Ancient DNA Speaks Volumes About New World Populations and Migrations

What can an ancient human genome tell you? Quite a lot when you leverage the speed and sensitivity of Illumina's Genome Analyzer and the power of Infinium® BeadChips.

Introduction

"I lived several thousand years ago in what is now Greenland. I had thick black hair with a tendency towards baldness, brown eyes, brown skin, type A+ blood, shovel-formed front teeth, dry ear wax, a stocky build adapted to living in a cold climate, and I am more closely related to people living in Siberia than I am to anyone now living in North America or Greenland." ~ 4,000 year old Saqqaq individual

This was more information than Eske Willerslev, Ph.D. and Morten Rasmussen[1] expected to uncover when they decided to sequence the DNA of a ~ 4,000 year old tuft of permafrost-preserved human hair. Their goal at the onset was to sequence the first complete ancient human nuclear genome and in the process, gain information about one of the first cultures to settle in the New World.

The hair sample had been at Denmark's National Museum since it was discovered in 1986 at a site on the southwestern slope of Greenland. The location and radiocarbon dating placed the individual in the Saqqaq culture which populated Greenland from 2500 BCE until about 800 BCE[2].

The Benefits of Speed and Quality

Dr. Willerslev and Mr. Rasmussen knew from sequencing the mitochondrial DNA[3] that the specimen was of reasonably good quality. While in ancient DNA terms, the sample looked promising, there would still be limitations and obstacles to overcome in sequencing a complete nuclear genome. "We were going from 16,000 base pairs for a mitochondrial genome to around 3 billion base pairs for a nuclear genome, said Dr. Willerslev. "We needed a system that delivered high output, with high quality. Once we verified that we could obtain a lot of useful sequence from a single run, it was clear the Illumina Genome Analyzer was the best choice. It made a resequencing project like this feasible."

It turned out that speed and output would be important for another reason. A team of German scientists was attempting to sequence another hominid ancient genome—that of a Neandertal[4]. To increase the chances of completing their research first, Dr. Willerslev approached the Beijing Genomics Institute (BGI) to enlist its help. BGI is one of the world's leading scientific organizations, operating seven genome research centers in mainland China for scientific collaboration and sequencing services. With help from BGI, the race was on to determine which team would be the first to sequence and publish an ancient human genome.

Accounting for Contamination and Age

With the sequencer and sequencing sites selected, the team turned to solving a critical issue that could impact the quality of the sequence—how to deal with the problem of contamination, both microbial and human. With keratins offering a protective coat around the internal DNA, any microbial contamination on the surface could be easily removed by bleaching and washing the hair before treatment with proteinase K.

"We were more concerned about human contamination, because we needed to know that we were actually sequencing an ancient human, not 90% modern contamination," said Mr. Rasmussen. "One of the beneficial aspects of Illumina's technology is its use of indexing adaptors and primers for library preparation." After the DNA extraction and library builds were performed in a clean laboratory in Copenhagen, the libraries could be indexed enabling the team to know exactly which sequences left the laboratory in Copenhagen. This made it easier to identify any contamination that might have entered the samples during analysis.

Dr. Eske Willerslev is leader of the Center of Excellence in GeoGenetics at the Natural History Museum of the University of Copenhagen, Morten Rasmussen is a Ph.D. candidate at the Center.

To account for DNA fragmentation, the DNA sample-to-adaptor volumes were adjusted for short insert DNA to approximate the 1:10 ratio suggested in Illumina's protocol, and the library protocol was modified to move the gel cutting step after library amplification to ensure the library would not be lost during purification. The amplification step was also adjusted to avoid misreads caused by deamination of cytosine, which accounts for 95% of ancient DNA damage and makes it difficult to distinguish between evolutionary derived substitutions and those caused by DNA damage. "As recommended in Illumina's protocols,

we used a Phusion™ polymerase[5] that does not amplify sequences that are damaged and successfully tested this approach on a number of ancient animal remains," said Dr. Willerslev. "As a result, we were able to almost completely avoid miscoding lesions in the data set."

From start to finish, the sequencing of the ancient Saqqaq genome took about 45 days. "At the beginning, we only obtained 10X coverage," added Dr. Willerslev. "Although this is way better than genome sequencing from any ancient material so far, we wanted an ancient genome comparable in quality to that from a modern specimen and continued sequencing until we had obtained 20X coverage. The only difference between this ancient human genome and that of modern humans is that we could only uniquely map some 80% of the genome because of fragmentation that caused the average read length to be about 55 base pairs."

