# Preliminariy  Report of DNA Study from Peruvian/Nazca Trydactil Mummies.

**Part One Principal Research and Part Two Canadian Study Compiler:**

**Ricardo Rangel Martínez**

**Biologist.**

# Part One.

**Public Hearing UAPS Phenomenon**

**Mexican Congress.**

*1sth Public Hearing UAPS Phenomenon. Mexican Congress*

*September 12, 2023*

*Statement by Biologist Ricardo Rangel Martinez:*

Continuing with the analysis of the biological entities referred to a few moments ago by Dr. Zalce, I have come here freely and voluntarily to give my interpretation of the studies carried out on the tissue samples obtained from the neck and hip of the individual called "Victoria" which were subjected to cleaning and extraction to DNA (deoxyribonucleic acid), to later carry out a DNA sequencing process, this in order to be able to identify the organism to which the specimen in question could be related.

Since these are samples of ancient biological tissues, which could be verified through carbon 14 tests, it was expected that the molecules that protect the genetic information of this individual could be altered, degraded or partially destroyed, this due to the effect of stress caused by multiple environmental variables.

Furthermore, it must also be considered that since the specimen was found in a cave, where it could have been in direct contact with microorganisms, insects, pollen, or any other living organism, the possibility that the DNA sample

Ricardo Rangel-Martínez

isolated from the aforementioned tissues could have been contaminated is very high.

Also, when interpreting the results, it must be taken into account that the people who found and handled the specimen did not do so with the necessary precautions to avoid contaminating it with saliva, respiratory droplets, or any other human biological tissue.

That said, a pilot study was carried out to verify the state of the DNA molecules, in order to subsequently carry out the sequencing process.

The extraction technique and its subsequent running on an agarose gel confirmed the existence of degraded DNA and small molecules of non-degraded DNA with the possibility of being able to carry out a sequencing process.

The technique of choice for sequencing ancient DNA, such as that of the sample in question, is the next-generation massive sequencing technique, which was carried out on an illumina® brand sequencer. This technology allows the simultaneous sequencing of millions of fragments obtained from the problematic DNA in a short time.

The sequencing results were as follows:

1.- From the specimen identified as WGS Ancient0002, (neck bone) 561,665,320 reading sequences were obtained.

2.- From the specimen identified as WGS Ancient0004, (muscle tissue of the hip) 501,700,245 reading sequences were obtained.

A reading sequence is made up of DNA fragments that are 150 nucleotides long.

These sequences were then uploaded to the platform called Sequence Reading Archive or SRA. This megadata analysis tool is free and is part of the National Center for Biotechnology Information of the National Library of Medicine of the USA.

This tool allows the comparison of all the sequences of readings obtained from each biological sample in question, and compares them with all the known genome sequences in the GenBank, totaling today more than 700 thousand. Although it is estimated that there are more than 7 million species on our planet, up to now at least one organism of all the existing taxa of living beings known

on the planet has already been sequenced. Therefore, with these 700 thousand genomes with which the comparison of the obtained readings is made, we are provided with a representative and significant result.

The aforementioned SRA tool provided us with the following results.

a) From the sample of neck bone tissue identified as WGS Ancient0002, 72.07% of the reading sequences were identified and 27.93% of the reading sequences obtained did not match the genomes of living beings known to date.

b) Of the 72.07% of the readings identified, 70.45% belong to contaminating DNA sequences from Homo Sapiens and the remaining percentage belongs to viruses and bacteria that also contaminated the sample.

c) From the sample of muscle tissue from the hip of the specimen identified as WGS Ancient0004, 36.28% of the reading sequences were identified and 63.72% of the reading sequences did not match the genomes of living beings known to date.

d) Of the 36.28% of the identified genomes, all turned out to be contaminating DNA from contemporary viruses, bacteria and plants, and the genome of no mammal, including humans, could be identified.


From this study, the following can be concluded:

1.- The two ancient biological samples contain contaminating DNA that could have been acquired during the manipulation of the specimen or over time, by organisms that surrounded it and by the people who manipulated the specimen.

2.- The sample identified WGS Ancient0004 showed the highest number of unidentified sequence readings, about 316 million readings.

3.- This high percentage of unidentified readings with the living organisms that we currently know should be the focus of future research to try to define the identity of the organism in question.

4.- There is a probability greater than 90% that this organism is not related to humans.

5.- There is a probability greater than 50% that this organism is not related to living beings known to date on our planet.


Ricardo Rangel-Martínez

6.- Genetic studies should be carried out on all specimens known to date in order to sequence their entire genome.

7.- At this time it is not possible to identify the specimen in question.

8.- All the gene sequences obtained in this study are public and can be freely accessed.

9.- It would be important to promote a multidisciplinary scientific committee to continue the study of anomalous biological entities, as well as for the study of UAPs

Thank you very much for your kind attention.

Ricardo Rangel-Martínez

*2nd Public Hearing UAPS Phenomenon. Mexican Congresss .*

*November 7, 2023.*

*Statement by Biologist Ricardo Rangel Martínez:*

Good afternoon everyone. Honorable members of the Chamber of Deputies. Honorable members of the presidium. Representatives of the media. Ladies and gentlemen. Thank you very much for your presence.

