contamination by a second individual.

If two or more profiles were present in a sample during mtDNA sequencing around mutation loci, superimposition of the bases will yield an appearance similar to heteroplasmy. In a contaminated sequence, a larger than expected number of heteroplasmic appearing bases would be seen. No heteroplasmy was observed in the mitochondrial sequencing. This further supports the fact that none of the samples utilized in this study were contaminated.

Determination of DNA Quality

Throughout this study, close monitoring of DNA quality and yields was necessary to determine if any PCR failures could be attributed to DNA degradation. Therefore, we utilized a laboratory that routinely extracted DNA from animal hair and hair root tags. In the experience of this laboratory, hair samples constitute a very reliable source of DNA. Almost all large animal breed registries utilize plucked hair samples as their primary source of DNA for their parentage testing programs. One horse registry alone processes over 100,000 hair samples per year. These hair samples are archived and are viable for many years after they are submitted. Therefore the hominin samples were well within our expectations as far as nuclear DNA yield and quality. Only samples with verified follicular material were extracted. After extraction, yield gels with 3 µL of the extracted DNA were utilized to determine if there was DNA present and whether it was degraded (Figure 7). Hairs without tissue or root material did not yield DNA in this study.

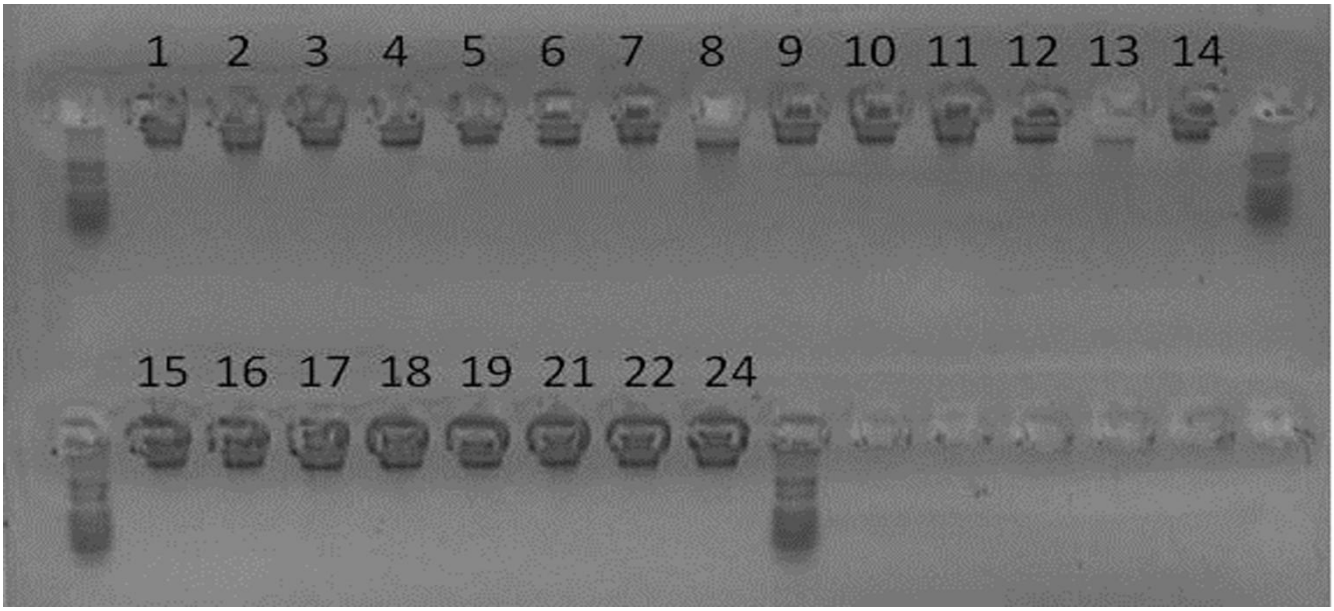

Figure 7 Yield gel for QC and quantification of extracted samples.

A small sample of skin with underlying structures from Sample 26 was submitted to the Texas Veterinary Medical Diagnostic Laboratory at Texas A&M University for the purpose of evaluating the sample for degradation and structure. Subsequently, the slides generated from this examination were submitted to Huguley Pathology Consultants for further evaluation. A detailed report of the histopathologic findings was generated from the second examination. The report confirmed the Texas A&M findings and revealed that there was little degradation observed and that the structures in the sample were visible and could be reviewed. There was no pathology present in sample 26, however the histology was deemed inconsistent with human skin. Examination revealed lesser numbers of eccrine glands and even sebaceous gland/pilosebaceous gland units than normally seen in human skin. Abnormalities such as abortive hair shafts and various alopecias were detected and hair follicle addition or extra follicles, clustered and deeper in the dermal region were noted. The clustering of follicles at a deeper level in the dermis than where most skin appendages usually occur was unusual and not generally associated with hair follicle loss as is seen with alopecia. (Figure 8, Supplementary Data 1)

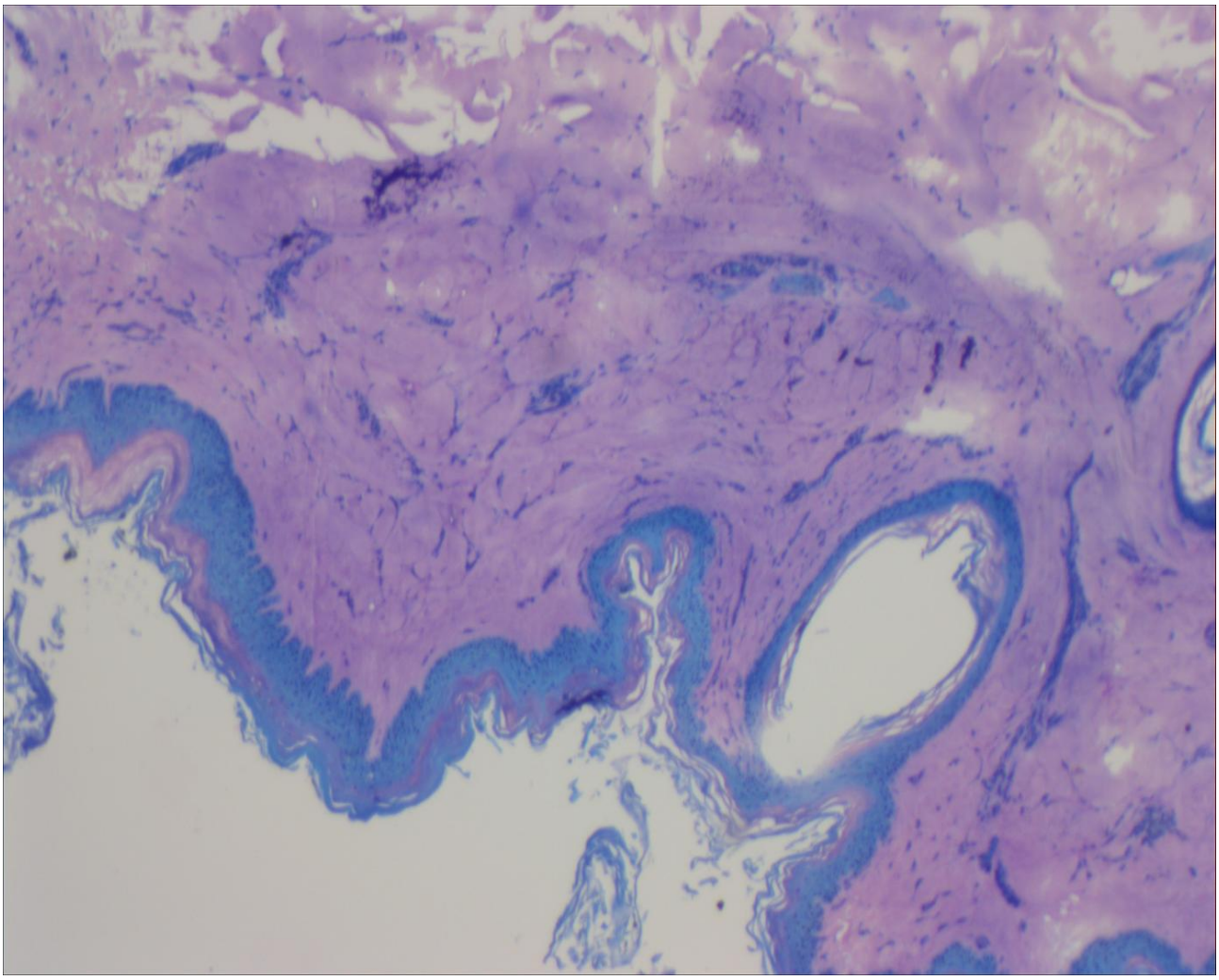

Figure 8 Histopathology slide of Sample 26.

Importantly, histological examination of the skin and tissue sample 26 confirmed that the cells were intact with few bacteria. The absence of contamination or degradation indicates that DNA extracted from this tissue was more than adequate for subsequent DNA sequencing.

Screening of the Hominin DNA Samples

Since the amount of hair available in the samples was finite, we opted to begin this project by screening tissue samples that had larger quantities of DNA. Universal mitochondrial DNA cytochrome b primers for species determination as well as universal mammalian primers designed for species identification in the hypervariable region 1 were utilized[21-28]. All 111 screened samples revealed 100% human *cytochrome b* and hypervariable region 1 sequences with no heteroplasmic bases that would indicate contamination or a mixture. These samples were then sent out to another laboratory for mitochondrial whole genome sequencing.

The hominin samples were also amplified using Promega PowerPlex® 16, a nuclear STR DNA multiplex platform that is specific for humans and higher primates. The submitters and the blank extraction controls were amplified to determine the presence of non-human DNA and to obtain a profile to compare with hominin DNA samples. PowerPlex® 16 amplification of the hominin samples yielded only partial profiles with off-ladder alleles while amplification of DNA from the submitters yielded complete profiles and the control blank indicated no amplification (Figure 9).

The preliminary screening, therefore, established that Sasquatch samples were not contaminated with submitters DNA and that Sasquatch samples comprised some human mitochondrial DNA but with nuclear DNA anomalies. Once it was established that the Sasquatch nuclear DNA did not conform to human DNA, the remaining samples were extracted.

Testing and Results

Mitochondrial complete genome and HV region sequencing results:

It was decided to sequence the entire mitochondrial genome from those hair samples with relatively greater yields of DNA and to sequence HV1 from those samples that yielded low quantities of DNA. The samples were sent to Family Tree DNA after initial screening for HV1 and cytochrome b at DNA Diagnostics The source of the samples was withheld from Family Tree DNA.

DNA samples were successfully amplified and sequenced across the whole mitochondrial genome and the HV1 locus using both human specific and universal primers. The sequences that were subjected to BLAST searches in GenBank®[40] showed consistent homology with human haplotypes. No mitochondrial DNA homology with apes, Neanderthal or Denisova cave sequences were found[21-39]. Mitochondrial DNA testing results are shown in Table 2 and sequences in Supplementary Data 2.

Table 2—List of Mitochondrial DNA Haplotypes

Sample Number	**Haplotype**
26	H1a, one novel SNP
1, 2, 12, 36,	T2b
28	H1
35	H10
29, 44, 46, 138	H2a2
39b,41, 42,	T2

Sample Number	Haplotype
43	
37	H12
11	A6L2c
31	LOd2a
38	V2
24	H1s
4,37	H3
33. 95	H
140,168	D
81	C
71,117,118	L3d
8, 139, 18*	HV2 (human specific) only
46-137	Partial HV1 (human specific) screened only

All 16 haplotypes from 20 completed whole mitochondrial sequences and 10 partial mitochondrial genomes have indicated 100% homology with human mitochondrial sequences without any significant deviation. Of the 16 haplotypes, most were European or Middle Eastern in origin. African and American Indian haplotypes were also observed. Those samples that did not give enough viable sequence to obtain a complete genome usually yielded sufficient data to delineate a haplotype from the mitochondrial hypervariable region or at least a human HV2 sequence. With the wide variety of haplotypes in the study and especially with the majority of the haplotypes being European or Middle Eastern in origin, migration into North America by these hominins may have occurred previous to the migration across the Bering land bridge. This previous migration is supported by the Solutrean Theory[41-42]. Three phylogenetic trees were subsequently constructed using the mitochondrial sequences obtained from samples 26, 31, and 140 using sequences generated at Family Tree DNA (Supplementary Figures 1, 2, 3). The trees were consistent with human mitochondrial sequences in Genbank®. Given the mitochondrial DNA results across all samples, nuclear genome testing was undertaken in order to address paternal origins of the DNA samples.

