Investigating DNA of pre-Iron Age hunter-gatherers from southern Africa

Proposal for obtaining a permit to sample pre-Iron Age skeletal material curated in the KwaZulu-Natal museum

The research team

This research will be conducted by an international, collaborative team including:

1. Prof Mattias Jakobsson, Director of the Jacobsson Laboratory at the Evolutionary Biology Centre and in the Eepartment of Evolutionary Biology at Uppsala University, Sweden [http://www.ebc.uu.se/Jakobsson/research/]
2. Drs Carina Schlebusch and Helena Malmström, gentic scientists in the Jacobsson Laboratory
3. Prof Himla Soodyall, genetic scientist at the Human Genomic Diversity and Disease Research Unit, Division of Human Genetics, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand and the National Health Laboratory Service
4. Prof Marlize Lombard, Director of the micro-TrACKS (Tracing Ancient Cognition and Knowledge Systems with micro-methods), Department of Anthropology and Development Studies, University of Johannesburg

Collectively this group of researchers has published on techniques for extracting and analysing ancient DNA, genetic variation of humans in prehistoric times, admixture events, cross diciplinary investigation of the human past in southern Africa. Full CVs are available on request.

Background to pre-colonial history of southern Africa and research aims

San and Khoe communities currently represent remnant groups of a much larger and widely distributed population of hunter-gatherers and pastoralists who had exclusive occupation of southern Africa before the arrival of Bantu-speaking groups since about 2000 years ago and sea-borne immigrants of the last 350 years. Mitochondrial DNA and Y-chromosome studies conducted on contemporary Khoe-San groups revealed that they harbor some of the most divergent lineages found in living peoples throughout the world (Behar et al. 2008; Karafet et al. 2008; Schlebusch 2010; Schlebusch et al. 2013). Recently, we conducted a high coverage autosomal genetic study in contemporary Khoe and San populations and showed that these populations form a common lineage, basal to all other modern human populations (Schlebusch et al. 2012). This basal Khoe-San lineage split from other human populations around 100,000 – 150,000 years BP (Gronau et al. 2011; Veeramah et al. 2011; Schlebusch et al. 2012). Thus genetics confirms a deep split between the ancestors of modern Khoe-San populations and all other humans, which imply isolation of Khoe-San populations from other populations prior to contact with Bantu-speaking people and colonists. It is not known to what extent admixture with Bantu-speaking people and non-Africans influenced the genetic makeup of current San and Khoe groups, in particular, events that occurred during early contact between Bantu-speakers and Khoe-San groups. This is largely due to the lack of completely non-admixed comparative/reference Khoe-San groups and Pickrell et al. (Pickrell et al. 2012) suggested that all current-day Khoe and San people represent groups with some admixture with Bantu speakers. Thus obtaining genetic material from individuals in time periods that precede these admixture events is critical to estimate exact admixture proportions and understanding the history of southern African peoples.

Furthermore, in the high coverage autosomal genetic study (Schlebusch et al. 2012), we found that Khoe-San populations are genetically distinct from each other and a clear geographic structuring among contemporary Khoe-San groups was observed. The northern (Angola and northern Namibia; Ju speakers) and southern (South Africa, Tuu speakers) Khoe-San groups were most distinct from each other with the central Khoe-San groups (Botswana, Khoe speakers) being intermediate. Population divergence within the Khoe-San group was approximately 1/3 as ancient as the divergence of the Khoe-San as a whole to other human populations (thus ~ 25,000-43,000 years BP – on the same order as the time of divergence between West Africans and Eurasians). Our study included one pastoralist Khoe group (KhoeKhoe linguistic division), namely the Nama from Namibia, and we identified a Nilo-Saharan (East African) ancestral component in the group (around 16% of the genomes of Nama individuals were assigned to the East African component), possibly related to the introduction of pastoralism to southern Africa (Schlebusch et al. 2012). However, the rest of the Nama genetic component was similar to other San groups and mostly similar to descendants of the southern San (Tuu speakers), the ≠Khomani and the Karretjie People. It is known that many other Khoe as well as San groups existed historically throughout the Cape region of southern Africa. In the last 400 years, however, it seems that they started to lose their socio-cultural identities as Khoe and San groups, and many of their descendants were integrated in mixed ancestry populations such as the Coloured and Griqua populations of South Africa. By studying ancient DNA of different sites in southern Africa (before the disruptive influences of colonialism) we could investigate if the contemporary structure among Khoe-San and Coloured groups can be linked to specific geographic areas and/or cultural complexes in the past. Thus, testing the high coverage autosomal data that point to genetic distinction and geographic structuring and contributing more accurate knowledge about the region's population history that cannot be obtained in any other way. We will identify similarities/differences between the historic individuals and different current-day Khoe, San and Bantu-speaking groups.