Once again, I present myself here freely and voluntarily to give my testimony and interpretation of the analysis of the sequences obtained from the DNA isolated from an organism identified as 0003 and which has been colloquially identified with the name "María".

As already stated in documentary evidence previously presented at this public hearing, tissue samples have been taken from this organism to subject it to a carbon 14 dating study that was carried out by UNAM, showing an approximate age of 1,700 years. The gold standard for the analysis of the deoxyribonucleic acid molecule from ancient samples, as well as its sequencing, is NGS or next-generation sequencing technology. These technologies allow us to perform parallel amplification of millions of sequences from pieces of DNA obtained from the ancient problem sample.

Again, it is important to emphasize that, in the genetic material obtained from ancient tissues, there is a high probability that it is degraded and also contaminated with DNA from microorganisms or people who have manipulated the specimen without the corresponding precautionary measures. Therefore, such events must be considered when interpreting them.

The tissue sample obtained from Maria's hand bone, which is identified as WGS ancient0003, was subjected to a DNA isolation protocol, to subsequently perform massive sequencing on an Ilumina™ brand device.

Subsequently, the sequencing results were entered into the public platform of the Reading Sequence Archives of the National Center for Biotechnology Information of the National Library of Medicine of the USA.

Ricardo Rangel-Martínez

The results obtained were the following:

As a result of the massive sequencing, 647,778,937 reading sequences were obtained, which in turn are made up of 150 nucleotides in length. Subsequently, each of these reading sequences was automatically entered into the SRA tool, which yielded the following phylogenetic construction.

Slide No. 1



Estudio filogenético de las secuencias aisladas de la muestra 0003
Momia "Maria". Muestra tomada del hueso de la mano.

As can be seen, of the 647,789,366, 97.38% of the reading sequences were identified, while 2.62% of the reading sequences could not be identified with the Gene Bank sequences. That is, 16,972,081 reading sequences did not match the living organisms known to date.

Of the 97.38% of reading sequences identified, a small percentage corresponds to DNA of microorganisms that contaminated the sample or that are part of the tegumentary microbiota, that is, 2.13% of the reading sequences isolated from the sample obtained.

Therefore, the remaining 95.25% correspond to DNA from the tissue obtained from the organism in question. As can be seen, there is a downward distribution in the percentages identified, as we move away from the ancestral lineages in the phylogenetic tree, until we observe the lowest percentage in the organism that is identified to genus and species. That is, this Reading Sequence Archive

tool manages to fully identify a large part of the genetic composition of the organism in question.

When paying attention to the final result, it stands out and draws attention, that 30.22% of the reading sequences are identified with the human genome, that is, Homo sapiens, but it also includes an anomalous taxonomic branch of the Pan genus, with 3.05% of the reading sequences obtained from the organism under study.

Slide 2.