Nuclear DNA Analysis:

Amelogenin:
Both DNA genotyping and sequencing technologies were used in the testing of the *amelogenin* gene. The genotyping of the *amelogenin* locus produced the most consistent results across the samples tested. The DNA samples yielded four types of results: *XX, XY, Y* and null. The dropout of the *X* amplicon was the most significant of the findings observed with the STR genotype analysis of *amelogenin*. (Figure 9, Table 3) This dropout was reproduced in several individual samples and was repeatable both in the multiplex of PowerPlex® 16 and the analysis of the STR locus, so it is unlikely to be an experimental artifact due to low quantity or degraded DNA (Table 3). *Y* peak height as measured in the electropherograms further supported that there was more than adequate DNA present and that allelic dropout should not have occurred. The repeatability

and number of samples exhibiting the *X* dropout is inconsistent with what would be expected with normal human allele dropout[43-46]. It is noteworthy that *Amel*^{X} allele dropout occurs in significant numbers of the unknown samples yet seldom occurs in normal human testing.

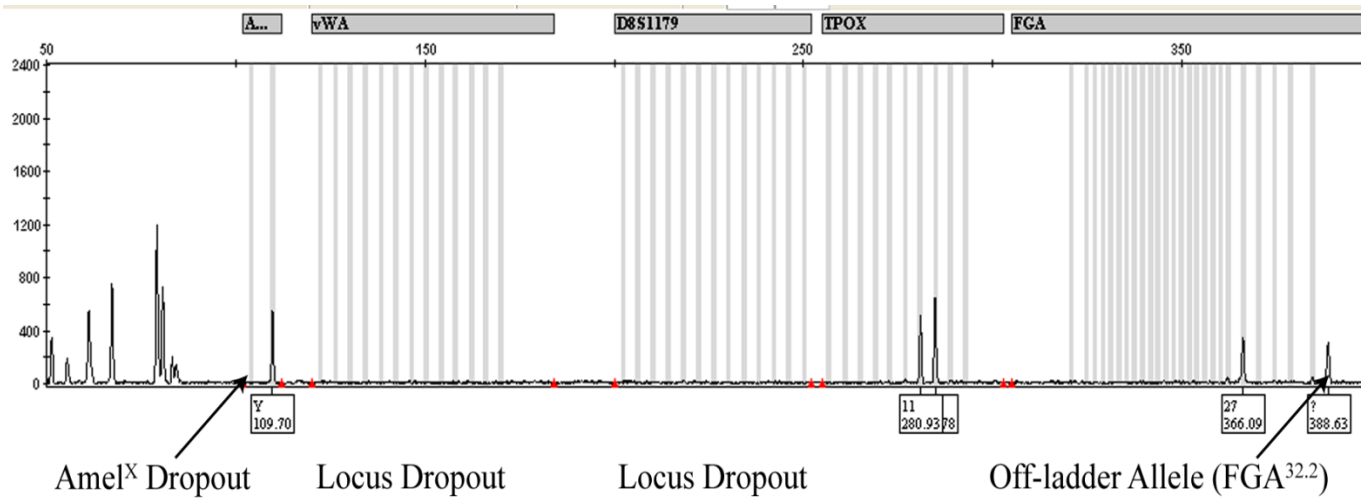

Figure 9 PowerPlex® 16 electropherogram (yellow channel) showing Amel ^{X} dropout, locus dropout, and unusual allele variants.

Table 3. Amelogenin STR results

Sample Numbers	Amelogenin Results
3, 7, 8, 10, 11, 12, 17, 26, 27, 28, 35	XY
2, 37, 21	XX
4, 9, 22, 23, 34, 36, 38	Y
1, 5, 6*, 13, 14, 15, 19, 24, 29, 30, 33, 130	Ø/Failed

* Ø, other PowerPlex® 16 markers amplified

Sequencing exons of the *amelogenin* gene also gave novel results. Some of the samples indicated a normal human *Amel*^{X}, yet others failed completely to amplify. Further, when the *Amel*^{Y} locus was sequenced across Exons 1, 2, 4/5 and 8, varied results were obtained, with no samples successfully amplifying and sequencing across all five exons (except the human controls). The resulting sequences ranged from totally non homologous matches, not found in Genbank® after numerous BLASTs (including dissimilar sequence BLASTs) to novel SNPs and even failure to sequence (Table 4). These findings were consistent at other loci tested. The documented anomalies came from DNA samples that yielded long sequences with pristine electropherograms at other loci including at least one *Amel*^{Y} exon. Notably, this indicated that the DNA was of high quality and that degraded DNA was not responsible for the anomalies.

Table 4. Amelogenin Sequencing Results

		Amelogenin Exons				
Sample Number	**Amel^{X}**	**1**	**2**	**3[1]**	**4/5**	**8**
87,95	H	*	*	*	*	*
26	*	*	UNK	*	H	H
27,130, 36, 39b, 42, 43, 89, 103	*	*	H	*	*	*
28, 83	*	*	H	*	H	*
33	H	*	H	*	H	AGAATATGAGACA GGAACTGA mutated to A/G
35, 38, 49	*	*	H	*	H	H
44	*	*	TTTCAGAACCA TCAAGAAATG mutated to A/G	*	AAATGTTTTAC CTTCTTCTTT mutated to T/T	H
54	*	*	UNK	*	H	H
71	*	*	H	H	H	H
72	*	*	*	*	H-27 variations	*
82	H	*	*	*	*	H
85	*	*	H	*	*	UNK
88, 91, 106	*	*	*	*	H	*
90	H	*	*	*	*	UNK
94, 98	*	*	*	*	*	UNK
96	*	*	UNK	*	*	H
100	UNK	*	UNK[4]	*	*	UNK
106	UNK	*	*	*	*	*
109	*	*	UNK	*	*	UNK
114	*	*	UNK	*	*	*
117	H	*	UNK[4]	*	*	*
99	*	*	H-Chromosome 13[5]	*	H	
122	H	*	H-Chromosome 13[5]	*	H	*
Cont-1[2]	H	*	*	*	*	*
Cont-2[3]	H	H	H	H	H	H

H= Sequence consistent with *Homo sapiens sapiens*
UNK BLAST yielded no matches
* PCR or sequencing failed
[1] Includes PowerPlex® 16 Y-specific
[2] Amp. Control-1 human female
[3] Amp. Control-2 human male
[4] 100, 117 align at Exon 2
[5] 99, 122 align at Exon 2

Promega PowerPlex® 16

The data utilizing Promega PowerPlex® 16 amplification of the *amelogenin* gene is reported separately in this manuscript. All samples that yielded results for PowerPlex® 16 gave only partial profiles with random dropout of alleles, off ladder alleles and/or allele frequencies for some markers inconsistent with those found in the human population (Table 5).

None of the data from PowerPlex®16 analysis in this study showed mixed profiles. DNA from the submitters and laboratory personnel were also amplified with PowerPlex® 16 or already had their PowerPlex® 16 profile on file. All submitters, laboratory personnel and human control DNAs showed complete profiles and were excluded as contributors to the profiles generated from the unknown samples. This, coupled with the lack of mixtures, novel profiles and failures with PowerPlex® 16 further eliminated the possibility of human contamination of the unknown samples.

Melanocortin 1 Receptor Gene (MC1R)

The *MC1R* gene on Chromosome 16 was targeted since hundreds of anecdotal eyewitness reports over the years have described bipedal hominins displaying red hair or having a reddish highlight to black, brown or blonde outer coat coloring[47-59]. Samples 28, 33, 35 and 37 had sufficient DNA extracted and were chosen for *MC1R* locus sequencing. The primers used in the sequencing of *MC1R* were designed by DNA Diagnostics and additional primers by SeqWright to correspond with great apes, humans and Neanderthals. Upon sequencing, it was found that the samples indicated normal human *MC1R* sequence and carried alleles for red hair color in humans. Sample 28 and 37 presented a C/T at base 478 of the *MC1R* control region and Sample 35 exhibited G/C at base 880. (Table 6)

Samples 28, 35 and others were then sent to SeqWright to have the sequences confirmed with the design of new *MC1R* primers. As with other loci analyzed, *MC1R* analysis at SeqWright found partial human sequences in some DNA samples, while others had novel sequences and still others failed to amplify. All human control DNA amplified and sequenced successfully as before.

Myosin 16 Heavy Chain (MYH16):

Human and ape specific primers were designed by DNA Diagnostics and SeqWright respectively for the *MYH16* gene on Chromosome 7. This gene is associated with the sagittal crest in apes and was analyzed in an effort to determine if the unknown hominin was related to apes. Interestingly, all DNA samples that successfully amplified yielded results consistent with human and aligned with the human reference sequence, including a 2 base deletion at bases 36334-5 [60-65]. SeqWright also reported human *MYH16*, Exon 18 sequences on the small number of samples.

Antigen peptide transporter 1(TAP1):

TAP1 is a gene that has been associated with breast cancer, diabetes and several other syndromes including ankylosing spondylitis [66-68]. The *TAP1* SNP, rs9276811, along with its flanking

sequences, were selected for sequencing when preliminary analysis of the Sasquatch samples failed to identify rs9276811 while uniformly identifying it in human controls.
Of the unknown samples, there were those that showed human *TAP 1* sequences, and those that failed even though all of the human controls sequenced. Novel SNPs were found and some amplifications with *TAP 1* primers yielded long sequence of unknown origin. Two sets of 2 questioned samples had long unknown sequences that failed to BLAST even in a dissimilar BLAST search in Genbank®, but aligned with each other (Table 7). Two of these samples, samples 33 and 44, aligned with one another although there were some SNPs that were inconsistent between the two sequences. These two samples were different in hair color and obviously were from different individuals and were retrieved at different times. Samples 33 and 44 were collected from the same general area and the donors are thought to be related. Samples 10 and 43 also aligned, each having different unknown *TAP 1* sequence from samples 33 and 44. Though sample 43 came from the same general area as 33 and 44, sample 10 came from a different location in Washington State.

Table 7. Alignment of several related samples

#	Sample Description	Gender	Amelogenin Exon 3	mtDNA Haplotype	TAP 1
33	WA hairs multicolored	male	Failed	H HV1 only	Unknown, aligns with 44
43	WA hairs brown/white	male	Unknown, aligns with 44	T2	Unknown, aligns with 10
44	WA hairs chocolate	male	Unknown, aligns with 43	H2a2	Unknown, aligns with 33
10	WA hairs blonde	male	Failed	HV2 only	Unknown, aligns with 43

In light of some of the findings being consistent with modern human DNA sequences, the rest of the unknown samples were extracted and submitted for whole genome SNP Bead Array Analysis.

Whole Human Genome SNP analysis:
Twenty-four samples were tested on the whole human genome (2.5 million SNPs) Illumina® Bead Array[69] platform using the Illumina® iSCAN instrument. Of these, in a clear departure from the results obtained with normal human DNA, 100% of the 24 samples failed to meet the human threshold of 95% SNP performance. The results ranged from 53% to 89% SNP performance. In the top 12 performing samples, only 45 SNPs out of the 2.5 million SNPs tested failed across all 12 samples, while simultaneously the human controls all yielded above 95% results on those SNPs.

In an effort to mimic severely degraded DNA that could explain the low SNP matches obtained, one of the human controls submitted along with the unknown hominin samples comprised non sterile blood that was purposely maintained at room temperature in a moist environment for 4 days in an effort to maximize degradation of the sample. Upon visual inspection, hemolysis of the sample had occurred and bacterial contamination, which often correlates with DNA degradation, was seen. An acrylamide gel was loaded with the degraded human sample to assess the degradation and was visualized with ethidium bromide. Smearing was observed (Figure 10).

