ARCHAEOLOGY BY MEANS OF DNA: HYPE OR HERE TO STAY?

NEGENENTWINTIGSTE KROON-VOORDRACHT

GEHOUDEN VOOR DE

STICHTING NEDERLANDS MUSEUM

VOOR ANTHROPOLOGIE EN PRAEHISTORIE

TE AMSTERDAM OP 30 MAART 2007

DOOR

PROF. DR. PETER DE KNIJFF

LEIDS UNIVERSITAIR MEDISCH CENTRUM

AFDELING HUMANE GENETICA

GERRIT HEINRICH KROON
(1868-1945)

INTRODUCTION

Suppose you live somewhere in Europe, you are young and eager to learn, and want to know everything there is to know about the history of your country. What you would have done when I was young is simply go to a public library or a specialized library and you would find shelves full of history books stretching for several meters. Now of course, you would initially ignore those dusty shelves and instead use Google or Wikipedia. In any case, within a few hours you can go back in time for many thousands of years and learn everything there is to learn. This plethora of knowledge about our past is something we take for granted without any further thoughts. The knowledge has been around for a long time and we simply can not imagine a life without it.

Still, even now there are countries with no archaeological record. There are still numerous isolated indigenous populations with a concept of the past only going back in time for a limited number of generations. For such countries or such groups, provided they do want to learn more about their past, there are two options: either put the spade in the ground and start digging or initiate a detailed DNA analysis. Here, I would first like to discuss the latter option: how DNA can be used to reconstruct our past.

CONTEMPORARY GENETICS

DNA BASICS

The human genome can be seen as a very complex chemical structure consisting of 3 billion stones (or as we geneticists say, base pairs, abbreviated as bp.). This structure is not one big unit, but consists of 23 different smaller units which we call chromosomes. In each cell of our body we find, tightly packed into the cell nucleus, two sets of 23 chromosomes, one set from each biological parent (Figure 1). Among these chromosomes we find 22 so-called autosomes numbered 1 to 22. In addition there

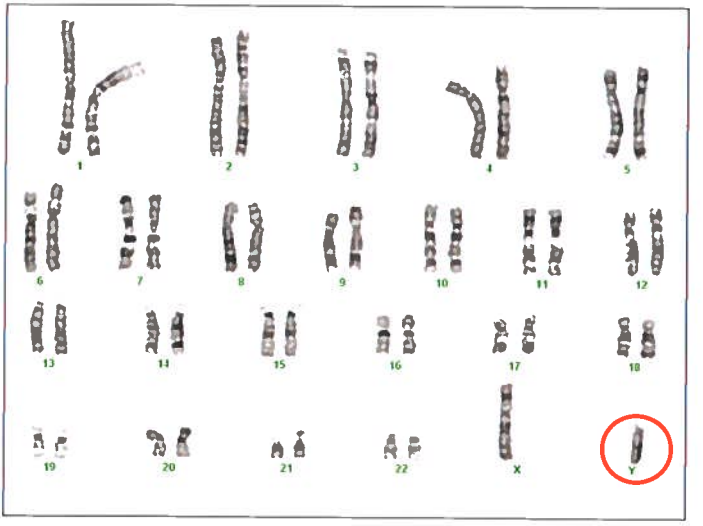

Figure 1. *A karyogram showing the double set of chromosomes present in most human cells. In the red circle the small Y-chromosome is depicted. Because of the presence of an X and a Y chromosome, and a total number of 46 chromosomes, this is a normal karyogram of a human male.*

are two sex-chromosomes called X and Y. Females have two X chromosomes, one of which they pass on to their offspring. Males have an X and a Y chromosome, one of which they pass on to their offspring. If during fertilization the embryo receives an X chromosome from both parents, it will develop into a girl. If the embryo receives an X and a Y chromosome it will develop into a boy. Of course, in the latter case the boy can only receive its Y-chromosome from its biological father.

On average, two human individuals are for 99.95 genetically identical. This still means that, because we have a genome of 3 billion bp., there are about 1.5 million random differences (0.05 x 3,000,000,000). These differences are there because, before each parent can pass on a full set of chromosomes to the next generation, the genome has to be copied and this copy process is not without errors. During each copy step mutations – or spontaneously genetic changes – occur. These mutations happen with a frequency of 2×10^{-8}. Thus, when a parent transmits a full copy of one genome to its child, the copy of the child will differ in a bout 10 – 20 positions from the parents original.
Another inconvenient aspect of this passing on of genetic information from one generation to the other is the random exchange of genetic information between each of the two pairs of all chromosomes 1-22. This process, called recombination, randomizes, given a certain amount of time (generations), the original order of genetic variants. There is, however, a single exception: the Y chromosome. Because males have a single copy of the Y chromosome, its contents can not be recombined, rendering the Y chromosome the perfect tool to reconstruct paternal pedigrees for many generations (Figure 2). Very rarely, also on the Y chromosomes spontaneous mutations occur between two generations. Because of these two proscesses combined (lack of recombination and rare mutations), the Y chromosome is also the ideal tool to reconstruct human evolution and human migration back in time.