Estudio filogenético de las secuencias aisladas de la muestra 0003
Momia "Maria". Muestra tomada del hueso de la mano.
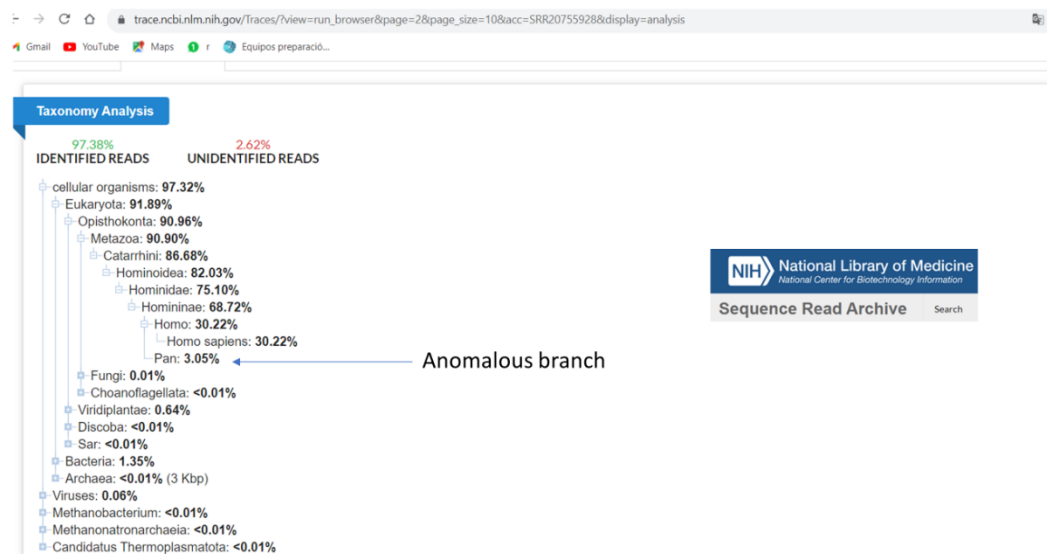
In order to investigate this result further, the crown analysis was deployed, which shows the percentage of reading sequences in a sector or Krona "pie" type graph. In this new diagram you can see the representation of the percentage of the genus Pan plus two hidden elements, which when you position the cursor over it and double click, the hidden information is displayed. Surprisingly, you can see that the new information displayed corresponds to the reading sequences identified in two species of apes, which are identified by genus and species, being Pan Troglodytes and Pan Paniscus, that is, the Chimpanzee and the Bonobo respectively. Both species are sisters.
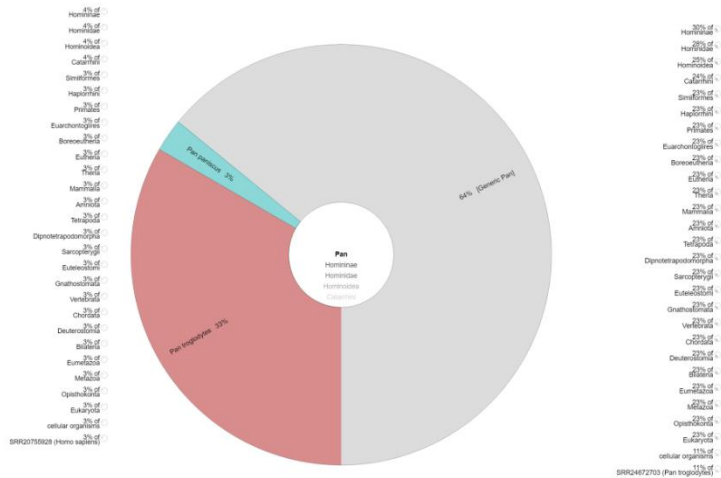
Estudio filogenético de las secuencias aisladas de la muestra 0003
Momia "Maria". Muestra tomada del hueso de la mano.
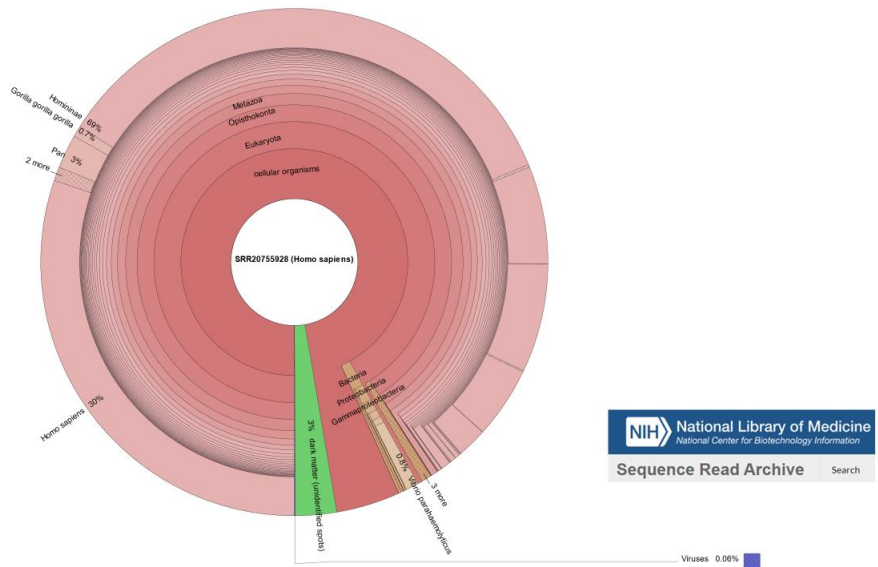


Espécimen 0003 "María"
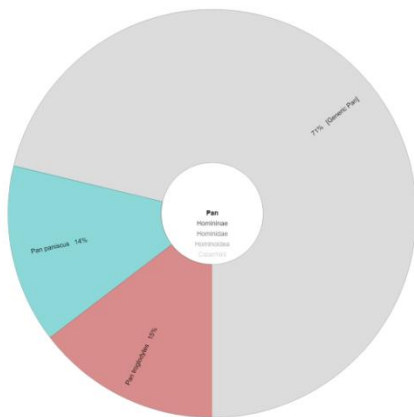


Célula Hibrida Chimpancé - Humano

These apes are only found on the African continent, so the origin of this species on the American continent more than 1700 years ago is a real mystery, which,

Ricardo Rangel-Martínez

according to the results of the sequencing, would have hybridized with a specimen of the genus Homo sapiens.

When faced with overwhelming evidence, we wonder if in this same file of reading sequences, there would be available readings of hybrid cells that have been carried out "in vitro" by some group of researchers. This was done in order to compare the pattern of the phylogenetic diagrams and the way in which the SRA program accommodates the data of a non-natural species, produced in the laboratory, using artificial techniques.

