Sequencing Tells a New Story About Native American Origins

Long ago archaeological find yields clues to an Old World connection that predates European explorers' voyages to the New World.

Introduction

About 70 years ago, an archaeological dig on the banks of Lake Baikal at the Mal'ta site in south-central Siberia yielded the skeletal remains of a young boy buried with various figurines and jewelry. Radiocarbon dating estimates the age of the bones to be about 24,000 years old, but conflicting hypotheses exist as to the ancestry of the remains. Researchers have analyzed the tooth and cranial morphology, with some suggesting that the boy's features are East Asian–like, while others believe that they are more European in origin.

In 2009, a team of international researchers led by Eske Willerslev, Ph.D., Director of the Centre of Excellence in GeoGenetics in the Natural History Museum of Denmark (Copenhagen), traveled to the Hermitage Museum in St. Petersburg, Russia with a request to sequence the remains of the Mal'ta boy. They were looking for clues about Native American ancestry. Using Illumina HiSeq® and MiSeq® systems, they identified its ancestry as being closest to modern-day Western Eurasians. Surprisingly, they also found that 14–38% of the Mal'ta boy's genome is shared with present-day Native Americans.

Until recently, most researchers agreed that the founding population of Native Americans was derived solely from ancestors of modern-day East Asians. Dr. Willerslev and the team's sequencing results replaced the old theory with a startling new one. In addition to representing a population that lived during the Last Glacial Maximum (LGM)*, the Mal'ta boy genome provides evidence that two peoples contributed to the founder population of Native Americans[1]—one with East Asian ancestry and one with Eurasian roots.

Maanasa Raghavan, Ph.D. is a post-doctoral researcher in Dr. Willerslev's laboratory at the Centre of Excellence in GeoGenetics at the University of Copenhagen.

iCommunity spoke with Maanasa Raghavan, Ph.D., a postdoctoral researcher in Dr. Willerslev's laboratory at the Centre of Excellence in GeoGenetics at the University of Copenhagen, about the research and how it offers proof of a more complex picture of the west-to-east human migration into North America across Beringia, the Ice Age land mass that lies beneath today's Bering Strait.

Q: Why are the Mal'ta boy remains important in your research of Native American ancestry?

Maanasa Raghavan (MR): The Mal'ta sample is quite famous in archaeological literature. The site itself is unique due to the presence of anthropomorphic Venus figurines, which are characteristic of Upper Paleolithic Europe but are rare in Siberian archaeology. Since it was unearthed more than 70 years ago, the Mal'ta boy has been radiocarbon dated and analyzed in various ways, primarily morphologically, but never sequenced. The origin of the boy remained contentious, with claims of both European and East Asian morphological characteristics. For our studies, the Mal'ta specimen offered a window into what the genetic structure was around 24,000 years ago in southern Siberia, an area that genetic literature suggests as a potential homeland of the first Americans.

Q: How did you gain access to the Mal'ta skeletal remains at the Hermitage Museum?

MR: Our collaborators from the United States, mainly Dr. Kelly Graf at the Center for the Study of First Americans and the Department of Anthropology at Texas A&M University, helped us gain access to the remains. In 2009, she and Dr. Willerslev planned a trip to St. Petersburg where Dr. Graf had contacts at the Hermitage Museum. They inquired about obtaining a bone sample from the Mal'ta specimen (MA-1) for sequencing and their request was approved.

Drs. Kelly and Willerslev picked a humerus bone and drilled out a smaller subsample. We had to make the most of the sample, since in addition to extracting and sequencing DNA from the bone we were also interested in obtaining a radiocarbon date from it.

Q: Was the DNA extraction performed at the Hermitage Museum or in your laboratory?

MR: Ancient DNA is prone to damage and degradation over time, and occurs in much lower copy numbers than DNA from fresh tissues. Therefore, ancient DNA research requires a sterile environment to control against contamination from exogenous sources. We have a dedicated ancient DNA facility at the Centre for GeoGenetics, which is where all the laboratory work was performed.

*LGM occurred between 25,000 and 19,000 years ago

Q: Radiocarbon analysis dated the sample at 24,000 years old, which is just before the LGM. In the area of Siberia where the sample was found, what was the environment like then?

MR: Glaciated areas were spreading southward during that time, so it would have been pretty cold in Siberia. Mal'ta and other contemporaneous settlements in Siberia thrived during pre-LGM times, despite the onset of environmentally stressful conditions that ultimately led to the depeopling of the area during the LGM.

Q: When you first sequenced the Mal'ta sample what did you find?