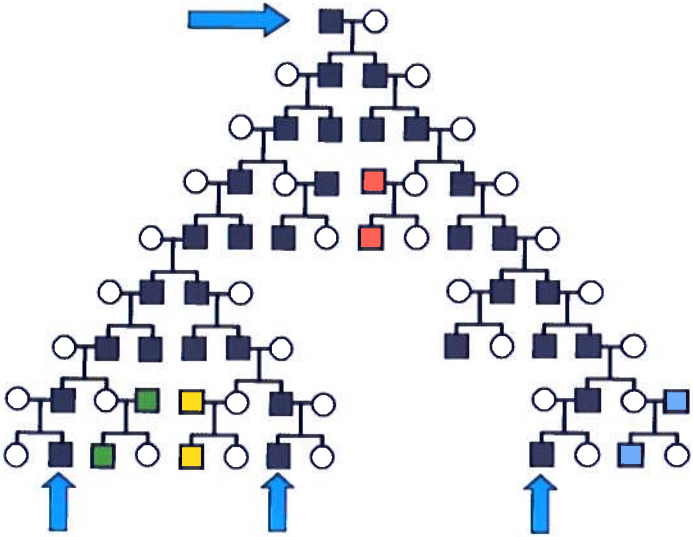

Figure 2. *The transmission of a Y chromosome in a deep rooting pedigree. Males are indicated by squares. Females are indicated by circles. The three males in the most recent generation (indicated by vertical blue arrows) share the same ancestral father 8 generations in the past (indicated by the horizontal blue arrow). The Y chromosome of this male was transmitted along the paternal lineage without any change (follow the dark blue boxes). Other males, marrying into this pedigree have a different Y chromosome indicated by different colors.*

Y CHROMOSOME POLYMORPHISM BASICS

If we would zoom in on a small part of the Y chromosome of a few unrelated males we would see two types of genetic differences (polymorphisms or variants) among these Y chromosomes (Figure 3). First, we would see differences at a single position. These polymorphisms are called single nucleotide polymorphisms (SNPs). Second, we would see stretches of small repeated sequences which are called short tandem repeats (STRs). From many studies we have learned that on the human Y chromosome, SNPs occur

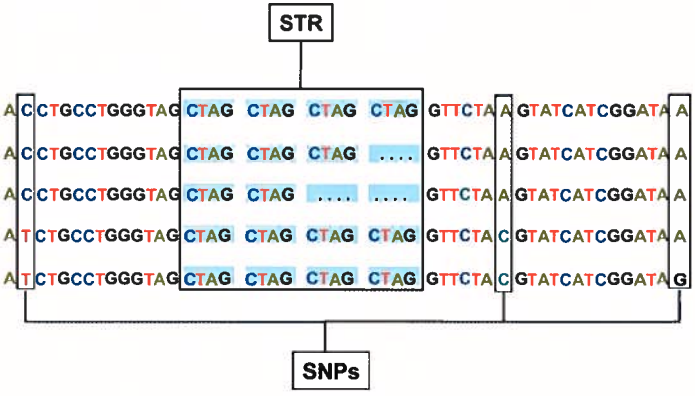

Figure 3. *A small fragment of the Y chromosome of five different males. Shown are two different genetic variations. The first, single nucleotide polymorphisms (SNPs) are genetic variations at a single position. Among the five males we detect three SNPs, from left to right a C/T, an A/C, and an A/G polymorphism. The second type of genetic variation is the short tandem repeat (STR) reflecting variation in the number of short repetitive genome fragments. Shown here is a tetra nucleotide repeat with the repeat motif CTAG repeated 2-4 times among these males.*

much less frequent when compared to STRs. On average, the mutation frequency of SNPs is 2×10^{-8} per generation compared to 0.2% (2×10^{-3}) per generation for STRs (Heyer et al., 1997), a difference of about a factor of 100.000. This makes Y-STRs ideal to infer population differences in a more historical context (say the last 2.000 years), whereas SNPs are more suitable at the

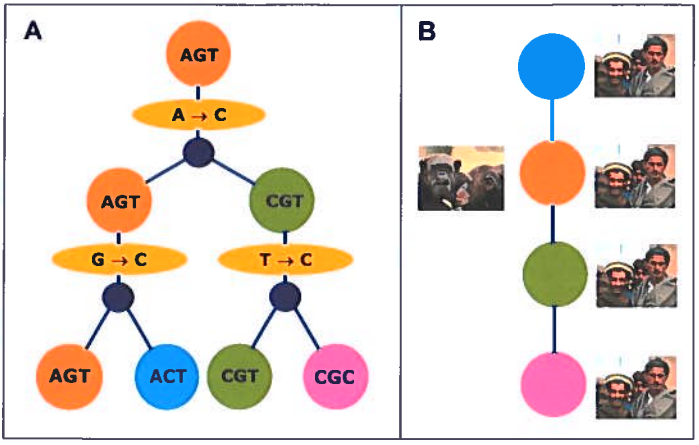

Figure 4. *Reconstruction of the evolutionary tree connecting four different human Y chromosomes characterized by different SNP types. Assuming a minimum number of single SNP mutation steps, the tree in (A) is the most optimal solution. However, from this tree it is impossible to infer which of the four distinct Y chromosome is the oldest or ancestral one. Each of the four variants could be the "founding father". Only when we type the same SNPs in a closely related outgroup of modern humans (here a Chimpanzee) we can reconstruct the ancestral sequence. In this case the tree in (B) indicates that the orange Y has been observed among men and ape, whereas the other types were only observed among humans. Thus, orange is the most likely ancestral type from which others human Y chromosomes arose.*

