Authors' Response to Review

Referee B

1. This is clearly important information that I hope the public will have access to soon. However, I was immediately taken by surprise after thoroughly reading the manuscript to see such a high reference to Hominins, including the title. I was surprised because there is no substantial evidence presented by the author that the species identified in the 3 whole genomes is a biped. Eye witness accounts are the only data presented to determine or substantiate a biped of any kind. It would be much more appropriate to delineate in this manuscript that novel genomic evidence highly suggests an unknown species living contemporaneously in the Continental and Subarctic United States and Canada within the Order: Primate. There is no conclusive evidence that the unusual data found in the follicular hair morphology or the tremendous genome results indicates a "hominin". The tribe Hominini (of Homininae) is reserved for Homo, post Panini divergence. Genetic information alone is not enough to classify the species within a tribe or clad under order - Primate. Though there is substantial alignment in some data relating to hair and other DNA studies, it is too presumptuous to purport the extant data included in the manuscript conclusively classifies this unknown species as hominin. An example (to the author) of how to, perhaps, revise:

---*The genomic data of sample 26 indicates the specimen allegedly collected from an unknown creature is certainly novel and indicates an unknown species. The phylogenetic trees of sample 26 indicate a species highly aligned with the order: Primate. If the alleged story related to us by the hunter who submitted the sample, the species is closely related to the hominin tribe because*

the hunter claimed to have seen the creature walking "upright". Our chromosomal analysis would substantiate a close hominini relationship. Further study and analysis will be ongoing to determine such classification. The phylogenetic trees are quite compelling and are probably the most substantial of the information given within the manuscript. However, the information and data is not satisfactory to hypothesize anything more than a living unknown primate. The molecular genetic data is quite compelling and enough to publish in and of itself. But in order to use taxonomic language in this manuscript, especially related directly to the species itself- is inappropriate. In order to include the level of specific taxonomy presented in the manuscript such as hominini, gross anatomical evidence along with molecular physiological evidence must be included in the manuscript. Even a clear video of this species actually walking bipedal would give some sort of gross anatomical evidence. But eye witness accounts are not enough to make such hypotheses. I do, however, believe the eye witness accounts are important and should not be excluded. They are just not appropriate to utilize for evidence to propose any hint of taxonomy, including hominini.

Author's Response:

Pan/*Homo* divergence occurred between 5.4 and 6.3 million years ago. Since the mtDNA speciation is not only unequivocally modern human in 100% of the samples included in this study, and furthermore, can be dated by the haplotypes obtained in this study to as late as 13,000 to 26,000 years ago including the mtDNA from the 3 Next Generation genomes, these individuals must be included as hominins. The human mitochondrial DNA haplotypes were

consistent between the Next Generation Sequencing and the original mtDNA whole genomes sequenced in the beginning of this study.

1. Bradley, B.J. (2006). "Reconstructing Phylogenies and Phenotypes: A Molecular View of Human Evolution". *Journal of Anatomy* **212** (4): 337-353. doi:10.1111/j.1469-7580.2007.00840.x.
2. Wood and Richmond.; Richmond, BG (2000). "Human evolution: taxonomy and paleobiology". *Journal of Anatomy* **197**: 19–60. doi:10.1046/j.1469-7580.2000.19710019.x. PMC 1468107. PMID 10999270.

2. The molecular genetics in this manuscript are the most important and it would be important to include information regarding the analysis of the whole genomes. The phylogenetic trees are exciting to look at and speculate about, but there is not enough analysis to determine phylogenicity. I highly suggest including language that 'it is recognized that continued analysis of the nuclear DNA will be required in order to determine phylogenicity. The genomic information is certainly impressive, but not as conclusive as the manuscript proposes. It is inappropriate to make connections between the various samples. In fact, I believe this weakens the manuscript due to the reality that the genomic data and mtDNA analysis does not align well against each other (except for 2 of the whole genomic sequences). There are similarities, and those similarities are important to note. But conclusions cannot be drawn about the inter-relatedness of these various samples. The number of samples do indicate high alignment with Homo along chrm11. However, there are many other novel sequences across the several samples

that creates confusion for the reader. Perhaps it may be helpful to split the manuscript up into several different studies- or simply divide the manuscript up into major sections to delineate the different evidenciary aspects. But to attempt to make links between the hair samples and the novel genomic sequences and other DNA evidence is premature. Tremendous work would need to take place with substantial evidence, including gross anatomical evidence, to make such linkages between the various samples in the study.