Slide 3.
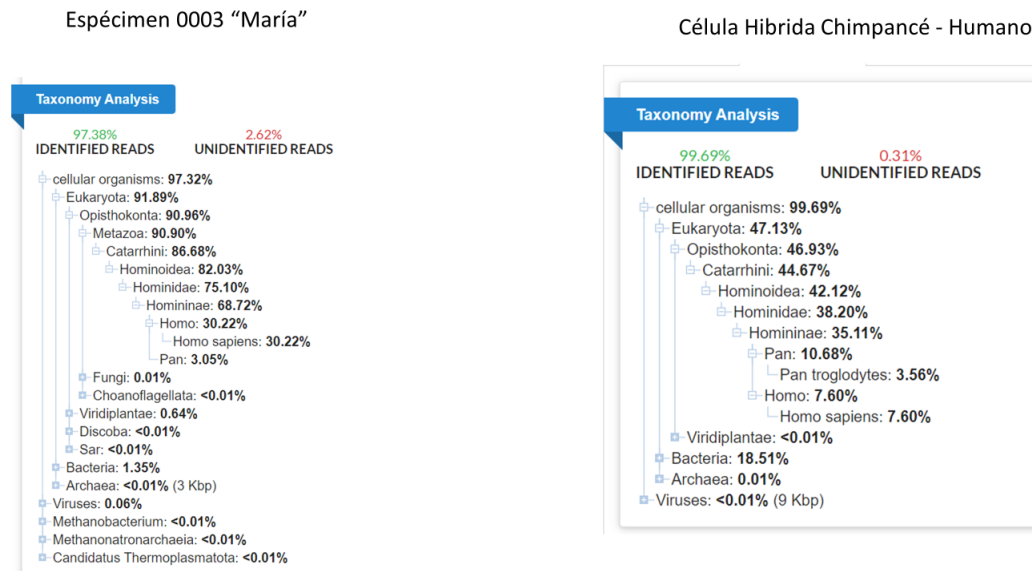


To our surprise, we managed to identify several experiments in which several hybrid neuronal cells between Homo sapiens and Pan troglodytes were obtained.

As can be seen, the treatment given to the data by the program and the phylogenetic analysis are very similar. Therefore, it can be concluded that there is a probability greater than 99% that the specimen in question named "María" is actually a hybrid organism.

**New taxonomic nomenclature proposed by Biologist Ricardo Rangel for the hybrid called Maria**

*Pan troglodites*          ***Homopan tridactyla***          *Homo sapiens*

**Possible convergence due to contemporary hybridization**

Possibly a new species, which presents unique characteristics and which could be explained by the 2.62% of unidentified readings which could be responsible for the peculiar phenotypic characteristics such as tridactyly of the upper and lower limbs.

**Possible genealogical tree of the María mummy**.



Ricardo Rangel-Martínez
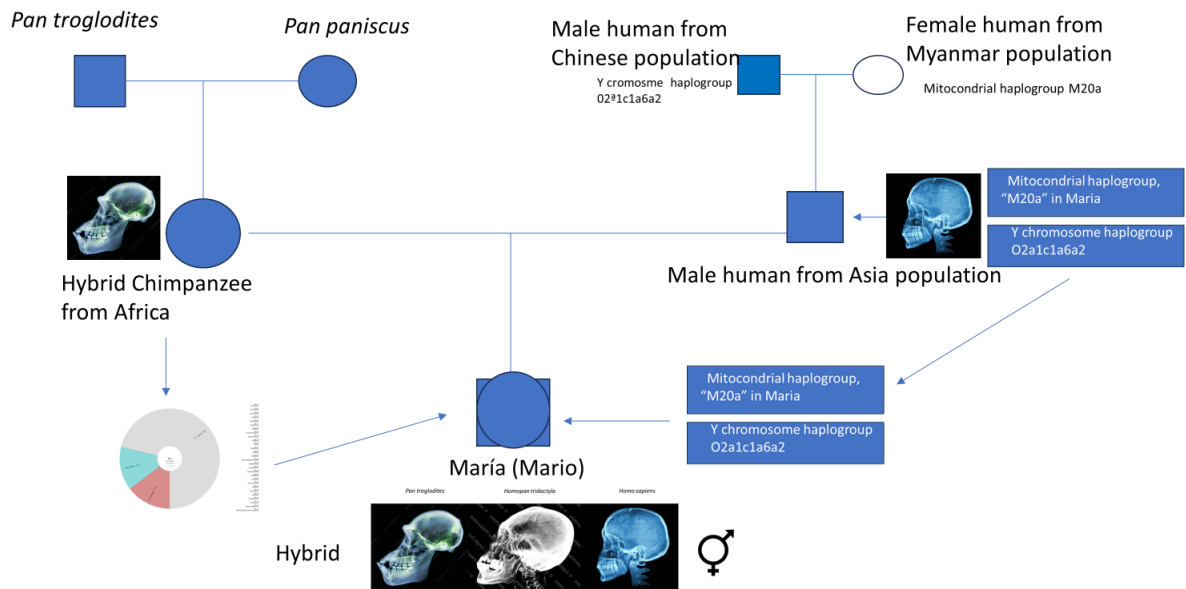
It will be important to create an international research group made up of anthropologists, primatologists, evolutionists, and scientists from other branches of science, in order, first of all, to ratify the data obtained up to this point, and to continue with the studies that will allow us to answer many questions about the origin of this organism, as well as to assign it a taxonomic nomenclature.