human evolutionary time scale (the last 200.000 years). However, ultimately, it is the combined use of Y-SNPs and Y-STRs which is most powerful (de Knijff, 2000).

How can we use these two different Y polymorphisms to reconstruct human migration? This is best explained by starting with the very slow mutating SNPs (Figure 4). Out of the first ancestral human Y chromosome, sooner or later a new variant will arise due to the first spontaneous mutation. Later in time, there will be other mutations, all of which lead to new unique Y chromosomes. If we sample human Y chromosomes we can identify such sequence variants and compare them with identical positions of the Chimpanzee Y chromosome. By means of simple logic, and assuming a minimum number of mutation steps, we can build a tree of human Y chromosome variants and identify the oldest – possible founding – Y chromosome (Figure 4). In our theoretical example we postulate that modern humans arose in Eastern Central Africa, and from there migrated into Western Asia. All males migrating out of Africa carried the same Y chromosome (Figure 5, the orange Y chromosome). While these males migrated out of Africa they accumulated new Y chromosome variants at different stages of their migration (green, blue, and pink in Figure 5). Assuming that there was no back migration, it can be easily seen from Figure 5A how Y chromosome variation among males sampled globally could result in region specific variants. However, Figure 5A is seriously misleading because it assumes that the connections between the dots (populations) reflect the original connections (i.e. routes of migration). What we actually only know for certain in modern evolutionary genetics is illustrated in Figure 5B: a number of populations with different Y chromosomes. Based on this information we would like to be able to reconstruct the most likely migration routes connecting the populations we sampled. Thus, how can we reconstruct the routes indicated in Figure 5A from information as in 5B. In order to do so with some accuracy we need to know (i) the evolutionary tree connecting the

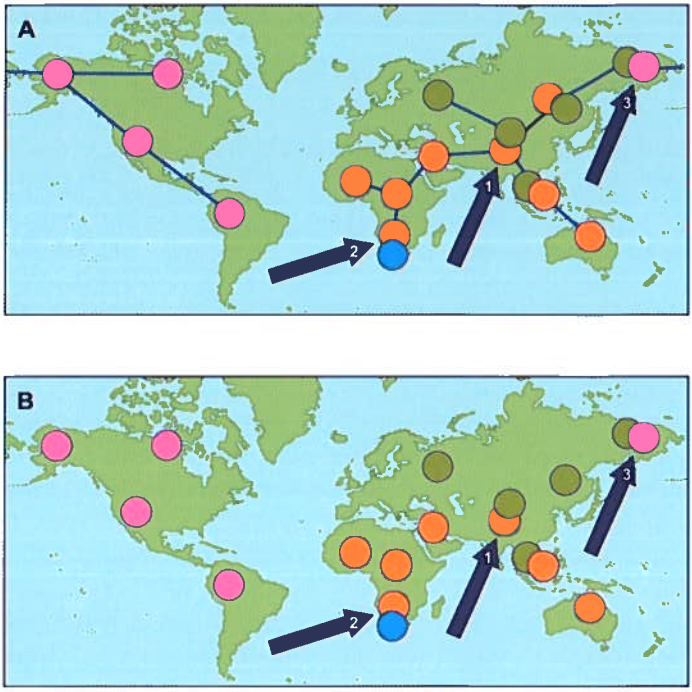

Figure 5. *(A) Theoretical model of the spread of modern humans out of Africa reconstructed by means of Y chromosome variation. All males leaving Africa had the same Y chromosome (indicated in orange). At three moments in time (see the numbered arrows) a new Y chromosome, indicated with different colors, arose due to spontaneous mutations. These new Y chromosomes are only observed at locations where it first occurred and where humans migrated to, not where they came from. Once we have reconstructed the tree of Y variants (Figure 4), our model makes perfect sense. (B) The big problem is that in real life we only have a map of globally dispersed Y variants. Only when we can date each of the three new variants accurately, we can start with the reconstruction of the migration routes. In addition, the more dense our modern sampling is, the better we can also indicate the region of origin of each new Y chromosome variant.*

Y chromosome variants, (ii) an estimate of the age of each new Y chromosome variant, and (iii) the place of occurrence of each new Y chromosome variant. If these conditions are met, a reconstruction of the human migration processes back in time solely on the basis of present day male sampling is relatively easy. The most reliable approach involves the combined use of Y-SNPs and Y-STRs (de Knijff, 2000). On the basis of Y-STR variation it is possible to infer the age of Y-SNPs (Figure 6). Ideally, all