The different contemporary Khoe and San groups also showed diverse histories of gene flow with surrounding populations, and through ancient DNA studies on dated material, hypotheses can be formulated about when this gene flow started. This can be especially useful in studying the two independent waves of pastoralist and farming practices to southern Africa; namely the Khoe lifeway, generally associated with the introduction of pastoralism, and the arrival of Bantu-speaking farmers in southern Africa.

For the KwaZulu-Natal region specifically, changes in the archeological record, morphological changes in skeletons and isotope changes relating to dietary subsistence indicates that this region was colonized by Bantu-speaking, Iron Age farmers around 400 AD (Ribot et al. 2010). Skeletal remains predating this date would thus be representative of indigenous hunter-gatherer San groups from the region. Since today San groups have completely disappeared from this region, and form most of the eastern parts of southern Africa, it would be invaluable to have genetic information of these eastern San groups to compare to genetic information from contemporary western San groups.

Background to new approaches to ancient DNA

Extracting DNA from ancient human remains or bone material was pioneered almost three decades ago, but the field of ancient DNA (aDNA) has been plagued with problems for many years, such as contamination from modern DNA, damaged DNA, and low levels of endogenous

DNA (Poinar et al. 2006). However, one problem after the other have been solved and in the last few years and the progress has resulted in a revolution of the aDNA field, most notably by the sequencing of the Neandertal genome (Green et al. 2010), the Denisova genome (Reich et al. 2010), sequencing of the genome of a prehistoric human from Greenland (Rasmussen et al. 2010) and sequencing of multiple Neolithic individuals from Scandinavia (Skoglund et al. 2012). Older material from animals has also yielded authentic ancient DNA, sometimes as old as 400,000 years (from sediments, (Willerslev et al. 2007); from cave bears, (Valdiosera et al. 2006); and from woolly mammoth, (Gilbert et al. 2008)). Working with human remains is particularly sensitive to contamination from modern humans. We have addressed these issues together with long-term collaborator Anders Götherström using molecular techniques (Malmstrom et al. 2007), and the Jakobsson laboratory has recently developed bioinformatic techniques to authenticate ancient DNA (Skoglund et al. 2012). Genetic analyses of ancient DNA has opened up the potential to answer questions about populations from the past such as their history, and relationship to living populations.

Project proposal

We propose to extract DNA from human remains (in particular bone and teeth) curated by the KwaZulu-Natal museum, originating from various sites in the KwaZulu-Natal Province. We have already viewed the material on a previous visit to the museum (on 16 May 2013) and found numerous specimens that could yield DNA (summarized below). To minimize impact on the remains, we will use a sampling strategy that causes minimal morphological alteration to the material. We will sand off a small portion of the surface and drill a small hole in the bones/teeth to extract small quantities of bone powder (100 mg for teeth, 200-400 mg for bone). If possible, we will extract bone/teeth-material from different individuals and different sources of each individual to increase the chance of successfully obtaining authentic ancient DNA. All, decisions regarding sampling area and quantity will be made in close cooperation with the museum curators. Sampling will take place at the museum, thus avoiding the removal of skeletal material from its original repository and curatorial authority. The samples will ultimately be destroyed during DNA extraction so that no material will be returned to the museum.