The opinion of leaders in the study of the hybridization process as a mechanism for generating new species, such as the doctor of science Eugene McCarthy, or the paleontologist Madelaine Bôhme, the anatomist Carol Ward and many other researchers who have dedicated their entire lives to research on the origin of man, would be very valuable.

We are faced with the most important biological evidence in modern human history; the international scientific community must be humble and not judge without delving into the case. The information on this case is public, as is access to the specimen, so that anyone who wants to join the research can do so from their area of expertise. This is what should unite us as human beings: the search for the truth.

Thank you very much for your kind attention.

# Part two

**Preliminary results of Canadian Study of DNA Peruvian Mummies.**

*Mummy's The Word: A Genomic Look at Peruvian Mummies*

## Abstract

Three ancient mummified remains, linked to the Nazca mummies, have prompted speculation about their origins, particularly in communities interested in the possibility of extraterrestrial life. We analyzed the DNA of these three remains to see what genetic information might be there.

All three samples showcased aged, degraded DNA, typical of ancient remains, and were riddled with contamination from minuscule organisms, mainly bacteria—common for environmentally exposed samples. Human DNA emerged in all three mummies, with one aligning quite significantly with the human genome, but in a way that creates more questions than answers. Diving deeper into the unmatched DNA snippets, we assembled them, finding that most that were classifiable matched with known bacteria.

For the mummy with substantial human-like DNA, an extra mitochondrial DNA check of maternally inherited DNA showed its membership in the human mtDNA lineage "M20a". Instead of linking to pre-European-contact Americas lineage, it flagged a specific southeast Asian maternal lineage, suggesting origins beyond a millennium-old Peruvian cave and opening up a range of further questions.

Through different alignments, assemblies, and analyses, our findings–encapsulated in SPAdes and Megahit assemblies, hg38 alignments, VCFs and reports, and available once we find hosting for it–suggest ancient DNA coupled with contamination. Nearly half the reads remained unclassified across all samples.

Ricardo Rangel-Martínez

Further analysis could yield more insights. This might, for instance, include BLASTn/p analysis of contigs, refining MEME motifs to exclude repeats, or running anomaly detection on the alignments, unmapped reads, and de novo assemblies—all of which we're sharing for further analysis.

While there's no evidence of extraterrestrial species, this project provides a clear roadmap and methodology for scientific inquiry into the unknown. We've shared our methods and findings on GitHub, and are seeking a secure digital haven for our cleaned-up data. This venture underscores science as an open door to probing some of the most interesting questions of our times, inviting anyone who is interested to join this genomic journey into the Nazca mummies.

## Introduction

## Background and context

Many early South American groups practiced mummification. In many arid and high-altitude climates, bodies would naturally desiccate, which several groups used to their advantage in preserving and honoring their dead. Several examples of these bodies have been discovered and examined by scientists since the 19th century (MUMMYSTUDIES).

Between 2016-7, multiple tridactyl—i.e., three-fingered—Nazca mummies were discovered in Peru. One group of five tridactyl mummies was found in a tomb in Nazca, Peru in January 2016, while another (named Maria) was found in a catacomb on the outskirts of Nazca, Peru in June 2017. Various reports suggest the mummies date back as far as 6,500 years, with some being found in a Peruvian tomb and examined by Russian scientists who claimed a different anatomical structure to humans.

The mummies were taken into custody by the Universidad Nacional San Luis Gonzaga in Peru, which aimed to exhibit them and promote their research. An event on November 6, 2019, at the university marked the official recovery of some of these mummies and launched an international call for researchers for the scientific study of these relics.

On September 13, 2023, the journalist Jaime Maussan presented two of these mummies to the Mexican Congress, claiming they were about 1,000 years old based on carbon dating process conducted by Mexico's National Autonomous University (UNAM). However, the scientists from UNAM only determined the age of the samples provided and did not draw conclusions about their origins.

The tridactyl Nazca mummies unearthed in Peru have captivated the imagination of both the public and certain factions within the scientific community. With their three-fingered hands and claims of extraterrestrial origin, these relics evoke the thrilling possibilities of the unexplored and the unknown, while also offering opportunities for serious and destigmatized scientific exploration.

The mummies' DNA data is freely accessible (Ancient0002, Ancient0003, Ancient0004), yet the expertise to interpret them is not as widespread. Thus, we identified an opportunity to apply robust scientific methodologies to explore the secrets within the DNA of these mystifying entities—and in doing so, demystify the narrative surrounding them.

Victoria's Mummie Secuencing Reading Archive:

Ancient0002        https://www.ncbi.nlm.nih.gov/sra/PRJNA869134

Ancient0004        https://www.ncbi.nlm.nih.gov/sra/PRJNA861322

Mary's Mummie Secuencing Reading Archive:

Ancient0003        https://www.ncbi.nlm.nih.gov/sra/PRJNA865375

Bioinformatics as a key methodological approach

Bioinformatics marries biology with computer science, unlocking the ability to dissect the genomic essence of these ancient beings. Because of its ability to process large amounts of biological data, bioinformatics can analyze dense and diverse samples as well as identify unexpected connections.