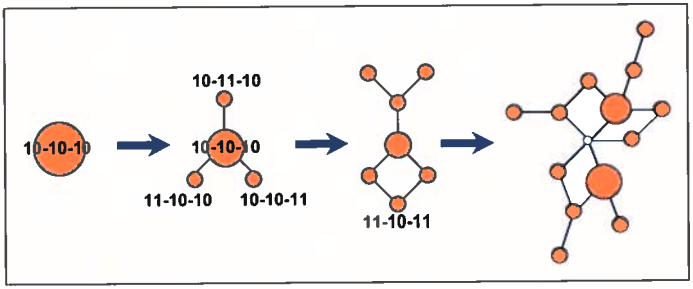

Figure 6. *(A) In order to date the ages of each new Y SNP variant we can make clever use Y STR variation. By means of deep rooting pedigree studies and the analyses of father-son pairs a fairly accurate Y-STR mutation rate of 0.2% could be obtained. This means that a single STR locus will spontaneously increase in length or decrease in length once every 500 generations. Normally these mutations involve a single repeat step. From this it easily follows that if we were able to follow the growth of Y-STR variation we would be able to detect a process as shown here. Out of a single Y chromosome with three STR loci of 10 repeats long, due to spontaneous mutations new Y chromosomes emerge, all with a slightly different Y-STR composition (in this example each of the STR loci mutated once from 10 to 11 repeats). Given time, slowly a more complex Y-STR network grows, containing Y chromosomes connected to each other with single STR repeat steps. Obviously, assuming a certain average Y-STR mutation rate we can estimate the number of generations it takes to accumulate a certain Y-STR variation out of a single Y chromosome.*

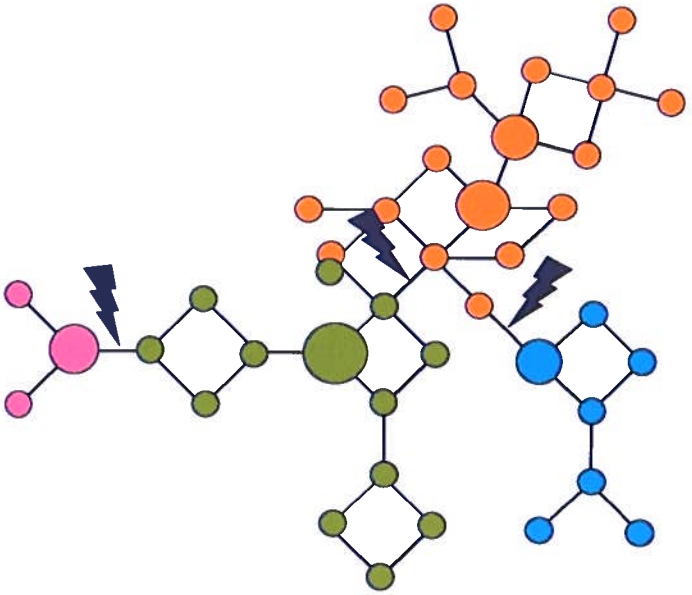

Figure 6. *(B) In real life, of course, a Y chromosome can accumulate both SNP variation and STR variation. Here we see this happening. The flashes indicate a new SNP due to a spontaneous mutation. This new SNP marks the beginning of the growth of a new Y-STR network connecting Y chromosomes which all share the same SNP background. Within each of these subgroups we can use the Y-STR variation to estimate the age of the marking SNP. What is also illustrated here is the notion that the younger the age of a SNP, the simpler the Y-STR network is. It is important to stress here that this rather simple approach only works in the absence of substantial immigration and only in the original population where the SNP arose. Dating an Asian specific SNP based on typing Chinese cooks in restaurants in Holland will not work because these cooks will certainly not represent the original population of origin.*

population studies should make use of the same set of Y-STRs and Y-SNPs. For this, there is also a commonly used nomenclature: different Y-chromosomes defined only by Y-SNP differences are called haplogroups, different Y chromosomes defined only by Y-STR differences are called haplotypes, and the combined use of Y-SNPs and Y-STRs allow the identification of Y-lineages (de Knijff, 2000).

Y CHROMOSOME POLYMORPHISMS IN REAL LIFE

As always, the best proof for a pudding is eating it, or, in our case, are there indeed sufficient Y-SNPs and Y-STRs available for our research purposes? To be short, yes. Because of a lot of effort by a small group of Y chromosome enthusiasts (Hammer et al., 1997; Jobling and Tyler-Smith, 2000; de Knijff et al., 1997; Roewer et al., 1996; Underhill et al., 2000), we do have at least 300 YSNPs and 100 Y-STRs available. On the basis of Y-SNPs a very robust human Y chromosome tree was reconstructed (Figure 7), with 18 major branches (Figure 7A) and 58 subgroups (Figure 7B). Since 2000, many detailed Y-SNP studies using the same Y haplogroup nomenclature (YCC, 2002) have been published. In addition, we have also screened a lot of globally dispersed males for Y-SNPs. All this information was collated into a Y-SNP database containing data from ~900 populations and close to 50.000 males (Figure 8). Here I discuss the first results based on information from ~43.000 males.