Ancient DNA from humans has the potential to answer a number of important questions including, assessing genetic variation of humans in prehistoric times, assessing population affinities of past populations, their relation to current populations, and assessing the genetic impact of subsequent admixture events. We have recently developed population genetic tools that are tailor-made for analyzing ancient DNA from human remains (Malmstrom et al. 2007) that can; *i*) solve the issue of potential contamination from individuals handling the samples, and *ii*) make inferences from ancient genetic data that are incomplete by using reference samples. We will compare the genetic composition of the ancient KwaZulu-Natal individuals to genome-wide (more than 500,000 genetic markers per individual) reference data from modern populations, including worldwide collections of samples (HGDP – (Jakobsson et al. 2008), HapMap III – (Altshuler et al. 2010)), and sub-Saharan samples (Henn et al. 2011). In particular, we will be able to compare the KwaZulu-Natal ancient DNA data to one of the best current collections of genotyped southern African indigenous populations (Schlebusch et al. 2012). This collection includes eleven populations (220 individuals), of which seven are Khoe and San populations, two are Bantu-speaking populations and two are Coloured groups from different locations; and all individuals have been genotyped for 2.5 million genetic markers. These reference data from modern humans will allow us to compare the ancient DNA from southern African remains to a

broad range of modern humans, including the indigenous inhabitants of southern Africa and potential sources of admixture.

We propose to start our investigation by attempting to extract DNA material from eight sites (Ballito Bay A, Ballito Bay B, Ballito Bay C, Doonside, Mfongosi, Eland Cave, Newcastle and Champaigne Castle) described below. The samples were selected after an initial visit to the Museum by the team of geneticists from Uppsala University, during which material was identified that might provide the best potential DNA results. The Ballito and Doonside samples were selected because they have already been dated to ages that indicate pre-Bantu speaking and pre-colonial populations. The other samples were selected based on their geographic distribution that covers different locations in KwaZulu-Natal. We will independently date them in an effort to start generating a comparable database for dated human remains from prehistoric times in southern Africa together with their genetic data with the aim to assess genetic variation of humans in prehistoric times (pre-Bantu speaking and pre-colonial times), to assess genetic population affinities of past populations, their relation to current populations, and to assess the genetic impact of subsequent admixture events. Where possible, correlations will be made to what is known of the archaeological record.

Ideally, we would like to take three samples for DNA analyses from each individual (two from teeth when possible and one from a bone), in order to be able to replicate our results. Approximately 100 mg of bone powder is needed from each tooth and between 200-400 mg of bone powder from bones is needed for one attempt to extract ancient DNA from these human remains. To visualize how the material is affected by our sampling we show pictures of bones and teeth from 5,000-10,000-year-old Scandinavian material that we have sampled previously by drilling out bone powder from the interior of the samples (Figure 1). Additionally, we show photos of the human remains from KwaZulu-Natal Museum and we have added arrows to the photos as examples of good places to sample for DNA. Since bones from Mfongosi, Eland Cave, Newcastle and Champagne Castle are not dated we furthermore want to apply for extracting additional material to be used for Radiocarbon dating. For the dating of these four sites, a single additional drilling will be considered for each specimen in collaboration with the museum curators. The dating of samples with unknown age will contribute to the completeness of the museum's records regarding some of its human remains, without cost to the museum.

Genetic scientists, Prof Jacobsson and Drs Schelebush and Malmström, will extract the samples and transport them to their specialised ancient-DNA Laboratory in Uppsala, Sweden, where the analyses will be conducted.

References

Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, et al. (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467:52-58

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S (2008) The dawn of human matrilineal diversity. Am J Hum Genet 82:1130-1140

Gilbert MT, Drautz DI, Lesk AM, Ho SY, Qi J, Ratan A, Hsu CH, et al. (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proc Natl Acad Sci U S A 105:8327-8332

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, et al. (2010) A draft sequence of the Neandertal genome. Science 328:710-722

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. Nat Genet 43:1031-1034

Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodriguez-Botigue L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK, Norman PJ, Parham P, Bustamante CD, Mountain JL, Feldman MW (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci U S A 108:5154-5162

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB (2008) Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451:998-1003

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830-838

Malmstrom H, Svensson EM, Gilbert MT, Willerslev E, Gotherstrom A, Holmlund G (2007) More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. Mol Biol Evol 24:998-1004

Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Guldemann T, Kure B, Mpoloka SW, Nakagawa H, Naumann C, Lipson M, Loh PR, Lachance J, Mountain J, Bustamante CD, Berger B, Tishkoff SA, Henn BM, Stoneking M, Reich D, Pakendorf B (2012) The genetic prehistory of southern Africa. Nat Commun 3:1143

Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311:392-394

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463:757-762

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053-1060

Ribot I, Morris AG, Sealy J, Maggs T (2010) Population history and economic change in the last 2000 years in KwaZulu-Natal, RSA. Southern African Humanities 22:89-112