For example, as we discuss later, the discovery of mtDNA haplogroup "M20a" in one of the samples was a notable find, linking it to an Asian lineage rather than the expected pre-European-contact Americas lineage. Analysis on this scale would not have been possible without the tools that bioinformatics offers.

**Research question**

Our quest was simple: Analyze the Whole Genome Sequencing (WGS) samples of these mummies, unearthing any genomic insights they may harbor.

Ricardo Rangel-Martínez

This reflects a broader twofold goal: to collect credible data about the genomic information contained in the samples, and to democratize scientific inquiry. The latter is a crucial project in an era where information is abundant yet misinformation is rampant, and we aimed to empower other people with the knowledge and tools to delve into such intriguing questions themselves.

**Methods**

Data preparation

We began with the three  samples of ancient DNA from the NCBI Sequence Read Archive (SRA), uploaded by the Instituto Politécnico Nacional of Mexico in 2022. The samples, named Ancient0002 and Ancient0004, came from a mummy nicknamed "Victoria," while Ancient0003 came from a Mary's mummy.

To prepare clean data for analysis, we followed a series of steps to improve data quality.

Using standard SRA-Toolkit prefetch, we pulled the data from SRA, and fasterq-dump helped split it into readable parts. To tackle duplicate reads we used BBMap, and trimmomatic helped trim away any undesired adapters from the sequencing process, leaving us with neat sequences ready for the analytical stage.

Then, to evaluate and control quality, we ran FastQC to assess the quality of our samples. Our deduplication was moderate, and so even after processing Ancient0002 and Ancient0004 still had relatively high duplication levels (31.62% and 39.79% respectively), which suggest degradation or contamination leading to artifacts in the amplification process. On the other hand, Ancient003 had a lower duplication level of 15.29%. Besides a small rise in GC content for Ancient0003—still consistent with human DNA—all other quality metrics stayed within normal ranges, indicating we were working with ancient DNA and could proceed.

Taxonomic classification

We chose Kraken2 (KRAKEN2) for taxonomic classification. It's fast and accurate, producing useful and easy-to-analyze data. In addition, Kraken2 is popular, well-understood, and actively maintained in the scientific community.

We chose the comprehensive "nt" database to run Kraken2 against. Using this well-established reference required a robust computer setup: provisioning a computer with over half a terabyte of RAM to hold the database and ensure the process ran smoothly and swiftly. The investment was well worth it as it maximized our chances of catching meaningful matches from our samples.

We visualized the results from Kraken2 using Krona (KRONA), creating an interactive chart that laid out the taxonomic classification. This enabled not only our analysis but future review by others.

This step informed us we were looking at metagenomic data, and helped tailor our choice of tools for the steps that followed, steering the analysis in the right direction. This helped us get a broad sense of the landscape and prepare for the deeper analysis ahead.

Alignment and assembly

The next goal was to analyze the DNA sequences from our ancient samples in two different ways: alignment and assembly.

Alignment

Alignment refers to matching the DNA sequences from our samples to a known reference, in this case, the human genome (hg38). We used Bowtie2 (BOWTIE2) for this task, a tool known for its speed and efficiency, making it a good fit for handling our data.

Since NCBI's classification hinted at a significant presence of human DNA in our samples, we chose to align to the human genome. This would help us explore the human DNA aspect and to evaluate claims of genetic engineering associated with these samples.

Assembly

Assembly refers to piecing together the DNA sequences to form longer, more complete sequences, without using a reference. We used two tools for this: SPAdes (SPADES) and MEGAHIT (MEGAHIT). Both are well-regarded in the community for their accuracy and speed. Running both allowed us to cross-verify the results, ensuring a more reliable analysis. MEGAHIT was set to handle large metagenomic data, and SPAdes was run as part of a pipeline called

shovill that also further cleaned the reads. This way, we could get a clearer picture of the DNA, especially the parts that didn't align to the human genome.

This repo was forked from the original shovill repo to add functionality that let us select only specific parts of the pipeline. We ran into resource issues without doing this.

We used these tools with standard settings, making only minor adjustments to Bowtie2 to use a --local setting for a more flexible alignment, and --presets meta-large on MEGAHIT to specify a high-diversity metagenome . This approach kept the process straightforward and effective, setting a solid foundation for the deeper analysis that followed.

Binning

Binning is a way to group together DNA sequences that likely come from the same organism. This step helped us organize our data into more manageable chunks, making it easier to handle, and enabling possible insights into what organisms might be present in our samples.

Workflow was as follows. We started with the assembled contigs. (These are longer stretches of DNA that we pieced together in the previous assembly step.) Then, we attempted to segregate them into bins, each representing a potential different organism.

We used a script to automate the binning process, which did three things:

1. Subsetting contigs: Before diving into binning, we took a random subset of our contigs using a tool called Seqtk. This step made the process a bit more manageable.

2. Binning with MetaBAT2: We then used a tool called MetaBAT2 (METABAT2) to perform the actual binning. This tool grouped our contigs into bins based on certain characteristics they share.