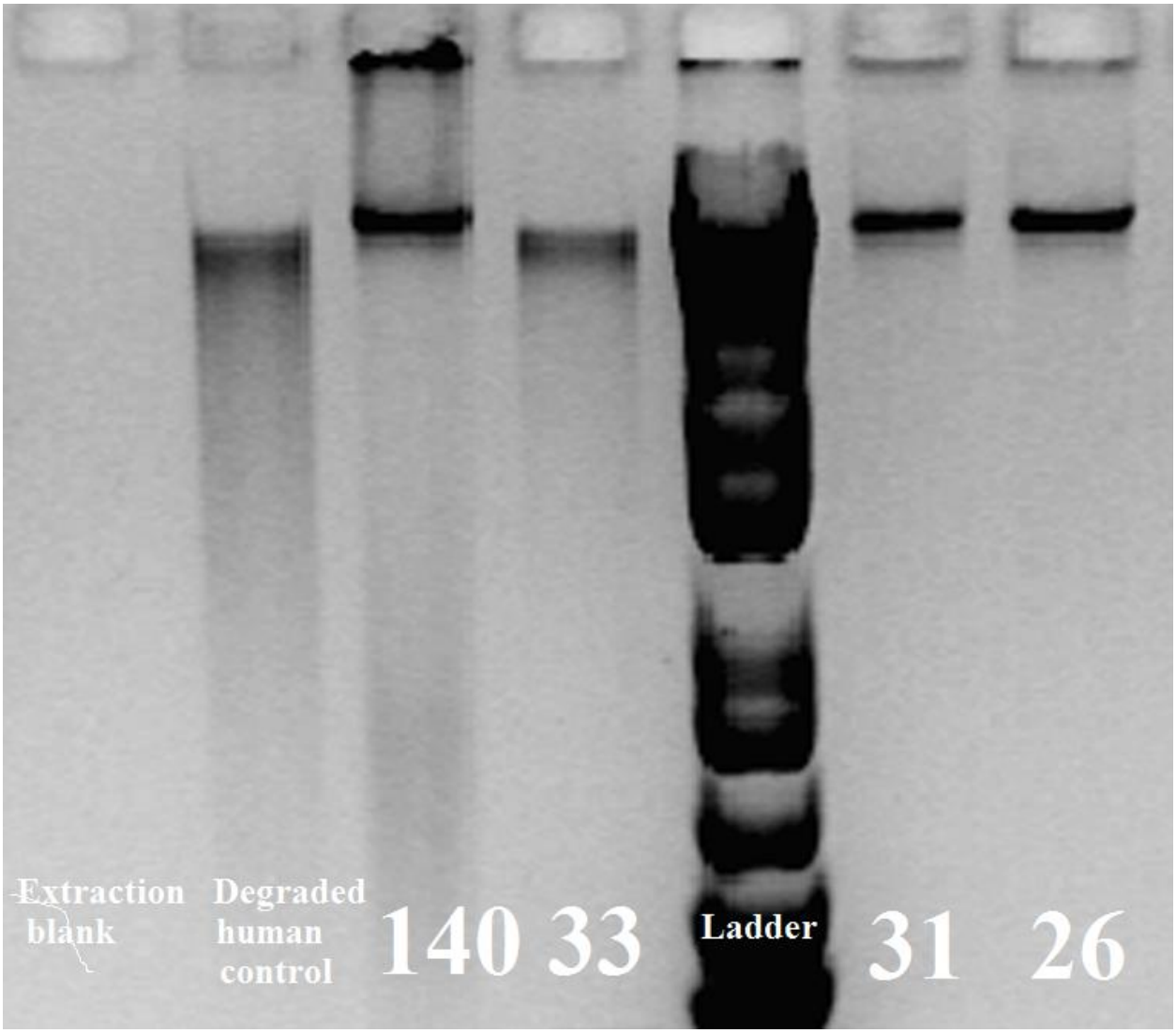

Figure 10 Yield gel utilized to QC candidate samples for next generation whole genome sequencing. Note: Degraded human control that was pictured in Figure 15 C EM.

Since ethidium bromide stained gel separation of the DNA obtained from the evidence samples showed no evidence of DNA degradation, in most cases, the contaminated, hemolyzed, degraded blood sample comprised a standard by which the unknown DNA obtained under more controlled conditions could be evaluated. The DNA extracted from the degraded human control blood sample tested 97.15%, SNP performance, well above all of the unknown hominin samples.

In order to further address the possibility that normal human DNA samples could produce such markedly low SNP performance in the Illumina® Bead Array assay, a separate group of human DNA controls was tested. These samples were extracted from buccal swabs obtained from sample handlers and laboratory workers and were also subjected to identical Illumina® Bead Array testing. These human control DNA samples yielded SNP performance above the 99% range. These data are summarized in Table 8.

Table 8. Percentage of successful SNPs using 2.5M human Bead Array testing

Sample Number	Percentage of SNPs successfully analyzed
26	89.24
34	88.49
35	83.46
33	67.18
39b	66.21
38	64.21
24	62.77
3	62.25
1	61.21
6	60.83
132	60.38
23	59.22
C-1 ♀	97.15*
C-2 ♂	99.63

*Control purposely degraded.

All of the above results eliminated the possibility of random operator error or faulty reagents during this procedure. Hence, the data obtained from the Illumina® Bead Array testing of the unknown samples can be considered highly unusual and not consistent with human.

The failures across all 2.5 million SNPs tested in the Illumina® Bead Array when compared with the human threshold designated above 95% SNP detection, even in severely degraded human DNA samples, suggests sequence variation from human in the actual DNA obtained from the unknown hominin samples.

Potential DNA Sequence Anomalies:
The DNA was amplified across loci such as *MC1R, TAP1* and *Amelogenin* and examined following separation via DNA agarose electrophoresis on yield gels. There were four outcomes to the testing across the examined samples (Figure 11). All human control samples were successfully amplified on all loci tested. In contrast, there were unusual PCR amplification findings from DNA obtained from the unknown samples. Some of the samples appeared to produce normal amplicons that resulted in bands consistent with the human controls. Other samples displayed clear bands that appeared to be of different sizes than those expected of normal human amplicons. Yet others had more than one band. Still other samples failed to amplify at all. Notably, since all human control samples, especially those obtained from severely degraded samples, behaved normally during the PCR analysis, operator and reagent error as well as sample contamination can be eliminated as explanations for the unusual PCR/electrophoresis migration anomalies found in the DNA from the unknown hominin samples.

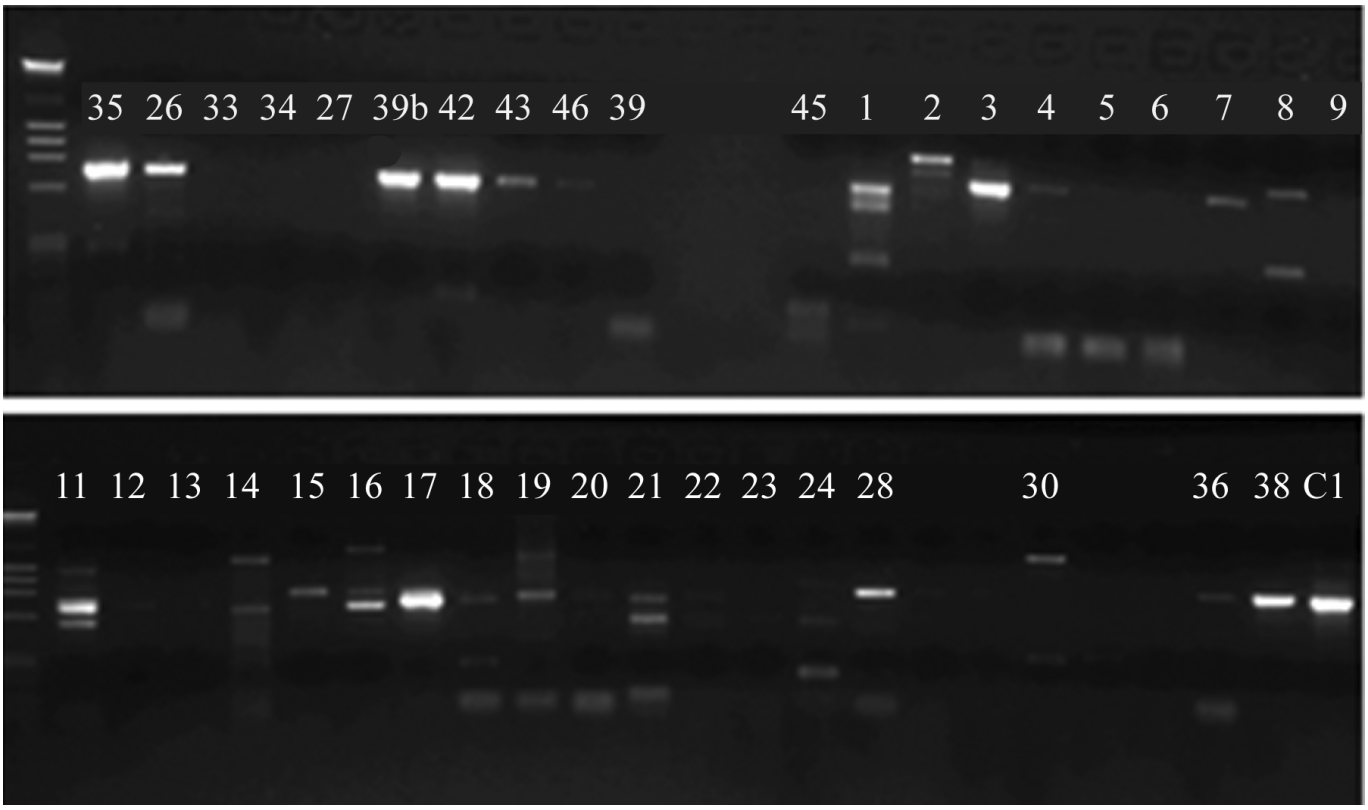

Figure 11 Agarose gel of MC1R locus showing unusual migrations of amplicons.

When the amplicons from various loci were visualized on agarose yield electrophoresis gels, the bands did not migrate consistently, but migrated at varying rates across the samples. Sequencing was attempted on samples that appeared to migrate differently on DNA agarose gels compared with human control samples. Some sequences produced no homology matches when BLAST searched against all primate, human, Neanderthal, Denisova, and other sequences in Genbank®. Since the DNA extracted from the hair and tissue purportedly obtained from a variety of unusual hominins displayed novel attributes as measured by several analysis techniques, it was decided to utilize electron microscopy to examine the structural properties of the DNA.

Electron Microscopy:
Electron microscopic characterization is considered superior to biochemical techniques when it comes to an accurate topographical analysis of single- and double-stranded DNA such as contour length/width, twisting, backfolding, nucleic acid-protein interactions, DNA-RNA heteroduplex analysis, R-loop mapping, degradation, replication, or packaging into higher order complexes.
Extracted DNA from sample 26 was previously found to have very little degradation and was analyzed by transmission electron microscopy with interesting results. These findings were replicated on two separate occasions using two separate DNA extractions from the same sample. Figure 12, panel A, shows long double stranded DNA up to 15-20 kb, but some of the DNA also comprised smaller fragments. In addition (Figure 12, panel B), most of the DNA showed clustering of material along the length or at ends and were typical of disordered single-stranded segments. There were also frequent single-stranded gaps and single-stranded ends which were not observed with the degraded human DNA control sample (Figure 12, panel C). Occasionally, with the single-stranded ends, a backfolding was observed that is typical for unspecific base

pairing that can occur in the absence of treatment with formaldehyde glyoxal or elevated temperatures[70-74].

The DNA utilized for electron microscopy originated from the same extractions that were combined and utilized for whole genome nuclear sequencing. High quality DNA is required for whole genome sequencing so we have no reason to believe that there was degradation or modification of the DNA pictured in Figure 12, panels A and B that would cause the atypical appearance of the DNA.