Schlebusch CM (2010) PhD Thesis: Genetic variation in Khoisan-speaking populations from southern Africa. University of the Witwatersrand, Johannesburg

Schlebusch CM, Lombard M, Soodyall H (2013) MtDNA control region variation affirms diversity and deep sub-structure in populations from Southern Africa. BMC Evol Biol 13:56

Schlebusch CM, Skoglund P, Sjodin P, Gattepaille LM, Hernandez D, Jay F, Li S, De Jongh M, Singleton A, Blum MG, Soodyall H, Jakobsson M (2012) Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. Science 338:374-379

Skoglund P, Malmstrom H, Raghavan M, Stora J, Hall P, Willerslev E, Gilbert MT, Gotherstrom A, Jakobsson M (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science 336:466-469

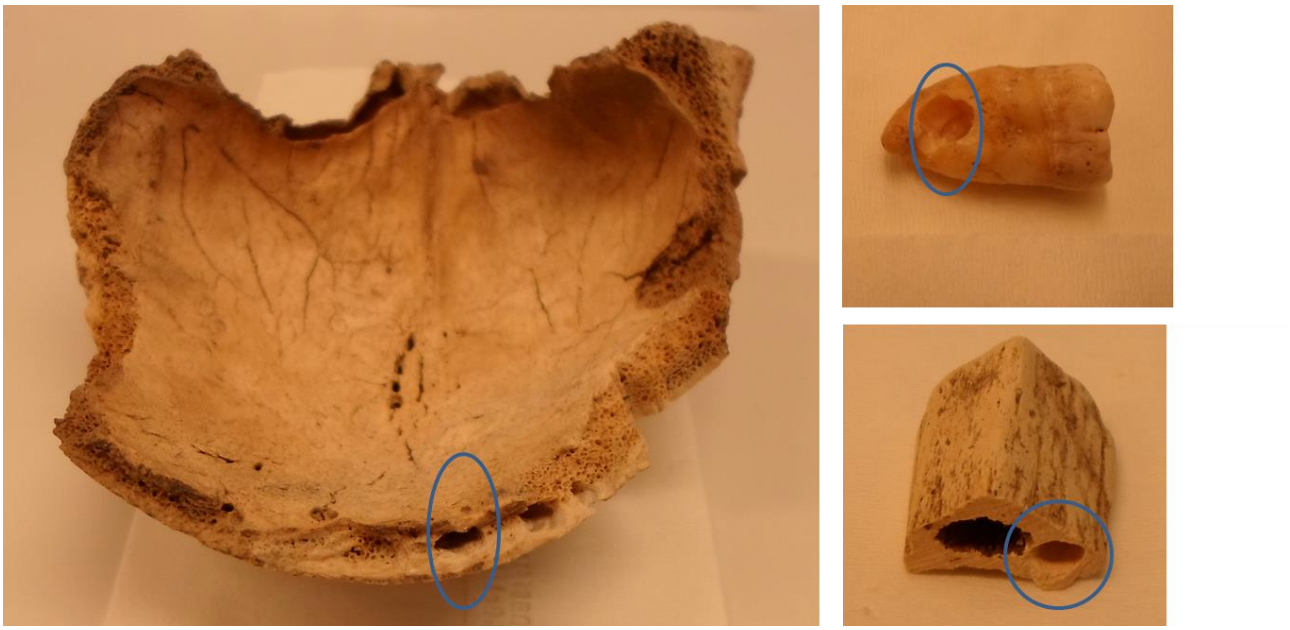
Valdiosera C, Garcia N, Dalen L, Smith C, Kahlke RD, Liden K, Angerbjorn A, Arsuaga JL, Gotherstrom A (2006) Typing single polymorphic nucleotides in mitochondrial DNA as a way to access Middle Pleistocene DNA. Biol Lett 2:601-603

Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, Destro-Bisol G, Soodyall H, Louie L, Hammer MF (2011) An early divergence of KhoeSan ancestors from those of other modern

humans is supported by an ABC-based analysis of autosomal resequencing data. Mol Biol Evol 29:617-630