MR: Initially, we decided to type the control region of the mitochondrial genome and obtained a typical European mitochondrial haplogroup U signature. Our first thoughts were that it had to be due to contamination. We decided to move forward anyway and see what the nuclear genome signal looked like. We had just received the HiSeq 2000 system, so we decided to use that to perform shotgun sequencing on the sample.

Q: Was the Illumina sequencing performed in your laboratory?

MR: The Danish National High Throughput Sequencing Centre is associated with the Centre of Excellence in GeoGenetics and it acts as our in-house facility. We performed the DNA extraction and library preparation at our ancient-DNA laboratory and then walked down the street to the sequencing center to get it sequenced.

Q: What were the results when you sequenced the MA-1 sample on the HiSeq 2000 system?

MR: Next-generation sequencing (NGS) provided us with orders of magnitude more sequences of the information-rich nuclear genome. It enabled us to observe on a much larger scale that the sample carried post-mortem damage, yielding a very typical ancient DNA signature. That told us there was definitely an endogenous signal and the European-like signature wasn't just contamination. We took it from there and generated more sequences and started performing downstream population genomics analyses.

Q: How does the MA-1 genomic signature compare with those of worldwide populations?

MR: We found through principal component analysis that MA-1 was intermediate between modern Native Americans and Western Eurasians. This was an intriguing signature and we decided to explore it further by performing model-based population genetics analyses. We used SNP array data and complete genomes from several worldwide modern-day populations, including Native Americans and Siberians. We also sequenced four new genomes from Eurasia with Mari, Avar, Indian, and Tajik ancestry. Further analyses demonstrated that even though MA-1 showed a genetic affinity to Native Americans, there was actually nothing East Asian about the signature. If MA-1 was a Native American or had received gene flow from Native Americans, then it would be expected to show some affinity to East Asians since Native Americans derive from ancestors of present-day East Asians. However, this was not the case. All our analyses indicated that MA-1 was of Western Eurasian ancestry and was either a part of or related to a population that had contributed genes to ancestral Native Americans.

Q: You also compared the genome of the Mal'ta boy to sequences derived from the remains of an ancient East Asian individual. What did the data show?

MR: The 40,000 year old individual from Tianyuan Cave in China has been found to be ancestral to modern-day Asians and Native Americans. We wanted to test if the greater genetic affinity of Western Eurasians to Native Americans over East Asians might be due to events in the recent history of East Asians. We used the available sequence data from chromosome 21 of the Tianyuan individual in lieu of present-day East Asians and found this was not the case. The Tianyuan individual and modern East Asians behaved similarly in these tests.

"Next-generation sequencing (NGS) provided us with orders of magnitude more sequences of the information-rich nuclear genome."

Q: You compared MA-1 with several Native American gene signatures. What did those results show?

MR: We compared MA-1 to published data sets, including the genome of a Karitiana individual from Brazil and SNP panels from several Native American and Eskimo-Aleut populations. We employed this complementary approach because the genome data set overcame inherent genotype data biases, while the genotype data provided us with a larger panel of New World populations. The other important consideration was that the SNP data had been masked for recent European admixture to avoid detecting a post-Columbian** European signal. Whether we employed the masked SNP data or the Karitiana genome, we found evidence of gene flow between Native Americans and the MA-1 lineage.

Q: How does MA-1 compare with modern Siberian gene signatures?

MR: Several Northwestern Siberian and Northeastern European populations showed slightly higher affinity to MA-1 than some of the central and south Siberian populations did. That makes sense because the more southern Siberian groups have received recent gene flow from East Asians. It goes to show that the population structure that is evident today is not necessarily reflective of the scenario in the past.

Q: Why did you compare MA-1 with the gene signature of younger ancient remains found at Afontova Gora-2 (AG-2), an archaeological site about 600 miles (965 km) from the Mal'ta site?

MR: The two specimens bridged the transition into and out of the LGM. We wanted to see if the gene signature in central Siberia changed after the LGM. The younger remains were radiocarbon dated to about 17,000 years before present. Although the sample was more heavily contaminated than MA-1, using the HiSeq 2000

**Post-Columbian = after 1492 CE

system and downstream bioinformatics tools we managed to isolate an endogenous signal, and that was exciting. Principal component analysis showed that AG-2 has a similar genetic signature to MA-1.

Archaeologically speaking, throughout the LGM we don't have much evidence of human activity in this region. The hypothesis is that entire areas of Northeast Asia and what used to be Beringia were depopulated during this time. It was too cold. Our data demonstrate that there was genetic continuity, with a similar genetic composition reentering the area after the whole region was depopulated. This obviously has a strong bearing on the first American gene pool that is proposed to have made its way into the New World from Beringia after the LGM, some 15,000 years ago.