3. Quality checking with CheckM: Lastly, we ran a tool called CheckM (CHECKM) on our bins to assess their quality. This step ensures that the bins we've created are reliable for further analysis.

We ran this script twice, once for each set of contigs we obtained from the two different assembly tools, SPAdes and MEGAHIT. This allowed us to compare and replicate results.

Ricardo Rangel-Martínez

The modified shovill script was flexible, allowing us to skip any of the steps if needed, thanks to the flags --no-subset, --no-maxbin, and --no-checkm. This enabled us to tailor the workflow to our needs.

Haplotyping

Haplotypes are a group of DNA variants along a single chromosome that are likely to be inherited together. Haplotyping—i.e., looking for these groups—can tell us about the genetic differences within a population, as well as the potential lineage of a sample's origins.

Here, we examined the Ancient0003 sample, the most likely candidate to yield insights given its strong alignment to the human genome, with potential richness of mitochondrial DNA (mtDNA). We hypothesized that if this sample is indeed ancient and was found in a cave in Peru, its mtDNA should reveal a lineage predating European colonization in the Americas.

Workflow

We initiated the process with variant calling to identify variations in DNA sequences. This was done using tools like samtools and bcftools. We needed to ensure the quality of these variants, so we calculated various statistics like mean and standard deviation of quality scores.

The next step was to filter these variants for quality, focusing on those with a quality score above the median score of 69, to narrow down to more reliable data. Following this, we filtered down further to only include SNPs (single nucleotide polymorphisms), the most common type of genetic variation among people.

We then moved on to phasing with whatshap. Phasing helped us determine which variants are associated with each other, a crucial step for understanding the haplotypes.

Finally, we focus on the mitochondrial variants using bcftools, and ran haplogrep to classify the haplogroups of the mtDNA against the MitoImpute panel. This would help us understand the lineage of our ancient sample.

**Results**

Overall, our analysis showed samples consistent with degraded ancient DNA. The population of each sample was consistent with the metagenomic profile one

would expect from a sample of tissue extracted from an ancient mummy. The classified DNA reads were largely identified as bacterial, with some presence of human-aligned DNA; in particular, Ancient0003 showed a very strong alignment with the human genome, leading us to investigate ancestral roots of its mitochondrial DNA.

## DNA composition and classification

Our quality control analysis showed high duplication levels, consistent with the DNA degradation found in ancient DNA samples. Our classification suggested a vibrant and rich community of microbial life that had made these ancient remains their home. Between 49-52% of the raw reads in each sample were unclassified, not unusual in ancient DNA samples. Our kraken2 analysis, visualized in Krona and available on our site, showed that the classified reads were dominated by bacterial DNA (Ancient0002, Ancient0003, Ancient0004). This is consistent with what might be expected from ancient DNA samples that have not been cleaned or processed, and does not indicate anything unusual.

## Mitochondrial DNA Analysis

Ancient0003, the sample with >95% alignment to the human genome, yielded a particularly interesting find. Using a MitoImpute reference panel constructed from 36,960 mitochondrial genomes, we identified a basic membership in haplogroup "M," a very broad maternal haplogroup commonly found across East and Southeast Asia. In that panel, haplogroup "M" is not mapped at a high resolution, so we investigated further and found a panel that identified a proposed haplogroup, "M20a", identified in a multiethnic population of 327 citizens of Myanmar located in Northern Thailand. All of Ancient0003's mutations were identified in samples MMR137 and MMR317 (raw data archived here), strongly matching proposed **haplogroup M20a** (Summerer et al., 2014). This was a compelling deviation from the expected result, as it hinted at a genetic thread that spanned across continents from South America to a very specific population in southeast Asia. There were no identified subclades that could link the mtDNA to the Americas, adding a layer of intrigue to the human lineage narrative.

## Assembly and alignment

The samples all showed some alignment with the human genome, with the two samples from "Victoria" showing less (10.88% of the deduplicated Ancient0002

reads aligned to the human genome, and 12.02% of the Ancient0004 reads aligned), while Ancient0003, from an unspecified mummy, showed 95.69% alignment to the human genome.

With the remaining unaligned reads, we conducted de novo assemblies using both Spades and Megahit. After binning both assemblies from each sample with MetaBAT2 (Ancient0002 megahit and spades; Ancient0003 megahit and spades; Ancient0004 megahit and spades), we found generally consistent binnings with lineages, though the percentages of the lineages differed between the SPAdes and Megahit binnings. In general, the Megahit assembly showed lower contamination, while SPAdes showed higher completeness.

These assemblies and alignments are now neatly packaged, ready to be delved into for anyone interested in extending this genomic narrative.

For all three samples, the contigs we acquired post-assembly were primarily classified (binned) as bacteria and archaea, aligning with the overarching microbial theme of our findings.

**Discussion**

Our initial investigation began with a simple premise: to sift through the DNA and see what the data suggest.