After genotyping was conducted, the team verified that what had been sequenced was indeed the ancient DNA. "While we had put procedures in place to avoid contaminating the DNA during analysis, we did not know if the same level of care had been followed by the team that excavated the sample in 1986," added Mr. Rasmussen. "However, we did know that all of the people handling it at that time were Europeans. Given that the mitochondrial DNA sequence demonstrated that the person was of Asian origin, we looked for private alleles of European origin in the nuclear genome sequence and found the level was a maximum of 0.8% of the raw sequence data. That level of possible modern contamination had no influence on the final results and interpretations."

Genotyping Elucidates an Unknown Migration

From the nuclear sequence, a high-confidence subset of 353,151 SNPs was defined and the team proceeded to conduct genotyping studies to determine the population genetics context of the Saqqaq individual. The first step was to compare the genome with those of populations that could possibly be descendents. Illumina Infinium Human660 and Human610 BeadChips provided comprehensive global coverage of modern genetic variation. To add depth within the Arctic and Native American populations that were potentially descendents of the Saqqaq individual, the team obtained fluid samples of 16 populations, including four native North American and twelve north Asian (Siberians and Inuits). These sample populations and that of the Saqqaq individual were run on the Infinium BeadChips, generating 197 new genome scans. Altogether, 95,502 SNPs were found in common.

The results of the SNP analysis were startling. It was long thought that inhabitants of Greenland were direct ancestors to present-day Inuit (Eskimos) of Alaska, Canada, and Greenland, or to the Native Americans that came in one or two migration waves completed around 13,000 years ago. However, this person shared very few markers in common with these population groups. Instead, his genome was more closely linked with native populations in Siberia.

"We used a population genetics model to obtain the maximum likelihood estimates of the divergence times between the Saqqaq individual and the reference populations," said Dr. Willerslev. "We found that the ancestral Saqqaq separated from their Old World relatives 5,500 years ago, about the time where you find the earliest archaeological evidence of humans in the New World Arctic. It means that the ancient genome represents a human migrating into the New World from Siberia that is independent of the migrations giving rise to present-day Native Americans and Inuit. At some point, the entire population that our Saqqaq individual represents became extinct, leaving no descendents in the New World today."

> "Once we verified that we could obtain a lot of useful sequence from a single run, it was clear the Genome Analyzer was the best choice. It made a resequencing project like this feasible."

"This unknown migration was the biggest surprise of our research," Dr. Willerslev added. "It means there is a great possibility that there have been several migrations coming into the Americas, with maybe only one of them leaving descendents in the New World. I think it will make people rethink their views about migration, specifically using the presence of ancient SNPs in modern populations as the only validation that the migrations occurred."

Key Issues in Sequencing Ancient Genome Samples

Not all ancient genome samples have been found in permafrost and are so well preserved, but that should not impact their ability to be sequenced. According to Dr. Willerslev, it is not the amount or condition of the ancient DNA that is the most critical factor, but the relative frequency between contaminant DNA and endogenous ancient DNA.

"The Illumina technology is so sensitive, that you can detect tiny amounts of DNA," Dr. Willerslev said. "It is therefore unnecessary to limit yourself to bone or tooth specimens where you find the most DNA per gram. Our hair sample contained much less DNA than a bone or a tooth, yet it gave us enough to complete a nuclear sequence. The key is to efficiently reduce the amount of contaminant DNA. We are confident that by using Illumina technology we can retrieve and sequence even very small amounts of endogenous DNA."

Dr. Willerslev and Mr. Rasmussen are moving forward with their research, looking at ancient human samples from outside the permafrost regions of the Americas and Europe for clues of unknown migrations to the New World. And yes, their team was the first to sequence and publish a complete ancient human genome.

References

1. Rasmussen M, Li Y, Lindgreen S, Pedersen J, Albrechtsen et al. "Ancient human genome sequence of an extinct Palaeo-Eskimo." Nature 463 (7282): 457–462.
2. Saqqaq culture profile, http://www.natmus.dk/sw18632.asp
3. Gilbert MT, Kivisild T, Grønnow B, Andersen PK, Metspalu E, et al. "Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland." Science 320 (5884): 1787–1789.
4. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U et al. "A draft sequence of the Neandertal Genome," Science 328 (5979): 710–722.
5. Phusion polymerase, Finnzymes, Espoo, Finland.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2010 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, and HiScan are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
*Phusion is a trademark of Finnzymes Oy.
Pub. No. 070-2010-015 Current as of 07 June 2010

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA