

# Sequencing Uncovers a 9,000 Mile Walkabout

A lock of hair and the HiSeq<sup>®</sup> 2000 system identify a human migration wave that took more than 3,000 generations and 10,000 years to complete.

## Introduction

Archaeological evidence dates the Aboriginal presence in Australia to ~50,000 years before the present (BP), making them one of the earliest known populations of modern humans outside of Africa. Recognized as Australia's founding population, scientists theorized that the ancestors of today's Aborigines arrived on the continent from a single-wave migration out of Africa into Europe and Asia. However, recent whole-genome studies date the Europe/Asia split to have occurred between 17,000 and 43,000 BP, more than 10,000 years after the earliest Aborigine archeological evidence. Could the founding Australian Aborigine population be the result of an earlier migration wave? The answer was foundi by sequencing a centuryold lock of hair with the HiSeq 2000 system<sup>1</sup>.

# Long Forgotten Sample Proves Valuable

Morten Rasmussen, Ph.D., postdoctoral fellow in Dr. Eske Willerslev's lab at the University of Copenhagen, stumbled upon the ancient hair sample during a visit to the University of Cambridge. The team had recently experienced success in sequencing DNA from the hair of a Saqqaq individual found in the Greenland ice, uncovering an unknown migration of Old World humans to the New World Arctic<sup>2</sup>. During a discussion of that research, the Cambridge scientists mentioned they had additional ancient hair samples in their archives, including several 100-year old Aborigine hair segments. "We were intrigued, since the samples were just old enough that we could assume they were likely from Aborigines of unmixed origin," said Dr. Rasmussen.

Most genome studies in the Aboriginal people have been done using SNP array analysis, which can introduce sampling biases. "The SNP array results showed only a few SNP positions for private



Dr. Morten Rasmussen is a postdoctoral fellow and Dr. Eske Willerslev is leader of the Center of Excellence in GeoGenetics at the Natural History Museum of the University of Copenhagen.

alleles in Aborigines, because they were not sampling a large enough population," said Dr. Rasmussen. "No one had yet taken advantage of whole-genome sequencing to find new variants. We decided to sequence the most suitable samples to see if they could provide data on the ancient origins of these individuals."

Prior to genome sequencing, the team screened four roughly 100-year old Aborigine hair samples for signs of mixture using the HiSeq 2000 system. Comparison with data previously generated on SNP arrays allowed the identification of admixed and non-admixed individuals. One individual with no measurable admixture and high levels of endogenous DNA was selected.

The sample was from a young, male Aborigine living near Kalgoorlie, Western Australia, approximately 500 miles east of Perth. Using 0.6 g of the hair for DNA extraction, the HiSeq 2000 system was used to sequence the genome to an overall depth of 6.4×, with ~60% of the genomic regions sequenced to an average of 11×.

According to Dr. Rasmussen, "HiSeq greatly sped up the project. We used the Genome Analyzer™ for the Saqqaq research, and while I think we generated roughly the same number of reads for the two projects, HiSeq was a lot faster, generating longer reads while still delivering the same high-quality data."

While the 100-year old sample did exhibit a high degree of fragmentation, the levels of cytosine-to-thymine misincorporation typical of ancient DNA were low and restricted to the read termini. These were trimmed to improve SNP call quality and the genome was assembled and genotyped, identifying over 2 million SNPs of which 449,115 were considered high confidence. Of these, 28,395 had not been previously reported.

# **Identifying Ancestry Clues**

To place the Aboriginal genome in the context of worldwide human genome variation, the high-confidence SNPs were compared to Illumina HumanHap650Y and Human660W-Quad BeadChip data from a reference group consisting of 1,220 individuals belonging to 79 populations. In addition to representative samples from people of African, European, and Asian descent, this reference population included 45 previously unpublished individuals from two huntergatherer groups, the Kusunda (Nepal) and Aeta (the Philippines), which were hypothesized to be relic populations from the proposed earlier wave of dispersal across eastern Asia. Principal components analysis results illustrated distinct genetic differentiation among the African, Asian, and Greater Australian populations, with the Aborigine genome clustering most closely with samples from Highland Papua New Guinea (PNG) and the Aeta, identifying the two as relic populations of the migration.

To determine shared ancestry, the team performed a D-test, also called an ABBA/BABA test, on the genotyping data. This comparative

analysis of multiple genomes involved counting the number of times a particular sequence was similar to a derived human allele (A) versus an ancestral chimpanzee allele (B), and enabled scientists to identify gene flow and isolation of different populations. "We found significantly larger proportions of shared derived alleles between the Aboriginal Australian and East Asians (Cambodian, Japanese, Han, and Dai) than between the Aboriginal Australian and Europeans (French)," Dr. Rasmussen said. "We also found a larger proportion of shared derived alleles between the French and Asian populations than between the French and Aboriginal Australians."

"We then analyzed whether the Aboriginal and PNG Highlands genomes were closer to the Native American or Asian genomes," Dr. Rasmussen added. "They appeared equally far apart, indicating that Asians and Native Americans were together equally long and that the Aborigine split off before that period. This suggests that before European contact occurred, Aboriginal Australian and PNG Highlands ancestors had been genetically isolated from other populations since at least 15,000 to 30,000 years BP."