Q: Based on the sequencing data and post-analyses, what did your study demonstrate about the human migration into North America?

MR: It's theorized that the entry of the first Americans into the New World was around 15,000 years ago. Our data shows similar signatures in the area south of Beringia after the LGM as we found in the Mal'ta boy. That Western Eurasian signature is detected in modern Native Americans as a result of the coming together and admixing of populations related to MA-1 and modern East Asians. I don't think a lot of people had expected an additional component to the first American gene pool. That's what made this study really interesting.

Q: Have any skeletal remains been found in Alaska or Canada that might represent the next part of the migration?

MR: Obviously, that's what a lot of people are focusing on. The problem is that much of what was once Beringia is under water now and we've lost whatever sites there might have been. At the moment, no ancient remains dating to immediately post-LGM have been found in far-east Siberia, Alaska, or northwestern Canada. That doesn't mean that they don't exist. We just haven't found them yet or perhaps there are archaeological samples lying in the basement of a museum that are yet to be rediscovered. That's what made the Mal'ta boy remains so important. They were well-documented and prominently featured in the literature.

> "It was incredible to get the HiSeq 2000 System and obtain data so quickly."

Q: Your data disproved the Solutrean hypothesis that the ancestors of Native Americans traveled in boats across the North Atlantic from Iberia to the northern shores of North America.

MR: The presence of a Western Eurasian-specific mitochondrial signature (haplogroup X) in modern Native Americans as well as discrepancies in the cranial morphologies of several first Americans have resulted in hypotheses of alternate, non-East Asian origins of Native Americans. For instance, the Kennewick man[2] remains, which are estimated to be between 9,300 and 9,600 years old, have long been labeled as a potential Caucasoid, or European. Our study presents an alternate route for Western Eurasian signatures in modern Native Americans than the Solutrean theory. Migration via Beringia into the New World is a more parsimonious explanation for the aforementioned signatures than rafting all the way from Iberia.

Q: How did NGS impact this project?

MR: NGS definitely enhanced the speed of the project. We really gathered momentum after we moved to NGS. It was incredible to get the HiSeq 2000 System and obtain data so quickly, both from modern and ancient sources. This was also aided by the fact that the biomolecular preservation of the Mal'ta sub-sample we analyzed was quite good and amenable to high-throughput sequencing. The challenge went from how to obtain genomic data to how to analyze it all.

> "The HiSeq System yielded high-quality sequences from both the ancient and modern samples."

Q: You also used the MiSeq and HiSeq 2500 systems to sequence the modern genomes. How did you split the sequencing tasks between the two systems?

MR: The MiSeq System was a very nice screening alternative for both the ancient and modern libraries, enabling us to perform pilot runs because of its smaller scale and faster turnaround time. We used the MiSeq System to evaluate the quality and overall characteristics of the libraries, and subsequently sequenced them on the HiSeq to generate the data we needed.

Q: What was the quality of the sequencing results for the ancient DNA?

MR: The HiSeq System yielded high-quality sequences from both the ancient and modern samples. The error rates for the ancient libraries were less than 0.3%, which is comparable to other ancient genomic data sets out there and enabled us to use the data sets for further analysis. The errors were primarily transitions deriving from deamination of cytosine (cytosine to thymine and guanine to adenine), which is the dominant source of ancient DNA damage.

Q: What's the next step in your research?

MR: In our study, we only sequenced MA-1 to 1× depth. It would be interesting to sequence MA-1 at a higher depth to increase the power to perform more in-depth analyses. Unfortunately, we're out of sample and sequencing libraries. Hopefully, the Hermitage Museum is happy with the results of our paper and will allow us to resample the specimen.

We also have the opportunity to move on and build upon this work. We've clarified the origins of the first Americans and would like to explore, using complete genomes, what happened after the first gene pools entered the New World and learn how the genetic diversification process occurred.

References

1. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505: 87–91.
2. Kennewick Man—The Ancient One, Burke Museum website, www.burkemuseum.org/kennewickman

Illumina, Inc. • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2014 Illumina, Inc. All rights reserved.
Illumina, HiSeq, MiSeq, and the pumpkin orange color are trademarks or registered trademarks of Illumina, Inc. in the U.S. and/or other countries. All other brands and names contained herein are the property of their respective owners.
Pub. No. 770-2014-005 Current as of 17 March 2014

illumina®