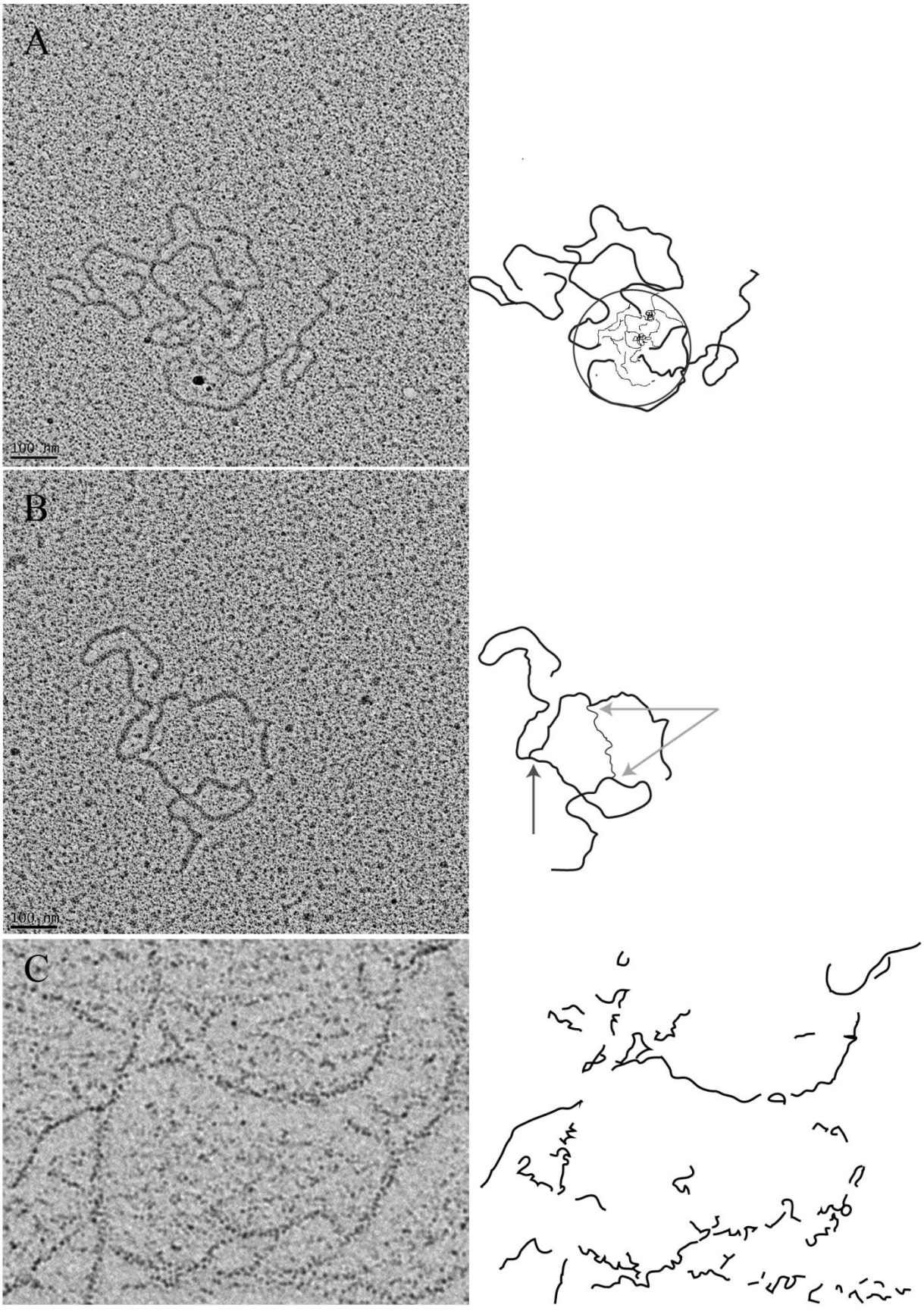

Figure 12 Transmission electron micrographs of extracted DNA after spreading and rotary shadowing. The light arrows highlight single-stranded DNA segments residing within double-stranded DNA and the dark arrow points at a Y-branch.

Next Generation Whole Genome Sequencing

In light of the large number of failures to amplify, failure to meet the human SNP threshold and unusual sequences obtained, three samples (26, 31 and 140) with large amounts of high quality DNA were selected for next generation whole genome sequencing. Sample 26 was a tissue sample purportedly from a Sasquatch. (Figure 13 and Supplementary Data 4). Sample 31 was collected from a research site using a food trap and forensic techniques to ensure that there was no human contamination (Figure 14 and Supplementary Data 5). Sample 140 was fresh dried blood from the inside of a vertical downspout that had been chewed and the individual was injured by sharp metal (Figure 15 and Supplementary Data 6). The inside of the downspout was clean and untouched by humans and the blood sample fresh and pristine. Large bite marks, hair, and fang marks were also evident on the downspout. The pipe was crushed to the point that the metal was torn with force.

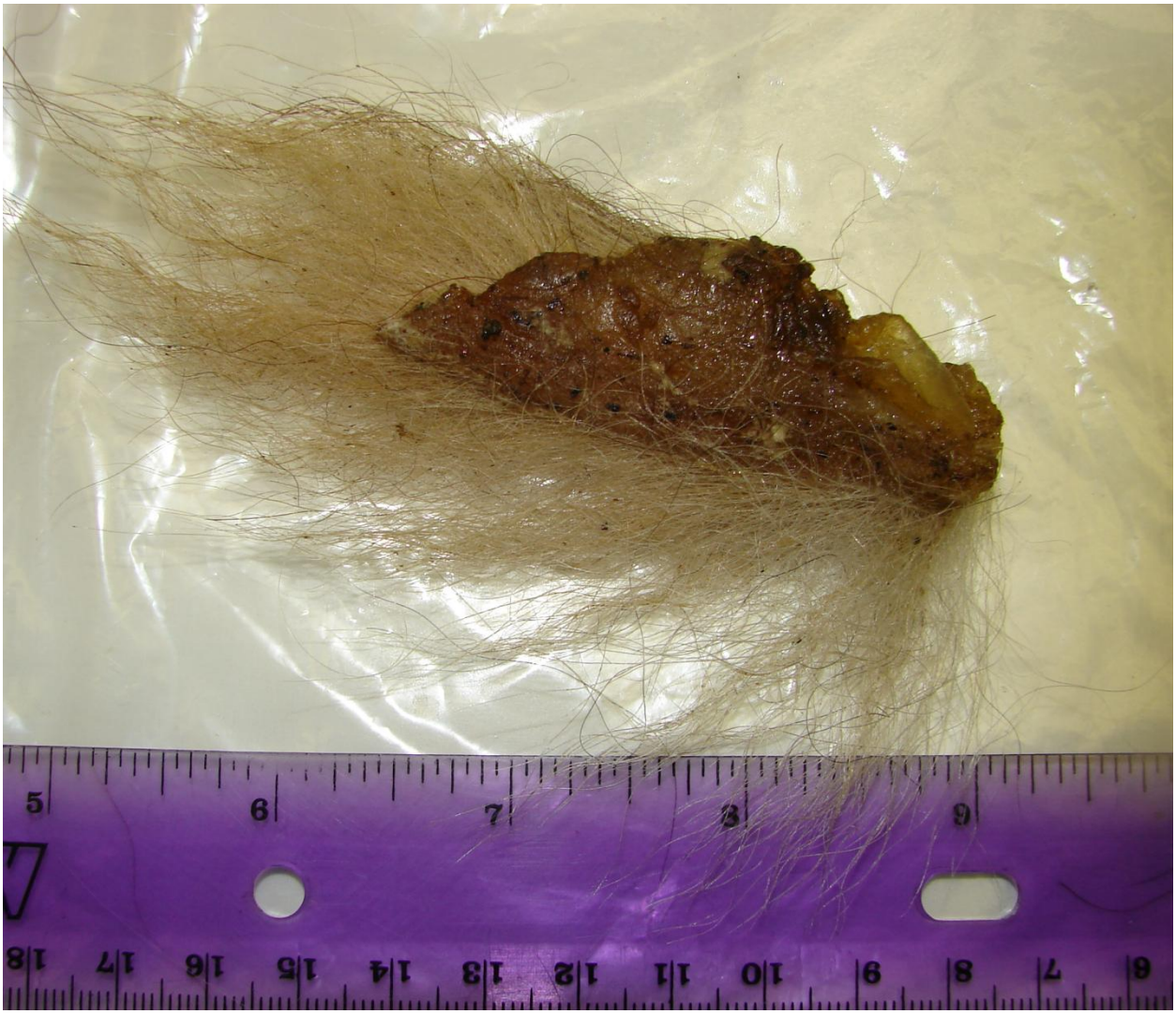

Figure 13 Sample 26, Tissue with hair, skin, subcutaneous tissues and some muscle.

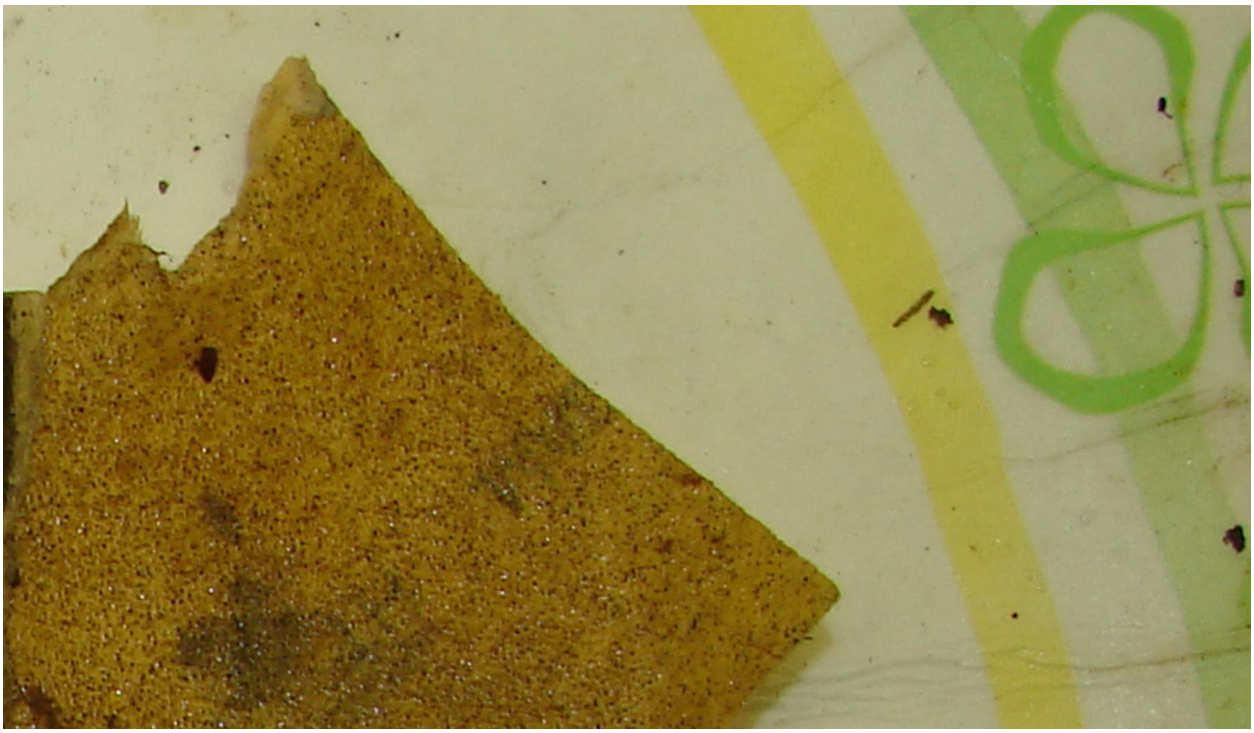

Figure 14 Sample 31, Plate with sandpaper used as a food trap.

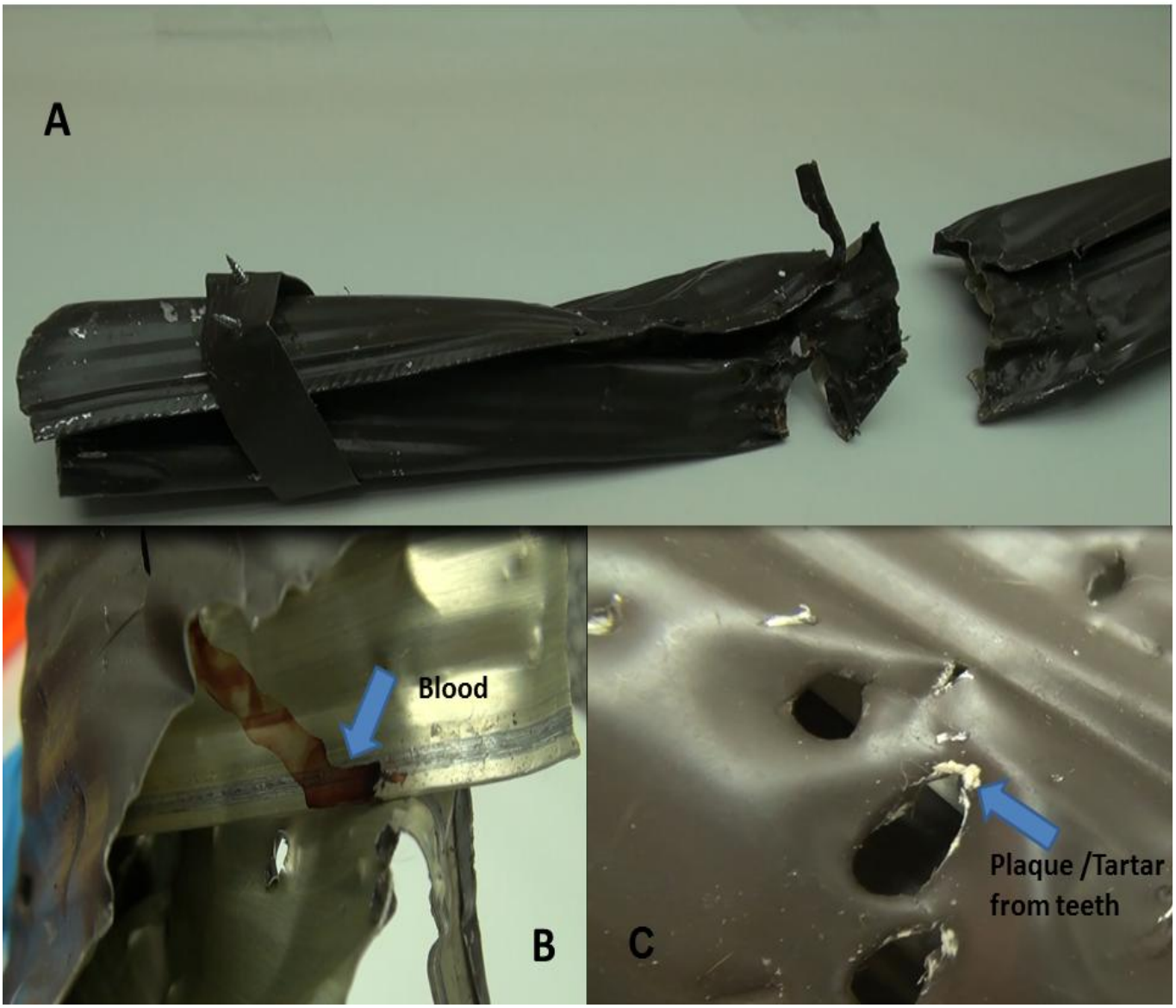

Figure 15 Sample 140, Chewed downspout with blood. (A) Overall view of Downspout. (B) Blood on the Downspout that was used for whole genome next generation sequencing. (C) Teeth marks in arched pattern with plaque and tartar residue left on the downspout.

The DNA from these three samples was sequenced using the next generation Illumina platform at the University of Texas, Southwestern in Dallas, TX, a laboratory that sequences human genomes[75-77]. On average, there were 70-110 million total reads for each sample in each lane, which is well over 90 Gb of raw sequence for each sample comprising greater than 30X coverage.

The run summary generated by the HiSeq 2000 next generation sequencer provides scores, Q30, for run quality[78]. Q30 can also be used to determine if there was any contamination (or mixture) found in the samples sequenced. According to Illumina, a pure, single source sample would have an Q30 score of 80 or greater with an average of 85. However, if there was contamination present in the sample sequenced, the divergent sequences would compete against one another

prior to sequencing causing a contaminated sample to have a Q30 score of 40 to 50. The Q30 scores for the three genomes sequenced had Q30 scores of 88.6, 88.4 and 88.7 respectively for samples 26, 31 and 140. The Q30 is the percent of the reads that have the statistical probability greater than 1:1000 of being correctly sequenced. Therefore, not only were the sequences from a single source, but the quality of the sequences were far above the average genome sequenced using the Illumina next generation sequencing platform. The high quality of the genomes can be attributed to the stringent extraction procedures utilized whereby the DNA was repeatedly purified. This ultra-purified DNA also allowed for greater than 30X coverage of the three genomes. The summary and of the next generation sequencing generated by the HiSeq 2000 Illumina sequencer is furnished as Supplementary Data 7-10.

In depth analysis of all three genomic sequences (samples 26, 31 and 140) was performed at the University of Texas, Southwestern and alignment confirmed by the University of North Texas Health Science Center. Using CLC Bio Genomic Workbench version 5.1, a subsample of extracted reads were assembled to create a consensus sequence using the human chromosome 11 as a reference. A set of 379 genes was analyzed among three different samples (26, 31, and 140). The coverage of a subset of 159 genes was compared among these samples, and a consensus sequence was generated.

For each sample, the subsets of genes were concatenated to produce a long, single sequence used to generate a supertree. The length of the concatenated sequences was 656,048 (26), 541,435 (140), and 74,589 (31). These concatemers (supercontigs) were used to find sequence homologs and generate phylogenetic trees. The genes represented in the selective supercontig were *DLG2, NTM, ODZ4, FAT3, CADM1, SOX6, DSCAML1, NCAM1, GRM5, MPPED2, PKNOX2, KIAA0999, ZBTB16, and SHANK2.* The lengths of the selective supercontigs were 293,249 (26), 235,738 (140), and 39,582 (31). The size-filtering generated supercontigs that only included genomic regions with long sequences, and created phylogenetic trees similar for all three samples, where the average branch length to Primates is 0.02 for all three samples. The Sasquatch consensus-selective supercontigs were used to create a set of phylogeny trees utilizing BLAST pairwise alignments (Supplementary Figures 4-6).

As shown in Supplementary Figures 4, 5 and 6, the Sasquatch consensus that showed homology to human chromosome 11 reference sequence is related to primate lineages including Homo sapiens, Otolemur garnettii, Pan troglodytes (Chimpanzee), Macaca mulatta (Rhesus Monkey), Nomascus leukogenys (White cheeked Gibbon) and Callithrix jacchus (Common Marmoset) and other primate species.

Additionally, a phylogenetic tree was constructed using the entire consensus sequences for Samples 26 and 140 (Figure 16). The lengths for these sequences were 2,726,786 and 2,101,957 bases respectively. Because the global BLASTn demonstrated statistically significant alignment across the Primate order; a Primate "Drill Down" utilizing BLASTn with inclusive Primate organism taxids was analyzed. This was intended to focus on the most statistically significant chromosomal aligned sequences and eliminated less significant mammalian sequences. A BLAST Tree using the entire 2,726,786 and 2,101,957 bases aligned with human chromosome 11 for Samples 26 and 140 respectively was generated from the next generation sequencing results[79, 80]. Figure 16 suggests that the genomes are closely related and are likely derived from

the same species. Sample 31 had a shorter aligned sequence so it was not included in Figure 19 to eliminate any potential sequence bias. Nevertheless, Sample 31 shared the same characteristics including the interspersed human sequence with novel primate sequence and aligned well with the other two genomes.

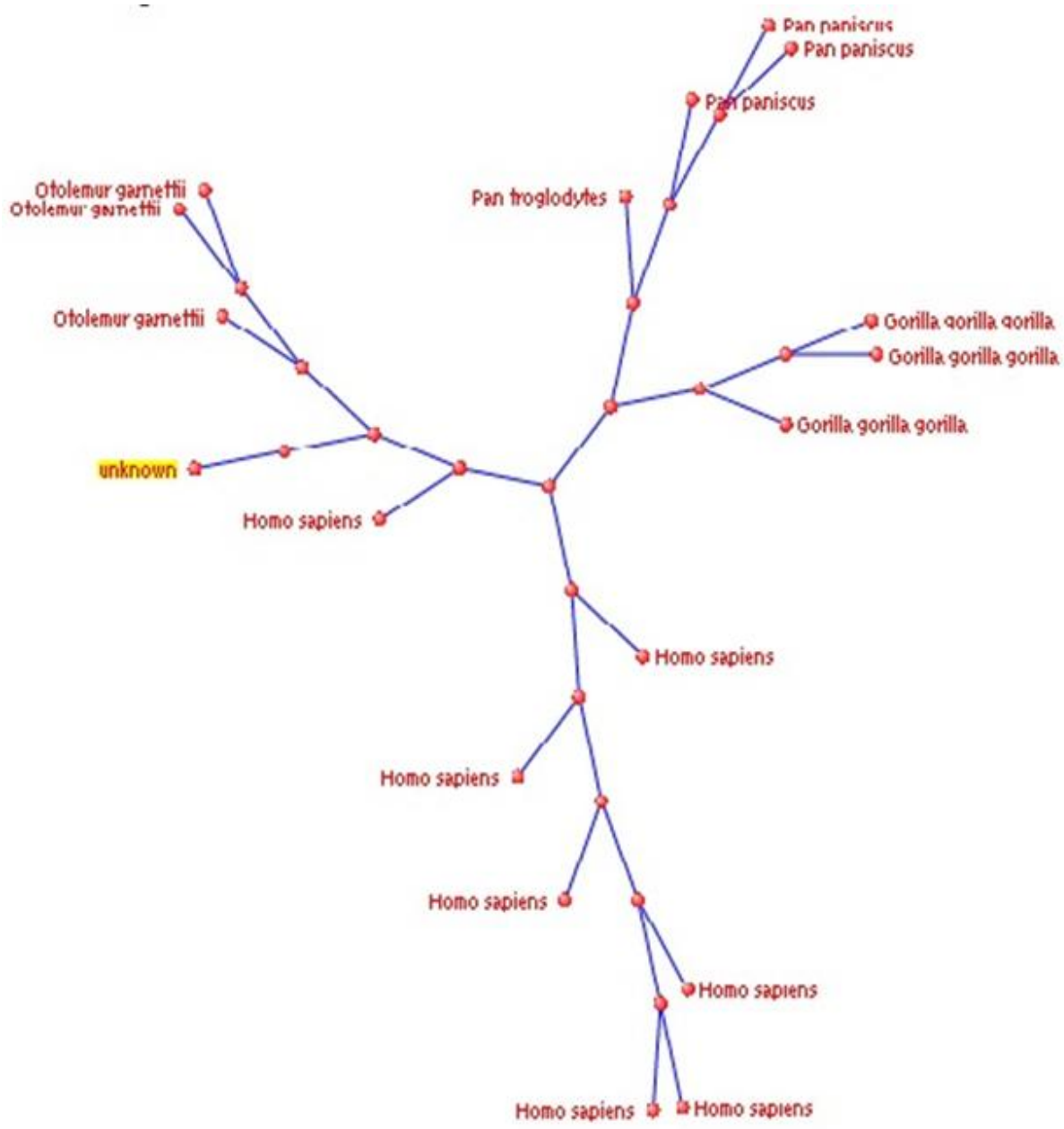

Figure 16 "Drill Down" analysis and resulting supertree of Global BLASTn sequence alignment results, Samples 26 and 140.

The same analysis method was used in an attempt to confirm homology to the human mitochondrial genome. The three Sasquatch consensus sequences all showed homology to the

human mitochondrial genome and were used to create a set of phylogeny trees utilizing BLAST pairwise alignments (Supplementary Figures 4, 5, and 6). The consensus sequences obtained from the next generation sequencing were consistent with the mitochondrial sequences generated for these three samples by Family Tree DNA in the early stages of the study (Supplementary Figures 1, 2 and 3).