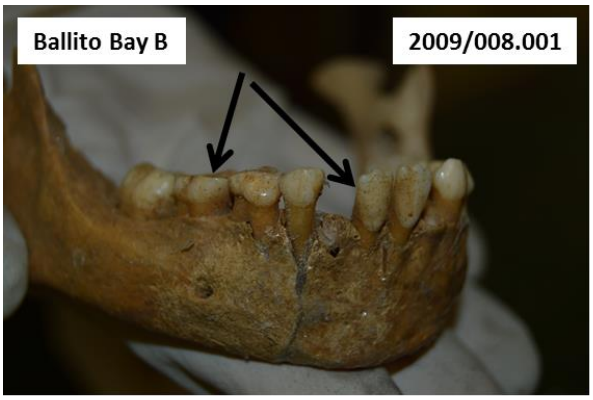
Willerslev E, Cappellini E, Boomsma W, Nielsen R, Hebsgaard MB, Brand TB, Hofreiter M, et al. (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. Science 317:111-114

Figure 1. Example of Scandinavian bones and teeth previously sampled for ancient DNA

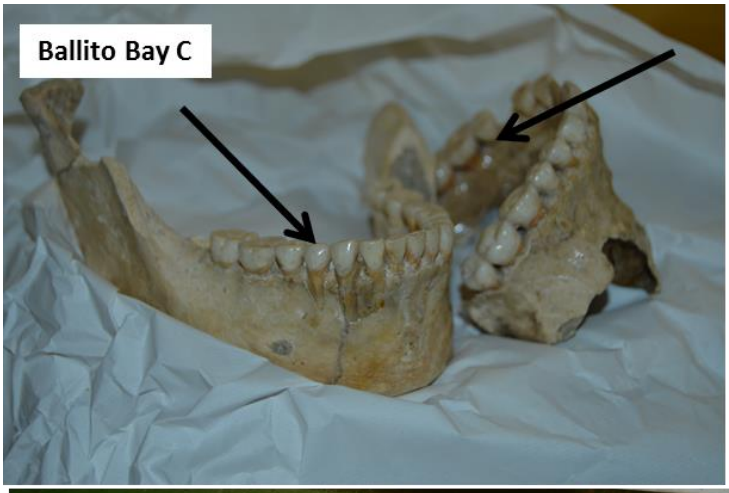


Ballito Bay A
Accession No. 2009/007
National Site No. 2931CA 116
Description: Skeletal remains from one human radiocarbon dated to 1980 ± 20 BP

Ballito Bay B
Accession No. 2009/008.001 and 2009/008.002
National Site No. 2391CA 047
Description: Skeletal material from two human remains of which 2009/008.001 is Radiocarbon dated to 2940 ± 50 BP.



Ballito Bay C
Accession No. 2009/009
National Site No. 2931CA 045
Description: Human remains from one individual Radiocarbon dated to 5900 ± 110 BP.

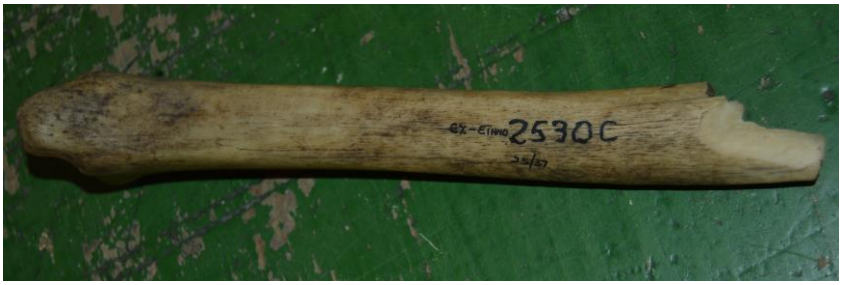


Doonside
Accesson No. 2009/010
National Site No. (not available)
Description: A human skeleton Radiocarbon dated to 2110 ± 50 BP.

No photo available.

Mfongosi
Accession No. 1925/036.002
National site No. 2830DB 023
Description: 5 human teeth, skull, 3 long bones, 1 lower jaw. Not Radiocarbon dated.

Eland Cave
Accession No. 1925/037
National Site No. 2929AB 022
Description: Four human bones found. Not dated.



Newcastle 2378
Accession No. 2007/006.001
National Site No. 2829CA 006
Description: Human remains are in archive. Not dated.

Champagne Castle
Accession No. 2009/023
National site No. (not available)
Description: Partial human skeleton. Not dated.