Along with varying alignments (ranging from 10% to 95%) with the human genome, we discovered a wide range of ancient microbial life and an unexpected genetic linkage to Myanmar, specifically through the mitochondrial **DNA haplogroup "M20a"** in the sample Ancient0003. The lack of American subclades and the Asian lineage proposes new questions about the origins of the human DNA in that sample. Was it the most likely explanation–a simple mis-handling of the samples along the chain of custody–or does it enrich the story of human migration? How did this DNA thread from Myanmar weave its way into a Peruvian mummy? This deviation from the anticipated American lineage invites a wealth of questions regarding human migration and potential intercontinental connections in antiquity.

The unclassified reads, contigs, and assemblies suggest several avenues for further analysis—BLASTing samples of the contigs, re-running the MEME motifs to exclude repeats, delving into anomaly detection on the alignments and assemblies, and so on.

Importantly, our inability to find evidence of extraterrestrial origin does not discredit the terrestrial nature of these mummies. DNA chemistry as we know it evolved on Earth, tailored by Earth's unique conditions. There's no given reason to believe that extraterrestrial life, if it exists, would share the same molecular structures. For instance, if we were to analyze a soft robot, we wouldn't find alien DNA. We might instead come across various microbial populations residing on or within it. Hence, the findings from this investigation hold significance only if the premise is that these mummies should possess recognizable DNA.

## Conclusion and future questions

The lack of evidence supporting extraterrestrial origin in no way undermines the terrestrial nature of these mummies. It merely underscores the limitations of our current understanding and the specific molecular framework we operate within.

It is also worth mentioning that these DNA samples pre-date the mummies that have so captivated researchers and enthusiasts since the Mexican hearing; we do not have samples for "Josephina" or "Maria"; and we do not know the attributed origin of the Ancient0003 sample. Our pipelines are reproducible, and we are eager to conduct further analysis on additional sample runs from the more recently reported mummies.

The results showing a Myanmar-related lineage on Ancient0003 rather than an indigenous Peruvian lineage are perplexing, and stand out as subjects for future investigation. If we were to take further steps in this direction, we would compare mtDNA and Y-chromosome haplotypes across all three samples as well as against sets of other ancient DNA samples, especially those from the Americas.

The assemblies and alignments of all samples, along with the unclassified reads, contigs, and potential for further detailed analysis, present a rich ground for continued exploration. Anyone interested can take a closer look, run BLAST on the contigs, and perhaps unveil further insights that we might have missed.

The genomic data we've unearthed serves as a stepping stone, inviting curious minds to extend the narrative, to probe deeper into the genomic essence of these

ancient remains. Our findings have broadened the discourse, offering a solid foundation for future investigations.

**Supplemental materials**

Project home page with reports: https://verbalcant.github.io/nazca_mummies.html

Project GitHub repo: https://github.com/VerbalCant/peru_mummy_pipeline

Project data files: https://usegalaxy.org/u/verbal_cant/h/perumummyphase1

References

(MUMMYSTUDIES) Lombardi G, Arriaza B. South American Mummies. In: Shin DH, Bianucci R, editors. The Handbook of Mummy Studies: New Frontiers in Scientific and Cultural Perspectives. Singapore: Springer Singapore; 2020. p. 1–14.

(MMRPAPER) Summerer M, Horst J, Erhart G, Weißensteiner H, Schönherr S, Pacher D, Forer L, Horst D, Manhart A, Horst B, Sanguansermsri T, Kloss-Brandstätter A. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. BMC Evol Biol. 2014 Jan 28;14:17. doi: 10.1186/1471-2148-14-17. PMID: 24467713; PMCID: PMC3913319.

(KRAKEN2) Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol 20, 257 (2019). https://doi.org/10.1186/s13059-019-1891-0

(KRONA) Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 2011 Sep 30;12:385. doi: 10.1186/1471-2105-12-385. PMID: 21961884; PMCID: PMC3190407.

(BOWTIE2) Langmead, B., Trapnell, C., Pop, M. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.Genome Biol 10, R25 (2009). https://doi.org/10.1186/gb-2009-10-3-r25

(SPADES) Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes de novo assembler. Current Protocols in Bioinformatics, 70, e102. doi: 10.1002/cpbi.102

(MEGAHIT) Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674-6. doi: 10.1093/bioinformatics/btv033. Epub 2015 Jan 20. PMID: 25609793.

(METABAT2) Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019 Jul 26;7:e7359. doi: 10.7717/peerj.7359. PMID: 31388474; PMCID: PMC6662567.

(CHECKM) Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015 Jul;25(7):1043-55. doi: 10.1101/gr.186072.114. Epub 2015 May 14. PMID: 25977477; PMCID: PMC4484387.

Addendum: Report updated as of August 6, 2024.

Recently, using the same methodology described for the identification of the mitochondrial haplogroup, the Y chromosome haplogroup was identified in the mummy named Maria. This haplogroup coincides with that of the Han population of China: Y chromosome haplogroup O2a1c1a6a2. This reinforces the origin of the human involved in this hybridization process, which comes from southeastern China and/or northern Myanmar.



Han-Chinese Phenotype

Mitocondrial haplogroup, "M20a" in Maria

Y chromosome haplogroup O2a1c1a6a2

Myanmar Phenotype

End of document

Ricardo Rangel-Martínez