Discussion

During this five year project, DNA was extracted and analyzed from over one hundred samples of hair, tissue, tree bark shavings, saliva, and blood submitted by a continent-wide team of dedicated collectors from 14 US states and two Canadian provinces. All of the samples originated from areas in North America that overlapped with substantial eyewitness reports of an unknown large bipedal hominin.

Prior to DNA analysis, all samples were screened to eliminate common wildlife species using a variety of methodologies including microscopic morphological examination and comparison against a large reference collection of known North American wildlife hair. Further, utilizing several morphological criteria, unusual characteristics were seen in hair texture, diameter ratios, medullary structures and cuticle patterns. The purported Sasquatch hairs also demonstrated elongated hair roots. Taken together, these morphologic differences indicate non-human hair that is also inconsistent with hair from known wildlife species. Over one hundred specimens survived this initial screen and were judged sufficiently different from human and wildlife samples to undergo DNA analysis.

In order to minimize concerns about bias, degraded DNA, operator error, bad reagents or contamination influencing the data during this study, sample integrity and cleanliness were prioritized. As a preliminary screen for possible human contamination of samples during the collection or laboratory analysis phases, control samples obtained from submitters and working with the collection of field samples were run in parallel. In all cases, as demonstrated by clean sequences without false heteroplasmic bases denoting mixture or contamination and single source profiles with the PowerPlex® 16 amplification kit, no evidence was obtained that the DNA extracted from collectors or scientists or any other secondary source was present as a contaminant in any of the samples. In the next generation whole genome sequencing, the Q30 scores provided definitive proof that the genomes were derived from single source of DNA, not a mixture of human DNA contaminated with animal DNA since the Q30 scores vastly decrease if contamination is present. Furthermore, the Q30 scores also indicated that the genomes were well above average quality (Supplementary Data 7-10) further eliminating the possibility of contamination.

Mitochondrial DNA analysis of all samples tested in three independent laboratories unambiguously confirmed exclusively human haplotypes from several different geographic areas. Further, Supplementary Figures 1, 2, 3, 7, 8 and 9 show six phylogenetic trees from whole mitochondrial sequences that clearly show human mitochondrial DNA. The Figures also support that the mitochondrial genomes derived from the three whole genome samples were consistent with the previous mitochondrial DNA sequences obtained at the beginning of the study.

It should be emphasized that there have been attempts by other groups around the world to obtain mitochondrial DNA sequence from Sasquatch (aka Yeti, Bigfoot) with remarkably consistent outcomes. Hence, the mitochondrial DNA findings have been unvarying between samples in this study and between seven different laboratories, four of which were independent of this project and predated it. Though the mitochondrial data has been consistent between all previous laboratories and throughout this study, the possibility of a human male/progenitor female mating cannot be excluded without testing larger numbers of samples.

NY University tested a well-known individual, Zana, and her hybrid son, Khwit, the account of which was validated with eyewitness reports from Russia where the incident occurred. Both the "wild man" mother and hybrid son yielded mitochondrial DNA H haplotypes with only 1 base variant from the human reference sequence[81-82]. Additionally, preliminary results of one Siberian sample tested in parallel, but not included in this study yielded a human T2b haplotype.

In a separate study, samples from several individuals were tested at Helix Biological Laboratory, one of which, Sample 134, is in this study. A scientific paper was written but never submitted in reference to the human mitochondrial DNA findings. (Supplementary Data 8) The results obtained from the mitochondrial DNA sequencing of Sample 134 were consistent between the studies as being human in origin. Other laboratories such as the University of Minnesota tested Sample 34 as having human mitochondrial results as seen on the television series, *Monster Quest.* Sample 37 was tested at New York University, Paleo Labs in Ontario, Canada and three laboratories in this study. The results obtained from all five laboratories were consistent with 37 having a human *H12* haplotype. Nuclear DNA testing of Sample 37 yielded novel sequence in addition to human sequence suggesting it originated from a novel hominin. Furthermore, testing confirmed this sample was a female as observed by the field scientists that collected the sample. A photo of this individual from which the sample was obtained can be seen in Figure 4 and Supplementary Video 1.

As detailed in the Results and Supplementary Material and Methods sections, novel properties of nuclear DNA was detected from analysis of Sasquatch nuclear DNA samples; none of these properties were detected in any of the control samples that were run in parallel.

Off-ladder alleles and allelic dropout were detected during analysis of short tandem repeats (STRs). The mutations were seen in 16 separate loci when PowerPlex® 16 amplification was used. *X* dropout was noted in a significant number of the Sasquatch samples due to novel sequence in the homologous *X* region of the *Y* chromosome at the *amelogenin* locus. This was supported by sequencing this region and the associated novel sequences and failures to sequence at this locus. Novel sequence, failures to sequence and interspersed human sequence was also observed when *amelogenin* locus exons 1, 2, 4/5 and 8 were sequenced. The same amplification/sequencing patterns were observed at the *TAP1* locus.

In order to further explore the reasons for the widespread novel amplifications described above, DNA from Sample 26 was imaged using electron microscopy. Electron micrographs of the DNA revealed unusual double strand – single strand – double strand transitions which may have contributed to the failure to amplify during PCR. The high quantities of single stranded DNA, interspersed with double stranded DNA seen in Figure 12 may suggest substantial structural

abnormalities of the DNA itself. The DNA was visualized twice by electron microscopy from aliquots of two different extractions from the same sample 26. Both of these extractions not only had yielded successful PCR amplifications previously, but were combined and utilized for whole genome nuclear sequencing at a later time. Alternatively, the abundance of single-stranded loci observed by EM could be interpreted as high numbers of replication forks or another structurally altering genomic process involving an increase in helicase activity.

In a further attempt to resolve some of the observed DNA anomalies, next generation whole genome nuclear sequencing was performed on three samples. Samples 26, 31 and 140 were selected because of the large quantities of high quality DNA extracted. All three samples were successfully sequenced and yielded, on average, 70-110 million total reads for each sample in each lane, indicating well over 90 Gb of raw sequence with coverage greater than 30X. Quality control indices including Q30 indicated that all three yielded high quality genomic sequence.

Utilizing a human chromosome 11 reference sequence, preliminary Sasquatch concatenated sequences of 656,048, 541,435 and 74,589 base pairs, were derived from the full genomic sequence from samples 26, 140 and 31 respectively. These supercontigs were used to find sequence homologies and to construct an initial set of phylogeny trees (Supplementary Figures 4, 5, and 6). The phylogeny trees clearly indicate relationships with primate sequences, including lemurs, chimpanzee, macaques, gibbons and marmosets and close relationship with humans.

As detailed in the Results section, the genes represented in the selective supercontig generated from samples 26, 140 and 31 were all on Chromosome 11 of hg 19. The genes represented on the selective supercontig were: *DLG2,* disc large homolog a membrane associated guanylate kinase family member involved in glucose homeostasis, *NTM* aka neurotrimin, a gene implicated in late stage Alzheimer's disease, *ODZ4* a member of the teneurin family of developmental regulators, *FAT3* a Cadherin family member and a putative tumor suppressor, *CADM1* a cell adhesion molecule and another putative tumor suppressor gene, *SOX6* an activator of chondrocyte differentiation, *DSCAML1* which is involved in hemophilic cell adhesion and spacing of neurons in the retina, *NCAM1* Neural cell Adhesion molecule, *GRM5* A metabotropic Glutamate Receptor family member, *MPPED* Metalloproteinase domain containing protein 2 and also a putative tumor suppressor, *PKNOX2 PBX/Knotted homebox 2* gene a nuclear transcription factor implicated in substance abuse and dependence in humans, *KIAA0999, ZBTB16/PLZFa* Glucocorticoid response element Zinc finger protein and *SHANK2* a scaffolding protein involved in neural synapses[83-94]. Thus, the selective supercontigs comprised an abundance of neural associated and putative tumor suppressor sequences all of which are highly conserved in primates and humans and clearly establish that the Sasquatch is closely related to humans. The high homology with primate lineages, including but not limited to, as demonstrated in phylogenetic trees (Supplementary Figures 4, 5, and 6) indicate that the supercontigs contain highly conserved human and primate gene sequences. Additional studies in the future will expand to the analysis of the entire genomes from samples 26, 31 and 140.

Additionally, the entire supercontigs for Samples 26 and 140 with lengths of 2,726,786 and 2,101,957 bases respectively were also utilized to make a phylogenetic supertree (Figure 16). These phylogenetic trees supported the original trees generated by using the selective supercontigs. Figure 16 supports that both Samples 26 and 140 are the same species. Further

studies with additional phylogenetic tree analysis will further elaborate the relationship between Sasquatch and other primate lineages.

Conclusions

In summary, we have extracted, analyzed and sequenced DNA from over one hundred separate samples of hair, tissue, toenail, bark scrapings, saliva and blood obtained from scores of collection sites throughout North America. Hair morphology was not consistent with human or any known wildlife hairs. DNA analysis showed two distinctly different types of results; the mitochondrial DNA was unambiguously human, while the nuclear DNA was shown to harbor novel structure and sequence. All known ape and relic hominin species such as Neanderthal and Denisovan were excluded as being contributors to both the nuclear and mitochondrial sequences. Analysis of whole genome sequence and analysis of preliminary phylogeny trees from the Sasquatch indicated that the species possesses a novel mosaic pattern of nuclear DNA comprising novel sequences that are related to primates interspersed with sequences that are closely homologous to humans. These data clearly support that these hominins exist as a novel species of primate. The data further suggests that they are human hybrids originating from human females. This hybridization can be likened to humans with Denisovan admixture resulting from Denisovan males mating with human females[103]. The same type of mating potentially occurred with Sasquatch; however, in the case of Sasquatch, the admixture is human. Though preliminary analysis supports the hybridization hypothesis, alternatively, it could also be hypothesized that the Sasquatch are human in origin, having been isolated in closed breeding populations for thousands of years. Nevertheless, the data conclusively proves that the Sasquatch exist as an extant hominin and are a direct maternal descendent of modern humans.

Future

At this time, analysis of the Sasquatch genomes is still ongoing. Further data will be presented as it becomes available. Additionally, analysis of hair purportedly from a Siberian Wildman is being tested in an effort to determine if relatedness exists between the Sasquatch and the Russian Wildmen. A species name has been applied for with ZooBank, *Homo sapiens cognatus*, LSID:40E2FA1F-10A1-4D42-8B02-A007347F1B43

References

1. ***The Epic of Gilgamesh***, Assyrian International News Agency Books Online http://www.aina.org/books/eog/eog.pdf (2011)
2. **Walker Elkanah and Mary Walker papers**, Mss 1204, Oregon Historical Society Research Library , Diary, (1838)
3. Boone Daniel **Boone: The Life and Legend of an American Pioneer**, Holt Paperbacks New York, NY (1992)
4. Wu, X, Zeng, X and Yao, H. **Analysis of a single strand of hair by PIXE, IXX and synchrotron radiation**. *Nuclear Instruments and Methods in Physics Research* **B75**:567–570. (1993)