Author's Response:

Mitochondrial DNA has been universally utilized to determine species by the scientific community. Since the mtDNA in Sasquatch is not only unequivocally modern human in 100% of the samples included in this study, that places the Sasquatch as human. However, since the nuDNA is novel to a large extent, the speciation of a potential progenitor is what is in question, especially since there are gene sequences that align 100% with human interspersed in the nuclear genome.

3. The other major problem with this manuscript is, again, the alleged specimens these samples came from. It is important to state the samples came from alleged Sasquatch specimens, which is done well throughout the manuscript, but not consistently. It would be helpful, however, to include additional evidence to corroborate the theory that the samples came from a large non-human biped. Video evidence of this species walking bipedal would be important to help strengthen the manuscript. In addition, it is inappropriate to make statements regarding the dynamic between Homo Sapien DNA and the whole genome sequence. It is appropriate to include statements about distant relationships with the order- Primate. But it is certainly

premature and the evidence is not conclusive enough to make such close connections between this unknown species and Homo Sapiens. I do believe that there is strong data in the whole genomes to suggest alignment with Homo Sapiens. However, there is not enough evidence to make statements about hybridization events. Hybrid phenomenon across the Homo/Pan/Gorilla taxonomy are currently highly debated in the scientific community. Making such claims is far too premature for a manuscript like this. The manuscript demonstrates a universal bias toward a hominini hypothesis. This anthropomorphic bias is problematic for this manuscript. A scientific manuscript is intended to present facts, not biases. If the editorial board decides to publish this manuscript with such a clearly "hominid" favoured bias, they certainly should proceed, but I recommend doing so with a strong disclaimer that an authored bias does exist in the manuscript.

Author's Response:

Since the mitochondrial DNA places them as human, in both the original sequencing of the mitochondrial whole genomes as well as the Next Generation Sequencing, it is not premature to align them with Homo. We did add another hypothesis in addition to the hybridization theory on lines 692-696. In consideration of the mitochondrial whole genomes, these are the only two viable hypotheses available.

5. The inclusion of the Q30 scores was very important for the author(s) to do. This data strengthens the manuscript more than anything else included (if editors wanted to publish based on the Q30 scores alone, then there would be no reason not to publish). Furthermore, I am also

most impressed with the presentation of the phylogenetic trees for the 3 whole genomes. They clearly identified primate relationships and alignment- though sample 31 seems to be very different from the other two whole genomes. That is concerning and confusing. It may make more sense to remove sample 31 altogether until the study can further understand and explain the huge difference with sample 31 (including conclusive evidence, not speculation). I am also pleased with the inclusion of the high definition pictures of human hair morphology versus the unknown species comparison. However, it would strengthen the manuscript to cross reference human morphology so as to substantiate the statements regarding the comparisons. I am also pleased to see the substantial amount of data and work taken to ensure high quality DNA samples with clear and appropriate techniques to reduce and eliminate contamination; this process was well done and substantiated.

Author's Response:

References 15-19 refer to hair analysis and human hair morphology as well as animal hair analysis.

As far as Sample 31, it should be included because it has the same characteristics as all of the other samples, human mitochondrial DNA and novel primate sequence in the genome as well as human sequence. In the original manuscript we discussed that eyewitness accounts clearly note physical differences related to geographic location in North America. We feel that 31 is such a variant. We had taken this information out of the manuscript but will re-introduce it if requested. The fact that overall, the genome is consistent with the others as far as its makeup clearly establishes its place in the manuscript.

6. The pictures included thus far of a supposed creature that is claimed to be a Sasquatch is of terrible quality and should be replaced with much more clear photographic evidence or video. These pictures weaken this manuscript. The inclusion of an image of sticks being put together in some kind of pyramid is inappropriate. I would strongly suggest removal of this image.

Author's Response:

We included the stick structure because one of the samples (168) was obtained within this structure. Not only did this document chain of custody for this sample but the fact that a viable sample was found within the structure supports the discussion of eyewitnesses encountering unusual stick structures within areas purported to be inhabited by the Sasquatch. As far as video, we are including Supplemental Video 1 of a juvenile female Sasquatch sleeping. This video is from the Sasquatch that Sample 37 was obtained from which also lends credibility and chain of custody to that sample. The still photos have been removed from the manuscript with the exception of Figure 4 which has now been edited to tie it to Supplementary Video 1 on lines 109-110 since the video is clearer and in hi def.