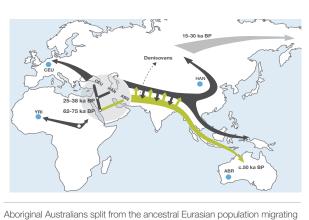
"HiSeq greatly sped up the project...it was a lot faster (than the Genome Analyzer), generating longer reads, while still delivering the same high-quality data."

#### One Migration or Two?

Given that the Aborigine genome was found in one analysis to be genetically closer to East Asians than Europeans, greater granularity was needed in order to determine whether the data supported the single-wave or multiple-wave migration theory. The team sequenced whole genomes of three Han Chinese (ASN), and together with 1000 Genomes data from two Europeans (CEU) and two Yorubans (YRI), used a test comparing patterns of similarity with the Aboriginal (ABR) genome. "The test is more robust than the D-test, pairing the genomes into four sets and comparing the similarities of Group 1 [(YRI, ASN), (CEU, ABR)] with Group 2 [(YRI, ABR), (CEU, ASN)]," stated Dr. Rasmussen. "In the single-wave model, the Aboriginal Australian would have a European allele (Group 1) as often as the Asian individual would (Group 2). Instead, our analyses showed that there was a significantly higher proportion of Group 2 over Group 1 alleles, something that could not occur in a single dispersal model, but is highly indicative of the multiple dispersal model."

A comparison of the Aborigine and African genomes also supported this theory. In contrast to other populations, the African genome has a higher ratio of singletons to heterozygous genotypes, with more lowfrequency SNPs. The team found the same to be true of the Aboriginal genome when compared to other Eurasian populations, suggesting that the Aboriginal Australians split from the ancestral population of Eurasians, rather than from modern East Asian populations.





Aboriginal Australians split from the ancestral Eurasian population migrating out of Africa between 62,000 to 75,000 years BP, more than 30,000 years prior to when European and Asian populations split from each other.

## A Migration of Epic Proportions

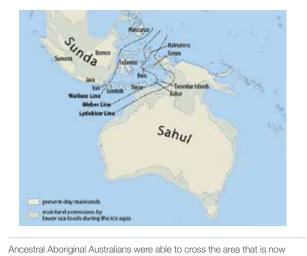
Estimating the times of divergence involved the use of population genetics to analyze the whole-genome data, measuring the rate of mutation, allelic frequency, and linkage disequilibrium. "We estimate that Aboriginal Australians split from the ancestral Eurasian population migrating out of Africa between 62,000 to 75,000 years BP," stated Dr. Rasmussen. "Our data also confirmed previous estimates that the European and Asian populations split from each other about 25,000 to 38,000 years BP." (Figure 1).

While their data provided an approximate time, the exact origin of the earlier migration wave out of Africa remains unclear. "The geography is a bit murky based on genetic data alone, but we have a few hints based on Neanderthal signatures in the Aboriginal genome," Dr. Rasmussen added. "We analyzed allele sharing with Denisovan<sup>3</sup>, an ancient hominin sample found in Siberia's Altai Mountains, and the original Neanderthal genomes."

"The data showed that the Aborigine genome shared the same genomic fragments of the original Neanderthal genome that all other out-of-Africa populations do. In fact, there is more allele sharing between the Aborigine and Denisovan genomes than between the European, Asian, and Denisovan genomes. It's likely that the Aborigine Australians split from Africans at about the same time as the Eurasians, but soon after took off on their own. Interestingly, we see the same Denisovan mixture signal in the PNG Highlands genome, which may mean that they were part of the same migration wave into Asia."

Allele sharing between the Aborigine Australian and PNG Highlands populations isn't too surprising given the close proximity of Australia and New Guinea. They were even closer during the time of the early migration, when the lower sea level merged what is now Australia, Tasmania, and New Guinea into one land mass called Sahul. (Figure 2) The estimated 50-meter drop in sea level also connected the current areas of Thailand, Sumatra, Java, and Borneo into one land mass called Sunda. Islands formed a stepping-stone path to Sahul, enabling the ancestors of the young man whose lock of hair sparked this research to become the first human inhabitants of greater Australia.

Figure 2: Sahul and Sunda Land Masses



Malaysia and reach Australia due to the lower sea levels that existed during the ice ages.<sup>4</sup>

### Sample Benefits Research 100 Years Later

The hair sample obtained by anthropologist Alfred Haddon during a trip to Western Australia in 1921 lay in Cambridge's specimen room for 100 years waiting for science to catch up. "The belief at the time was that you could study anthropologic characteristics through hair morphology, something that's proven difficult to do," said Dr. Rasmussen. "While it didn't benefit his research, that hair sample proved to be valuable almost a century later, enabling us to prove the occurrence of one of the first human migrations out of Africa." While the sequencing for this research project was performed at the Beijing Genomics Institute (BGI), the team's current research into migration pattern in the Americas is being powered by the new HiSeq 2000 system in their lab. "We're looking forward to using it to uncover more such migrations."

#### References

- Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 334: 94-98.
- Rasmussen M, Li Y, Lindgreen S, Pedersen J, Albrechtsen, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463: 457–462.
- Reich D, Green RE, Kircher M, Krause J, Patterson N, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468: 1053–1060.
- 4. Image source: Map of Sunda and Sahul, downloaded from Wikipedia, licensed under http://creativecommons.org/licenses/by-sa/3.0/deed.en

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

#### FOR RESEARCH USE ONLY

#### © 2012 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 070-2012-003 Current as of 11 April 2012