5. Milinkovitch, M C, Caccone, A and Amato, G. **Molecular phylogenetic analyses indicate extensive morphological convergence between the ‘‘yeti’’ and primates.** *Molecular Phylogenetics and Evolution* **31**:1–3. (2004)
6. Coltman, D and Davis, C. **Molecular cryptozoology meets the Sasquatch**. *TRENDS in Ecology and Evolution* **21**:60–61. (2006)
7. Lockley, M, Roberts, G and Kim, J Y. **In the footprints of our ancestors: an overview of the hominid track record.** *Ichnos* **15**:106–125. (2008)
8. Kim, J Y, Kim, K. S, Lockley, M G and Matthews, N. **Hominid ichnotaxonomy: an exploration of a neglected discipline.** *Ichnos* **15**:126–139. (2008)
9. Lozier, J D, Aniello, P and Hickerson, M J. **Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modeling**. *Journal of Biogeography* **36**:1623–1627. (2009)
10. Bindernagel, J A. ***North America's Great Ape: The Sasquatch - A Wildlife Biologist Looks at the Continent's Most Misunderstood Large Mammal.*** Beachcomer Books. Courtenay, B.C., Canada. **(**1998)
11. Bindernagel, J A**.** ***The Discovery of the Sasquatch***, Beachcomer Books. Courtenay, B.C., Canada. (2010)
12. Krantz, G S, ***Bigfoot Sasquatch: Evidence***. Hancock House Pub Ltd. Surrey, BC, Canada. 2 Revised edition. June, 1999.
13. Krantz, G S, ***Big Foot Prints a Scientific Inquiry into the Reality of Sasquatch.*** Johnson Books; Boulder**,** CO First edition September, (1992)
14. Mizokami, K, Franzoni, H and Glickman, J. **Native American Sasquatch Names".** Sasquatch Research. http://www.sasquatchresearch.net/sassynames.html. Retrieved 2008-08-18.
15. Bisbing, R E. **The forensic identification and association of human hair.** In: *Forensic Science Handbook*; Saferstein, R, ed. Prentice Hall, Englewood Cliffs, NJ. (1982)
16. Robertson, J, ed. ***Forensic and Microscopic Examination of Human Hair***. Taylor and Francis, London. (1999)
17. Moore, T D, Spence, L E, Dugnolle, C E and Hepworth, W G, ed. ***Identification of the Dorsal Guard Hairs of Some Mammals of Wyoming*****.** Wyoming Game and Fish Department, Cheyenne, WY. (1974)
18. Hicks, J W. ***Microscopy of Hairs: A Practical Guide and Manual*****.** Federal Bureau of Investigation. Washington, DC, January 1977.
19. Linch, C A, Prahlow, J A and Smith, S L. **Evaluation of the human hair root for DNA typing subsequent to microscopic comparison**. *J Forensic Sci*. **43(2).** (1998)
20. ***PowerPlex®16 System Technical Manual.*** Promega Corp., Madison, WI. (Revised 5/08)
21. Krings, M, Geisert, H, Schmitz, R W, Krainitzki, H and Pbo, S. **DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen**. *P Natl Acad Sci-Biol.* **96**:5581-5585. (1999)
22. Cooper, A, *et al*. **Neandertal genetics**. *Science.* **277**:1021-1023. (1997)
23. Gagneux, P, *et al.* **Mitochondrial sequences show diverse evolutionary histories of African hominoids.** *P Natl Acad Sci-Biol.* **96**:5077-5082. (1999)
24. Anderson, S, *et al*. **Sequence and organization of the human mitochondrial genome**. *Nature.* **290**:457-474. (1981)

25. Tamura, K and Nei, M. **Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees**. *J Mol Evol.* **10**:512-526. (1993)
26. Excoffier, L and Yang, Z. **Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees**. *Mol Biol Evol.* **16**:1357-1368. (1999)
27. Burckhardt, F, von Haeseler, A and Meyer, S. **HvrBase: compilation of mtDNA control region sequences from primates**. *Nucleic Acids Res.* **27**:138-142. (1999)
28. Caramelli, D**,** *et al*. **A 28,000 years old Cro-Magnon mtDNA sequence differs from all potentially contaminating modern sequences**. *PLOS One.* **3:**7, e2700. (2008)
29. Gabunia, L and Vekua, A. **A Plio-Pleistocene hominin from Dmanisi, East Georgia, Caucasus.** *Nature.* **373**:509-512. (1995)
30. Denisova Reich, D., *et al.* **Genetic history of an archaic hominin group from Denisova Cave in Siberia.** *Nature.* **468:**1053-1060. (2010)
31. Brown, P. *et.al*., **A new small bodied hominin from the late Pleistocene of Flores, Indonesia**. *Nature*. **431**:1055-1061. (2004)
32. Morwood, M J, *et al*., **Archaeology and age of a new hominin from Flores in eastern Indonesia.** *Nature*. **431**:1087-1091. (2004)
33. Aiello, L and Dean, C. ***An Introduction to Human Evolutionary Anatomy***. Academic, New York, (1998).
34. White, T D, Suwa, G, Simpson, S and Asfaw, B. **Jaws and teeth of Australopithecus afarensis from Maka, Middle Awash, Ethiopia**. *Am J Phys Anthropol*. **111**:45–68. (2000)
35. Tobias, P V. ***The Skulls, Endocasts, and Teeth of Homo habilis***. University Press, Cambridge, MA. (1991)
36. Walker, A and Leakey, R, eds. ***The Nariokotome Homo erectus skeleton***. Harvard Univ. Press, Cambridge, MA. (1993)
37. Vekua, A, *et al.* **A new skull of early Homo from Dmanisi, Georgia**. *Science.* **297**:85–89 (2002)
38. Barras, Colin "**Stone Age toe could redraw human family tree**", *New Scientist* (2011)
39. Reich**,** D, *et al.* **Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania**. *The American Journal of Human Genetics*, doi:10.1016/j.ajhg. (2011)
40. **GenBank®** *Nucleic Acids Research , (Database issue)* ***D32****:7.* http://www.ncbi.nlm.nih.gov/genbank/ (2011)
41. Bradley, B and Stanford, D. **The North Atlantic ice-edge corridor: a possible Palaeolithic route to the New World, World Archaeology 36(4):** 459 – 478 Debates in World Archaeology # 2004 Taylor & Francis Ltd, London DOI: 10.1080/0043824042000303656
42. Stanford, D J and Bradley, B A. **Across Atlantic Ice: The Origin of America's Clovis Culture,** University of California Press, Berkley and Los Angeles, CA (2012)
43. Commane, M, Hren, C, and Warshawsky, I. **A rare mutation in the primer binding region of the amelogenin gene can interfere with gender identification**. *J Mol Diagn*. **6(4)**. (2004)
44. Buel, E, Wang, G, Schwartz, M. **PCR amplification of animal DNA with human X-Y amelogenin primers used in gender determination.** J Forensic Sci. **40(4):**641-4. (1995)

45. Shadrach, B, Commane, M, Hren, C and Warshawsky, I. **A Rare Mutation in the Primer Binding Region of the Amelogenin Gene Can Interfere with Gender Identification**. *J Mol Diagn*. **6(4)**:401–405. (2004)
46. Maciejewska A, Pawłowski R. **A rare mutation in the primer binding region of the Amelogenin X homologue gene.** *Forensic Sci Int Genet*. **4:**265-7. (2009)
47. Lalueza-Fox, C, *et al.* **A Melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals.** *Science.* **318**:1453. (2007)
48. Makova, K and Norton, H. **Worldwide polymorphism at the *MC1R* locus and normal pigmentation variation in humans**. *Peptides.* **26:**1901–1908. (2005)
49. Kanetsky, P A., *et al.* **Population-based study of natural variation in the melanocortin-1receptor gene and melanoma.** *Cancer Res.* **66:**18. (2006)
50. Hongmei, N, *et al.* **Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians**. *Int J Cancer*. **125**:909–917. (2009)
51. Rees, J L. **Genetics of hair and skin color**. *Annu Rev Genet.* **37:**67–90. (2003)
52. Carroll, L, Voisey, J and van Daal, A. **Gene polymorphisms and their effects in the melanocortin system**. *Peptides.* **26:**1871–1885. (2005)
53. Sturm, R A, *et al.* **The role of melanocortin-1 receptor polymorphism in skin cancer risk phenotypes.** *Pigm Cell Res*. **16:**266–272. (2003)
54. Gerstenblith, M R, Goldstein, A M, Fargnoli, M C, Peris, K, and Landi, M T. **Comprehensive evaluation of allele frequency differences of MC1R variants across populations**. *Hum Mutat.* **28(5)**: 495-505. (2007)
55. Makova, K, and Norton, H. **Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans.** *Peptides.* **26**:1901–1908. (2005)
56. Bastiaens, M T, *et al.* **Melanocortin-1 receptor gene variants determine the risk of nonmelanoma skin cancer independently of fair skin and red hair**. *Am J Hum Genet.* **68**:884–894. (2001b)
57. Pastorino L, *et al.* **Novel MC1R variants in Ligurian melanoma patients and controls.** *Hum Mutat.* **24**:103. (2004)
58. Rana B K, *et al*. **High polymorphism at the human melanocortin 1 receptor locus.** *Genetics.* **151**:1547–1557. (1999)
59. Rees, J L. **Genetics of hair and skin color**. *Annu Rev Genet.* **37:**67–90. (2003)
60. Stedman, H H, *et al.* **Myosin gene mutation correlates with anatomical changes in the human lineage.** *Nature.* **428:**25. (2004)
61. Acakpo-Satchivi, L, *et al*. **Growth and muscle defects in mice lacking adult myosin heavy chain genes**. *J Cell Biol*. **139:**1219–1229. (1997)
62. Martinsson, T, *et al*. **Autosomal dominant myopathy: missense mutation** (Glu-706 ! Lys) in the myosin heavy chain IIa gene. *P Natl Acad Sci-Biol.* **97**:14614–14619. (2000)
63. Korfage, J A and Van Eijden, T M. **Myosin heavy chain composition in human masticatory muscles by immunohistochemistry and gel electrophoresis.** *J Histochem Cytochem*. **51**:113–119. (2003)
64. Hohl, T H. **Masticatory muscle transposition in primates: effects on craniofacial growth.** *J Maxillofac Surg*. **11**:149–156. (1983)
65. Allen, D L, Harrison, B C, Sartorius, C, Byrnes, W C and Leinwand, L. **A mutation of the IIB myosin heavy chain gene results in muscle fiber loss and compensatory hypertrophy.** *Am J Physiol-Cell PH.* **280**:C637–C645. (2001)

66. Bodmer, J G., *et al.* **Nomenclature for factors of the HLA system,** 1991. WHO Nomenclature Committee for factors of the HLA system. *Tissue Antigens*. **39(4):**161-73. (1992)
67. Schölz, C, Tampé, R. **The Intracellular Antigen Transport Machinery TAP in Adaptive Immunity and Virus Escape Mechanisms.** *J Bioenerg Biomembr*. **37(6):**509-515. (2005)
68. Jackson, D G, Capra, J D. **TAP1 alleles in insulin-dependent diabetes mellitus: a newly defined centromeric boundary of disease susceptibility**. *Proc. Natl. Acad. Sci. U.S.A.* 90:23. (1994)
69. Steemers, F J and Gunderson, K L. **Whole genome genotyping technologies on the BeadArray™ platform**. *Biotechnol J*. **2**:41-49. (2007)
70. Haselkorn, R and Doty, P. **The reaction of formaldehyde with polynucleotides**. *J. Biol. Chem*. **236**:2738-2745. (1961)
71. Kleinschmidt, A K and Zahn, R K. **Über Desoxyribonucleinsäure-Molekeln in Protein-Mischfilmen.** Z. *Naturforsch*. **14b**:770-779. (1959)
72. Kleinschmidt, A K. **Monolayer technique in electron microscopy of nucleic acids and molecules.** *Methods Enzymol*. **12B**:361-377 (1968)
73. Spiess, E and Lurz, R. **Electron Microscopic Analysis of Nucleic Acids and Nucleic Acid-Protein Complexes.** *Method Microbiol* (ed. F. Mayer) **20**:293-323. (1988)
74. Hoppert, M, and Holzenburg, **A. Electron Microscopy in Microbiology. *RMS Microscopy Handbook*** 43, BIOS Scientific Publ. Ltd., Oxford, (1998).
75. Shendure, J and Ji, H. **Next-generation DNA sequencing,** *Nature Biotechnology* **26**:1135 - 1145 (2008) doi:10.1038/nbt1486.
76. *Metzker, M.* **Sequencing technologies — the next generation**. *Nature Reviews Genetics* **11**:31-46 (2010) doi:10.1038/nrg2626
77. Larkin, M A, **Clustal W and Clustal X version 2.0.** *Bioinformatics*, **23**:2947-2948. (2007) www.clustal.org
78. **Quality Scores for Next-Generation Sequencing. Technical Note Publication #770-2011-030**. Illumina, Inc. http://www.illumina.com/Documents/%5Cproducts%5Ctechnotes%5Ctechnote_Q-Scores.pdf
79. US. National Library of Medicine, National Institutes of Health. (2009). ***National Center for Biotechnical Information*.** Retrieved from http://blast.ncbi.nlm.nih.gov/Blast.cgi
80. Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "**A greedy algorithm for aligning DNA sequences**", J Comput Biol 2000; 7(1-2):203-14.
81. Eberhart, G.M. Mysterious Creatures: **A Guide to Cryptozoology**. ABC-CLIO, 2002. ISBN 1576072835, 9781576072837
82. **Ancient DNA, A Compilation of DNA Haplotypes Extracted from Ancient Remains**. http://www.isogg.org/ancientdna.htm Retrieved January 2012.
83. Smith, S, et.al. Isolation of a Gene (DLG3) **Encoding a Second Member of the Discs-Large Family on Chromosome 17 q12-q21**. *Genomics* **31(2)**: 145-150. (1996) http://dx.doi.org/10.1006/geno.1996.0025.
84. Ben-Zur, T, Feige, E, Motro, B, Wides, R. **The Mammalian Odz Gene Family: Homologs of a *Drosophila* Pair-Rule Gene with Expression Implying Distinct yet Overlapping Developmental Roles.** *Developmental Biology*. (2000)

85. Mitsu, K, Nakajima, D, Ohara, O, and Nakayama, M. **Mammalian fat3: A Large Protein That Contains Multiple Cadherin and EGF-like Motifs.** *Biochemical and Biophysical Research Communications*. **290**:1260-1266. (2002) http://dx.doi.ort/10.1006/bbrc.2002.6338
86. Michels, E, et.al. **CADM1 is a strong neuroblastoma candidate gene that maps within a 3.72 Mb critical region of loss on 11q23.** *BMC Cancer*, **8**:173. (2008) doi:10.1186/1471-2407-8-173
87. Smits, P, et.al. **The Transcription Factors L-Sox5 and Sox6 Are Essential for Cartilage Formation.** *Developmental Cell.***1:**277-290. (2001) http://dx.doi.org/10.1016/S1534-5807(01)00003-X.
88. Fuerst, P G, et.al**. DSCAM and DSCAML1 Function in Self-Avoidance in Multiple Cell Types in the Developing Mouse Retina.** *Neruon.* **64**:484-497. (2009) http://dx.doi.org/10.1016/j.neuron.2009.09.027
89. Atz, M E, Rollins, B, Vawter, M P. **NCAM1 association study of bipolar disorder ans schizophrenia: polymorphisms and alternatively spliced isoforms lead to similarities and differences.** *PsychiatrGenet.* **17(2)**:55-67. doi: 10.1097/YPG.0b013e328012d850
90. Devon, R S, et.al. **The genomic orgainsation of the metabotropic glutamate receptor subtype 5 gene, and its association with schizophrenia**. *Molecular Psychiatry*. **6(3)**:311-314. (2001) doi:10.1038/sj.mp.4000848
91. Dermol, U, et.al. **Unique Utilization of a Phosphoprotein Phosphatase Fold by a Mammalian Phosphodiesterase Assocated with WAGR Syndrome**. *J Mol Biol*. **412(3)**:481-494. (2011) http://dx.doi.org/10.1016/j.jmb.2011.07.060
92. Imoto, I, et.al. **Identification and Characterization of Human PHNOX2, a Novel Homeobox-Containing Gene**. *Biochemical and Biophysical Research Communications*. **287(1)**:270-276. (2001) http://dx.doi.org/10.1006/bbrc.2001.5578
93. Wasim, M, et.al. **PLZF/ZBTB16, a glucocorticoid response gene in acute lymphoblastic leukemia, interferes with glucocorticoid-induced apoptosis**. *Journal of Steroid Biochemistry and Molecular Biology*. **120(4-5)**:218-227. (2010) http://dx.doi.org/10/1016/j.jsbmb.2010.04.019
94. McWilliams, R, et.al. **Characterization of an ankyrin repeat-containing Shank2 isoform (Shank2E) in liver epithelial cells**. *Biochem J.* **380(Pt 1)**:181-191. (2004) doi:10.1042/BJ200311577
95. Schneider, P M. **Recovery of high-molecular-weight DNA from blood and forensic specimens. Forensic DNA profiling protocols 1-7**. *Method Mol Cell Biol.* **98** (1998).
96. Sambrook, J. ***Molecular Cloning: a laboratory manual***. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001).
97. Sadaki S and Shimokawa, H. **The amelogenin gene**. *Int J Dev Biol*. **39**, 127-133 (1995).
98. Delgado, S, Girondot, M and Sire, J. **Molecular evolution of amelogenin in mammals.** *J Mol Evol*. 60(1)**:**12-30. (2005)
99. Sire, J, Delgado, S and Girondot, M. **The amelogenin story: origin and evolution.** *Eur J Oral Sci.* 114(s1):64–77. Article first published online (2 MAY 2006).
100. Gibson, C W, Collier, P M., Yuan, Z A. & Chen, E. **DNA sequences of amelogenin genes provide clues to regulation of expression**. *Eur J Oral Sci.* **106(1)**:292-8. (1998)

101. Blin, N, Stafford, DW, **A general method for isolation of high molecular weight DNA from eukaryotes**. *Nucl. Acids Res*. (1976) 3 (9): 2303-2308. doi: 10.1093/nar/3.9.2303
102. ***Quantifiler® Duo DNA Quantification Kit User's Manual*** © 2012 Life Technologies Corporation.
103. Matthias Meyer *et al.,* **A High-Coverage Genome Sequence from an Archaic Denisovan Individual.** *Science* 338, 222 (2012). DOI: 10.1126/science.1224344

Supplementary Information is linked to the online version of the paper at ?????????????.

Acknowledgements: Role of the Funding Source -We wish to acknowledge and thank Mr. Wally Hersom of Henderson, NV. Without his extremely generous financial contributions and his absolute commitment to seeing this research through to its completion, this project would not have been successful. We also would like to thank Mr. Adrian Erickson of Osoyoos, British Columbia, Canada for his generous funding. The small initial project would not have grown into this study without his generous support. Furthermore, we appreciate the other donors: David Paulides, Robert Densford, Richard Stubstad and Larry Surface who contributed both financially and in the collection of hominin samples in the field.

We would like to thank Adrian Erickson of the Erickson Project for his generosity in providing the still and the video included in this paper (Figure 4, Supplementary Video 1). We also thank Bob Durdella for his pictures of the collection site for Sample 168 and Bruce McDonald for the linear trackway photo.

We thank Dr. Colm Kelleher for critical review of and for his efforts in editing and revision of the manuscript. We also thank in alphabetical order Carlos Arana, Dr. Igor Burtsev, David Claerr, David Daleiden, Adrian Erickson, Clint Faile, Jane Harris, Tony Harris, Wally Hersom, Troy Hudson, Cecile Jernigan, JC Johnson, and Robin Lynne, Tim Olson, and David Paulides, for their help in providing Internet support, books, field reports, moral support and numerous enlightening discussions pertinent to advancement of this study and the development of http://www.sasquatchgenomeproject.org .

Also, we would like to acknowledge the three groups that provided the DNA for the 3 whole genomes sequenced in alphabetical order:
Stan Courtney Research Group (Stan Courtney)
The Erickson Project (Adrian Erickson, Dennis Pfohl, Dr. Leila Hadj-Chikh, Randy Brisson, Raymond Brisson and Zoltan Bartha)
The Olympic Project (Wally Hersom, Derek Randles, Richard Germeau, Steve Mattice, David Ellis)
We also would like to thank the field researchers and associated groups in alphabetical order for their patience and generous sample contributions:
Arizona Cryptological Research Organization, (Alex Hearn)
Crypto Four Corners (JC Johnson, Brenda Harris)
North America Bigfoot Search (David Paulides, *et al.*)
Stan Courtney Research Group (Stan Courtney)
TexLa Cryptozoological Research Group (Robert Densford, *et al*.)

The Erickson Project (Adrian Erickson, Dennis Pfohl, Leila Hadj-Chikh, Randy Brisson, Raymond Brisson and Zoltan Bartha)
The Olympic Project (Wally Hersom, Derek Randles, Richard Germeau, Steve Mattice, David Ellis)
WRAITH , (Dr. Al Guinn, *et al*.)
Additionally, we would like to thank the individuals who actually collected the samples in alphabetical order: Dr. Robert Alley, Zoltan Bartha, Susan and Raymond Bilbrey, Randy Brissom, Dr. Igor Burtsev, Dr. Angelo Capparella, Scott Carpenter, Janice Carter, Stan Courtney, Robert Densford, Bob Durdella, P. Freeman, Richard Germeau, Mary Green, Dr. Al Guinn, Darwin A. Greenwell, J. (John) Richard Greenwell, Brenda Harris, Jennifer Hilaman, Makayla Hilaman, Larry Jenkins, Shawn Jones, Dr. Melba Ketchum, W. Laughery, Steve Mattice, Bruce McDonald, Dr. Curt Nelson, David Paulides, Michael E. Pentycofe, Dennis Pfohl, Derek Randles, Dr. Webb Sentell, James Shirley, Justin Smeja, Kirk Stewart, W. Sumerlin, Larry Surface, Raven Ulibarri, Doug Waller.

Analysis for Hire Laboratories Used in the Blind Study:
The following laboratories provided sequencing and analysis of samples in the study on a work for hire basis and were not initially told the origin of the samples being tested until after the samples were tested:
Family Tree DNA Genomics Research Center, 1445 North Loop West, Suite 820, Houston, TX 77008
SeqWright, Inc., 2575 W. Bellfort St. Suite 2001, Houston, TX 77054
UT Southwestern Medical Center, 6000 Harry Hines Blvd. NA7.116, Dallas, TX 75235-9093
USC Norris Comprehensive Cancer Center**,** 1441 Eastlake Avenue, Los Angeles, CA 90033
Texas A&M University, Microscopy & Imaging Center, Department of Biology and Department of Biochemistry & Biophysics, College Station, TX 77843-2257
Texas Veterinary Medical Diagnostic Laboratory, College of Veterinary Medicine, Texas A&M University, College Station, Texas 77843
Southwestern Institute of Forensic Sciences, 2355 North Stemmons Fwy., Dallas, TX 75207

Author Contributions: M.K., P.W., D.S., A.H., R.S., and R.S. performed experiments. M.K, and F.Z. analyzed the genetic data. M.K., A.W., and P.W. wrote and edited the manuscript. A.H. analyzed and wrote the EM portion of the manuscript. D. S. analyzed and wrote the hair analysis portion of the manuscript. D.T. analyzed and wrote the histopathology portion of the manuscript. A.W. also researched pertinent additions to the manuscript and helped with data collection. M.K. distributed samples, collected and combined data from the blind studies.

Author Information: The raw sequence data utilized in this manuscript from the Sasquatch samples has been provided in the Supplemental Data. Reprints and permissions information is available at ***Denovo*** (www.denovojournal.com). The authors declare no competing financial interests.
Correspondence should be addressed to M.K. (science@sasquatchgenomeproject.org) or A.W. at (awatts02@charter.net)

Supplementary Figure 1 Phylogenetic tree generated from sequencing the whole mitochondrial genome of Sample 26.
Supplementary Figure 2 Phylogenetic tree generated from sequencing the whole mitochondrial genome of Sample 31.
Supplementary Figure 3 Phylogenetic tree generated from sequencing HV1 of Sample 140.
Supplementary Figure 4 Phylogenetic Tree generated from selective supercontigs for Sample 26.
Supplementary Figure 5 Phylogenetic Tree generated from selective supercontigs for Sample 31.
Supplementary Figure 6 Phylogenetic Tree generated from selective supercontigs for Sample 140.
Supplementary Figure 7 Phylogenetic tree generated from the mitochondrial genome extracted from the next generation whole genome from Sample 26.
Supplementary Figure 8 Phylogenetic tree generated from the mitochondrial genome extracted from the next generation whole genome from Sample 31.
Supplementary Figure 9 Phylogenetic tree generated from the mitochondrial genome extracted from the next generation whole genome from Sample 140.