Referee A

I appreciate having had the opportunity to review this interesting investigation into such a controversial subject and recognize the enormous constraints under which the authors and their collaborators have labored.

Major concerns with this manuscript:

1. A difficult two-part thesis is posed that is inadequately substantiated by the analysis presented in the manuscript. Both parts (i.e. Part 1 - previously uncharacterized hominins exist in North America; Part 2 said hominins are the descendants of a putative hybridization event involving an ancient uncharacterized hominin and a modern human). A thesis this complex and counterintuitive requires significantly more in-depth analysis and consideration and should be developed through a series of peer-reviewed publications. Add in the idea that more than one species of hominin may be present in North America and the effort to make a convincing case is multiplied.

Author's Response:

We removed the discussion of the different variants of Sasquatch living in North America simplifying the manuscript and added an alternative hypothesis outside of hybridization to the discussion so further discussion of their origins can be fully developed in future manuscripts. Lines 688-696.

2. The presentation of the thesis is overburdened in three ways: a) the considerable inclusion of unsubstantiated eyewitness accounts and folkloric/pop-cultural ideations and their presentation as foundational facts that validate the analytical data; b) the presentation of a series of anomalous results generated using routine methodologies resulting in a sub-narrative that is curious and interesting, but unnecessary given the fact that cutting edge technology has been employed to generate sequence data sets that should allow for incredibly detailed analyses yet lead to relatively unambiguous conclusions; and c) the results from Next Generation Sequencing

methodologies, especially related to the analysis of the nuclear genome are fairly superficially treated so as to be unclear whether the authors have selectively aggregated sequences to support a favorable placement among primates. As stated in Perleman et al. (2011; PLoS Genet 7(3): e1001342. doi:10.1371/journal.pgen.1001342, "...primate taxonomy is both complex and controversial, with marginal unifying consensus of the evolutionary hierarchy of extant primate species," and given the expertise of one of the authors in this particular area, it is a bit astonishing that a more in-depth and stepwise treatment is not provided. The decades of work published by the Goodman lab provide an excellent roadmap for a convincing analysis. In summary, this work would achieve a much higher degree of analytical credibility by a) limiting folkloric/cultural references to the minimum necessary, b) moving many of the references to anomalous results to a supplementary role (2nd publication, perhaps), c) fully leveraging the treasure trove of information yielded by the Next Generation Sequencing.

Author's Response:

a) The introduction with popular information was requested by another reviewer. Since there is broad interest in this manuscript, that reviewer felt that the inclusion of such information was important. I will leave it up to the editor whether this should be included or not. It should be noted that 2 of the authors of this paper did have eyewitness encounters at the invitation of people that have interaction with this species so it is not entirely unsubstantiated.
b) The previous results were left in the manuscript because they supported the next generation sequencing results and also because they were obtained from a large number

of purported Sasquatch samples, which behaved in the same manner as the three samples utilized for Next Generation Sequencing.

c) The purpose of the manuscript was to prove that the Sasquatch exist. We chose chromosome 11 as our starting point for building our supercontig in order to generate the phylogenetic tree since the scope of the manuscript was to determine whether a novel primate did exist in North America. This was the logical way to proceed and with the length of the contigs and also with more than 379 genes analyzed (for another manuscript in the future) we feel that this was the direction to go to prove that they exist and as a preliminary study. In the beginning of the study, we had already ruled out non-primate species with not only the mtDNA findings but also with PP16 since it only gives results for ape and human with the exception of the 103 peak at Amelogenin that appears when there is animal DNA present in a sample. The 103 peak was not found in any of the purported Sasquatch samples that were screened using PP16. Furthermore, fully leveraging the genomic data will take years. This manuscript is just the beginning of the study.

3. A significant amount of work leading to the generation of the data central to the thesis was outsourced. It is very difficult to manage the quality control and release standards of contract laboratories from the outside. This does not automatically call such data into question. However, when coupled with a significant amount of anomalous and ambiguous data, such concerns must be considered. This would be especially true concerning Next Generation Sequencing of unknowns where heterologous libraries may result from contamination. This technology is much more sensitive to the detection of contaminating gene sequences than others, and the possibility that contaminating sequences may affect the development of a consensus sequence should be

given considerable attention. Reference bias should also be given consideration in the development of a consensus sequence. Similar work typified by that of Pruefer et al. (Nature, 2012/06/13/online) could provide good guidance concerning appropriate quality control analysis.

Author's Response:

The outsourced laboratories were accredited and work with human samples, which require regulation. We have correspondence where the laboratories were confused about the results and in some cases doubled checked their findings. Furthermore, more than one laboratory repeated the analysis at least to some degree supporting the data of other laboratories. They all utilized the same extractions that had been aliquoted. As far as the Next Generation data, it confirmed the mitochondrial findings from previous laboratories as well as novel sequence. That is why it was important to leave the original data in the manuscript. Furthermore, according to supervisors in data analysis at Illumina, the Q30 quality scores would have been remarkably lower if contamination had been present in the genomes due to competition by the various sequences. That also would have affected the mitochondrial findings which were consistent with the original mtDNA sequencing.

4. Conflicting statements occur concerning the validation against submitter contamination:

"Control DNA was obtained from the majority of the submitters and was profiled using Promega PowerPlex® 16 (20). All submitters yielded complete profiles."

Author's Response:

This has been corrected to clarify on line 200: All submitters "tested"

At the very least the STR profiles for the submitters of Samples 26, 31, and 140 should be presented, as these samples yielded the most significant data of the study.

Author's Response:

Since the submitters' names are given per their request contractually in Table 1, it would be inappropriate and unethical to publish their DNA profiles as it would link to their identity. However, we have them on file.

5. Sample 26 - It is stated in the study that Sample 26 was derived from a shooting incident. The inclusion of such a sample in the study may be inconsistent with contemporary scientific ethics concerning the treatment of both human and animal study subjects and the procurement of research specimens from such. Moreover, this description casts the provenance of such a sample as murky, at best.

Author's Response:

We have removed any references to the alleged shooting incident. Line 481

6. The bioinformatics should include gene sequences from expected outlier species that may also be capable of contributing contaminating nucleic acids. For example, a BLASTN search using Sample 26 does turn up some exceptionally strong homology with a gene from Ursus americanus

(DQ240386.1). This would support the idea that the consensus sequence may have been affected by contaminant sequences.

Author's Response:

There will always be some homology with other species when short random sequences are chosen, however, your example of bear contamination can be completely ruled out considering none of the laboratories handling the samples have bear samples. Furthermore, forensic screening would have given a 103 peak at amelogenin on PP16 should there have been non-primate contamination. Sequencing of the mitochondrial DNA with universal primers also would have shown any contamination of the original extractions with non-human DNA. Additionally, that is a single isolate from the black bear 7193328 brain-derived neurotrophic factor. Not only is it preset in Sample 26, but it is just as present in Sample 140. Furthermore, DQ240386.1 is 489 bases. I wonder why such a small sequence, ie 489 bases out of 2.7 million bases was the focus of this critique. DQ240386 is statistically significantly aligned with primates and carnivores. In fact, BLASTing DQ240386- ring tailed cats of the raccoon family and seal have as much alignment as Ursus americanus. The maximum score for raccoon and seal are about 850. Maximum score for Ursus americanus is about 900. Max score for Sample 26 is 538. This shows contamination bias.

Minor concern

1. The use of "hominin" in the introduction, as opposed to hominid, creates the perception of bias toward folkloric and unverified eyewitness accounts that seem to emphasize human

characteristics. The general discussion of such creatures tends to split along two lines: human or human-like and ape or ape-like. The term hominid, as related to the Family Hominidae, is the higher order, and, thus, the broader grouping. Visual description from a distance might allow for a reasonable description of a hominid, but close examination of the organism or its remnants is required for its placement into the Hominini. Thus, to establish initial objectivity, the authors should use a qualifying adjective such as "putative."

Author's Response:

Pan/*Homo* divergence occurred between 5.4 and 6.3 million years ago (refs below). Since the mtDNA speciation is not only unequivocally modern human, but can be dated by the haplotypes obtained in this study to as late as 13,000 to 26,000 years ago, these individuals must be included as hominins.

We did add the word "putative" in line 85.

Bradley, B.J. (2006). "Reconstructing Phylogenies and Phenotypes: A Molecular View of Human Evolution". *Journal of Anatomy* **212** (4): 337-353. doi:10.1111/j.1469-7580.2007.00840.x.

Wood and Richmond.; Richmond, BG (2000). "Human evolution: taxonomy and paleobiology". *Journal of Anatomy* **197**: 19–60. doi:10.1046/j.1469-7580.2000.19710019.x. PMC 1468107. PMID 10999270.