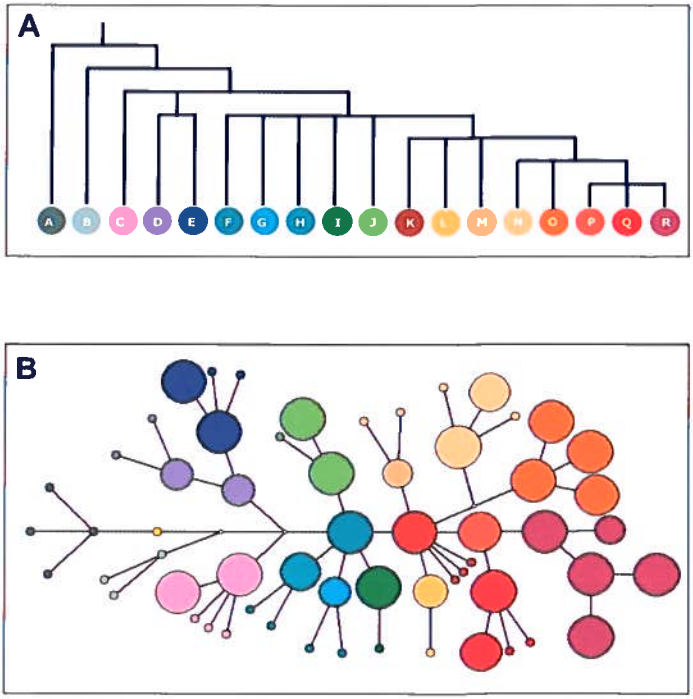

Figure 7. *(A) Simplified Y-SNP tree in which the 116 distinct Y haplogroups from the first complete Y-SNP tree (Underhill et al., 2000) are condensed into the 18 major Y haplogroups. The Y-haplogroup names A-R are according to the YCC consensus nomenclature (YCC, 2002). (B) These 18 major haplogroups can be further divided into 58 common subgroups. Here the diameter of each colored circle corresponds to a somewhat inflated relative frequency based on Y-SNP information from ~43.000 globally dispersed males (our own database). Small circles represent relative frequencies of 0.1% or less. The large circles all have a relative frequency of 1% or more (thus observed in a minimum of 430 males). Note that the colors in 7B correspond with those in 7A.*

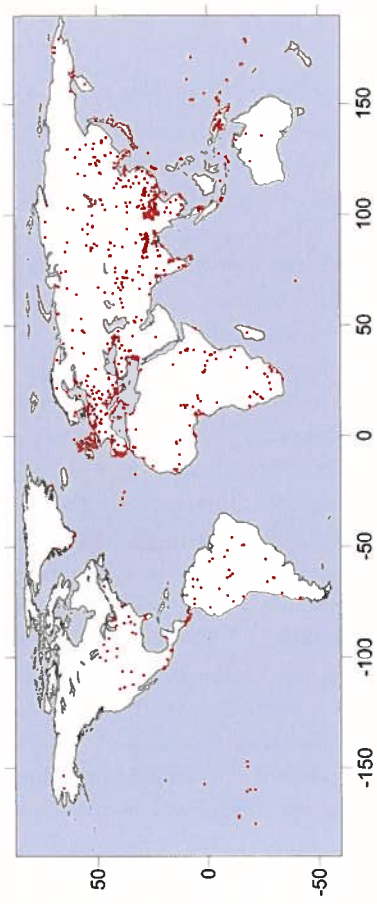

Figure 8. *Distribution of the ~900 populations included in the FLDO Y-SNP database. Only indigenous populations are included in this database. For example, information from Europeans or Africans in the USA is not included. Y-SNP information from a total of ~43.000 males is included in our Y-SNP database. Each dot represents a single population sample. It is obvious that information from large parts of the world is still missing. This includes, rather surprising, also France, one of the largest European countries with virtually no documented Y-SNP information.*

If we further simplify the Y-SNP tree (Figure 9) and plot all regions where the Y haplogroups A - F are currently present in frequencies of 10% or more, it immediately clear that the first two major Y-haplogroups A and B only occur among Africans, predominantly south of the Sahara. Two other haplogroups, C and D only occur outside Africa, and both have a scattered distribution. Haplogroup E occurs in rather high frequencies (up to 50 %) in all African populations, but also in all countries bordering the Mediterranean Sea. Finally, haplogroup F, combining all other Y haplogroups, predominantly occurs outside Africa, but does also occur among northern African populations, most likely due to relatively recent migration into Africa. Assuming a Y-SNP mutation rate of $1.2*10^{-9}$, and a Human-Chimpanzee divergence of 5 million years ago, I estimate the times back to the most recent common human male ancestor of 84.000 ± 27.500 years. This puts the age of the M168 mutation at 55.000 ± 17.800 years, and of the M89 mutation of 46.000 ± 16.200 years. From this, and the specific global distribution of these six basic Y haplogroups we can conclude a few things. First, modern humans indeed started to expand in Africa about 84.000 years ago. Second, The M168 mutation arose within Africa around 55.000 years ago, shortly before the out-of-Africa expansion. Third, the haplogroups A and B are the oldest Y-haplogroups and males carry-

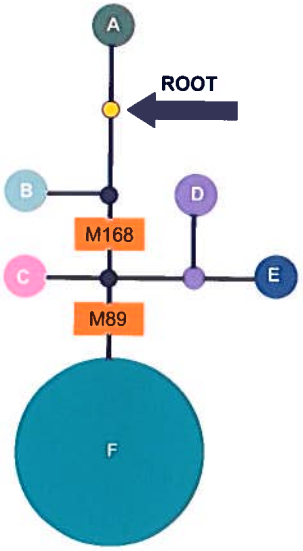

Figure 9.

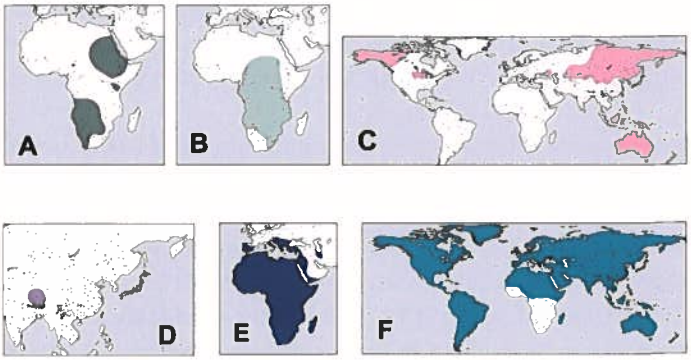

Figure 9. *A simplified Y-SNP tree (see also Figure 7) of the six major Y haplogroups and maps showing their global distribution. From this condensed tree and the corresponding maps, a complex picture, illustrating to the out-of -Africa expansion of modern humans, emerges. Although there are a number of different possible scenario's, the most parsimonious (simple) model is the one where M168 – the Y-SNP marking all Y-chromosomes outside Africa – occurred in Africa shortly before humans migrated into Asia. The migrating males carried two different Y-chromosomes: (i) those with the with M168, but still not developed into C, D+ E, and F, and (ii) those with the immediate ancestral type of D and E, but still not developed into either D or E. What remained in Africa where those with haplogroup A, haplogroup B, and also part of those with the immediate ancestral type of D and E, but still not developed into either D or E. Once outside Africa, the M168 chromosomes soon accumulated new mutations leading to the haplogroups C and F. Also, from the DE ancestral type haplogroup D arose. Within Africa, haplogroups A and B further diversified and from the DE ancestor the very successful African haplogroup E arose. The presence of haplogroup F in Northern Africa with low frequencies is best explained by a relatively recent back-migration.*

ing these two haplogroups never left Africa. Fourth, haplogroups C, D, and F, arose outside Africa, not long after the out-of-Africa migration. Fifth, The complete lack of haplogroup C and D carrying males in Africa, indicates that around the time of the out-of Africa migration, no – or very little – back migration occurred.

How exactly these first humans migrated out of Africa is still very much open for debate, although slowly, a clear picture starts to emerge from the accumulating data. Based on detailed Y chromosome studies, mtDNA studies, and archaeological data (Mellars P, 2006a, 2006b; Richards et al., 2006; Sengupta et al., 2006; Underhill et al., 2001, 2003) it becomes more and more clear that early modern humans could have followed a coastal route along the southern fringes of the Arabian peninsula and the Asian continent. They soon arrived in the Indian subcontinent where they successfully settled. From this homeland they migrated into the Asian subcontinent in different directions at different timescales. Although it is difficult to see in Figure 9, two ancient non-African haplogroups C and D do occur in India, although very local and in very low frequencies. Also, haplogroup F and some of its sublineages are common throughout India. This strongly suggests that indeed, already 50.000 years ago, modern humans were present in India. Exactly how the haplogroups C, D, and F spread over Asia and beyond is still unclear because of lack of data from vast and important regions (Tibet, Mongolia, Siberia, and Central Asia) and insufficient detailed genetic information from many of the already sampled populations. We do know, however, that within Asia, the Himalayan mountain range must have been a formidable geographical barrier. Our own research (unpublished, but see below) categorically rules out any substantial ancient migration from south to north through this region. Accordingly, early modern humans must have migrated from India into the rest of Asia around the Himalayas, either though South-East Asia, through Central Asia, or both.

A GENETIC GIFT FROM THE PAST

As stated before, contemporary DNA research, as outlined above, can sometimes be an ideal method to reveal something from your unknown ancestors. We can illustrate this with an example from the Himalayan region which we have analysed in detail.

Although there are some claims for archaeological artefacts dating back to 2000 BC, and some historians have theorised that a state of such kind might have existed between 500 BC and 600 AD, the first transcribed evidence for Bhutan as an entity comes from around 800 AD, mainly from early Buddhists transcripts stating Tibetan origins. Before this date, almost nothing reliably is known. We were able to study the genetics of modern Bhutanese, as part of a big European Science Foundation project "*Languages and genes from the greater Himalayan region*". One of the goals of this project was providing more clarification into the early human settlement of modern humans in Bhutan. Crucial in this respect is the understanding that all of Bhutan lies immediately South of the highest peaks of the Himalayas. From North to South, there is an altitude gradient from ~7500 to ~ 300 meters above sea level over a distance of just under 200 kilometres (Figure 10). In the North, crossing into Tibetan/Chinese territory is extremely difficult since there very few passes, and none of them are easy-going. In the South, the extensive sub-tropical Bhutanese forests are abruptly replaced by the almost completely cultivated Indian Duars.

On the basis of a detailed Y-SNP survey, we discovered that among the Bhutanese (and most of the Nepalese) populations, haplogroup O3e was by far the most frequent one, with frequencies exceeding 50%. Just across the border, in India, O3e is exclusively found among a limited number of isolated aboriginal populations. O3e is completely absent among the vast majority of Hindu and Muslim Indians. In contrast, in India haplogroup H is one of the most important ones. In Nepal, H is only seen in among relatively recent Hindu and

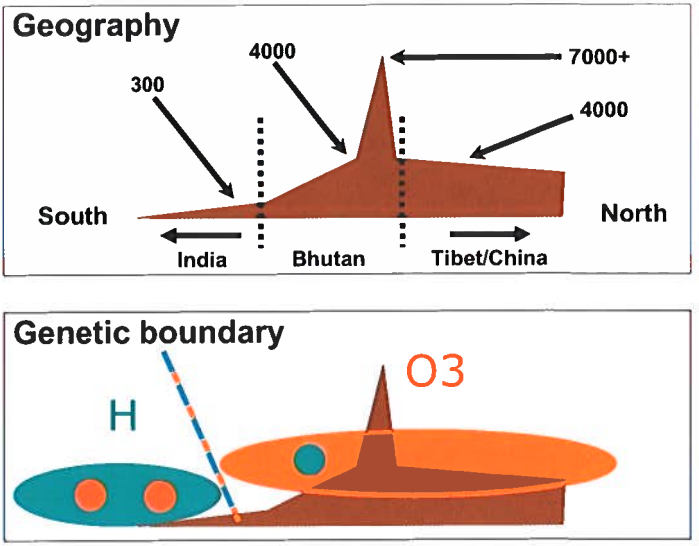

Figure 10. *The geography of Bhutan and neighboring countries. Also shown is the distribution of two Y-haplogroups O3e and H which show a remarkable discontinuous distribution. In the vast territory of O3e, the region north of the Indian/Bhutanese border and stretching over almost all of Tibet and China, there is only one small isolated population (in the centre of Bhutan) with the predominantly Indian-subcontinent specific haplogroup H as the most frequent Y-haplogroup. In present day India, there are a number of isolated and aboriginal populations with almost exclusively haplogroup O3e, reflecting a ancient close genetic relationship with their Tibetan ancestors.*

Muslim immigrants in the south. In Bhutan, haplogroup H is completely absent, apart from one small isolated population in the centre. When mapped together (Figure 11), these two haplogroups nearly perfectly fit together without any substantial overlap, thus demarcating the perfect genetic border. What is puzzling is the position of this

border, 200 km. South of the most obvious geographical border in this region, the highest Himalayan peaks. What we are relatively sure of, is the origins of nearly all Nepalese and Bhutanese populations. They have a close genetic connection with populations from Tibet and further east in Asia, and not, as could be expected on the basis of geography and the antiquity of modern humans in this region, from the Indian subcontinent. Why we see this, and when and how humans migrated into Bhutan and Nepal is still open for discussion, and awaits further clarification. For this we need substantial more genetic information from this entire region. At least, our preliminary results already provide a good indication for the people of Bhutan and Nepal with respect to their remote ancestors.

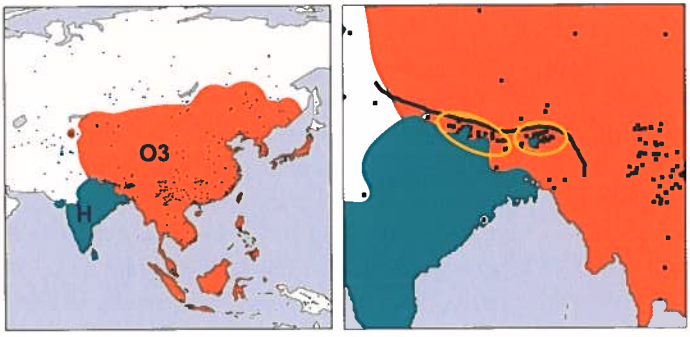

Figure 11. *Maps of South Asia and East Asia showing the distribution of two Y-haplogroups O3e and H. In the right panel the approximate position of the highest peaks in the Himalayas (indicated with a dark blue line) and of Nepal and Bhutan (yellow ellipses) are shown. Although not completely exact, the position of the border separating both Y-haplogroups runs roughly parallel with the Indian/Nepalese and Indian/Bhutanese borders. The exception is formed by a few east Indian aboriginal populations which are also much closer related to Bhutanese populations on the basis of language and culture. Such a sharp genetic border is very unique. In this figure, each small dark square is a single population from which Y-haplogroup data could be collected.*

DIRTY HANDS AND CLEAN PROFILES

When compared with contemporary DNA research, ancient DNA research is very complex and often very unreliable. Many experts will agree with me that DNA research on ancient skeletal remains is as difficult and challenging as it can get in human genetics (Gilbert et al., 2005). This is primarily caused by a number of factors:

1. DNA in ancient skeletal remains is heavily degraded and often bound to skeletal matrix.
2. Ancient skeletal remains are heavily contaminated with non-human DNA (e.g. from plants and, bacteria).
3. Most of the ancient skeletal remains in collections are heavily contaminated with contemporary DNA.

The first two factors can be easily recognised once you start with ancient DNA research. However, the third factor is sometimes impossible to identify for the simple reason that those who handled the bones have DNA profiles which can be nearly identical to the profile of the bone itself. In this respect, a Neanderthal bone would be easier to analyse because it has a recognisable different DNA profile. Despite these difficulties, DNA research on ancient remains of modern humans is becoming more and more popular. If performed under stringent conditions, it is possible to obtain reliable DNA profiles, and from these we can learn a lot. Here, I will not explain in detail how ancient DNA should be done. For that there are numerous sources. I will, however, highlight a number of exciting possible applications. Below, I will draw a seemingly unrealistic optimistic picture of the future of the field of archaeogenetics. For the uninitiated I will shortly try to explain why I am so overtly optimistic.

Until recently, ancient DNA research involved a lot of manual handling steps and complex molecular biological techniques including cloning and sequencing many individual clones. In short, it was extremely hard work resulting in very little sequence information. Another limitation was the lack of sensitivity. Essentially,

only mtDNA sequence information could be extracted from ancient bones with some reliability. Without doubt, the first ever published mtDNA Neanderthal sequence information (Krings et al., 1997) marked the beginning of a new phase in ancient DNA research. Although these results were obtained the old fashioned way, for the first time routine ancient DNA analyses seemed possible. A second major breakthrough was the release of a completely new DNA sequence technology. In 2005, the USA-based company 454 published a revolutionary new approach (Margulies et al., 2005). Soon, it became obvious that this technique would revolutionise ancient DNA research. Less than a year later this prophecy became truth by means of the first two studies revealing genuine ancient autosomal DNA sequence information from Mammoth (Poinar et al., 2006) and Neanderthal (Green et al., 2006). These studies will without doubt pave the way to many ancient DNA studies reporting not only mtDNA data, but also autosomal DNA from unexpected sources.

Why is this new technique so soon so influential? To cut everything short, with this method it is possible to generate reliable sequences from any biological source, without suffering from the fact that most of the ancient DNA is severely degraded into fragments of 20-150 bp. More specifically, this technique can only be used for such extremely short fragments, and with this, seemed like specially designed for ancient DNA studies (which was not the case). Another major benefit is the fact that up to 250.000 individual short sequence reads can be obtained in a single experiment which can be completed in only a few days. In the good old days of cloning this would simply not be possible in terms of years!

CONTRASTS OVER TIME

Once we have sufficient DNA results from ancient modern humans at our disposal, we can learn a few important things. First of all, it will be possibility to identify spatiotemporal changes in the genetic struc-

ture of a regional population. Contemporary population genetics, as described in the first part of this document largely assumes that human populations with their present day distribution reliably reflect their original historical/ ancient distribution. This is of course a rather naïve assumption which will be violated at many places. For this, there is some archaeological evidence from central Asia, present day China, Africa, but also from Europe. Even in recent times, say the past 2000 years, we know that large scale population migration events (either forced or voluntary), have more or less destroyed the original population structure in eastern Asia. It will be fascinating to be able to reconstruct the original human founders in situ. Although it will still take some time, eventually this information will lead to a better understanding of human migration.

Second, results from ancient DNA studies, given the ideal study design, will also provide a better understanding of the flexibility of the human genome to accumulate new genetic variants. These could be either beneficial or detrimental for human health. This approach can be illustrated by means of the recently started collaborative project between the department of archaeology of the city of Eindhoven and my laboratory. For this project, well over 400 primary skeletons were excavated in font of the Catharinakerk in the centre of Eindhoven. During excavation, multiple dental elements were removed from each skeleton, under very stringent DNA contamination prevention measures. The skeletons at this site cover a timeframe from ca. 1250 to 1850 AD and will serve a genetic reference collection, together with samples from present day citizens of Eindhoven. In theory, the scientific approach is rather simple (Figure 12), although to my knowledge it has never been tested at the scale we are now working on. It is still too early to present hard evidence, but from a small pilot project performed in collaboration with Prof. Joachim Burger from Mainz, and the company 454, we already know that we can obtain sufficient DNA from skeletons covering the entire time scale of the Catharinakerk project.

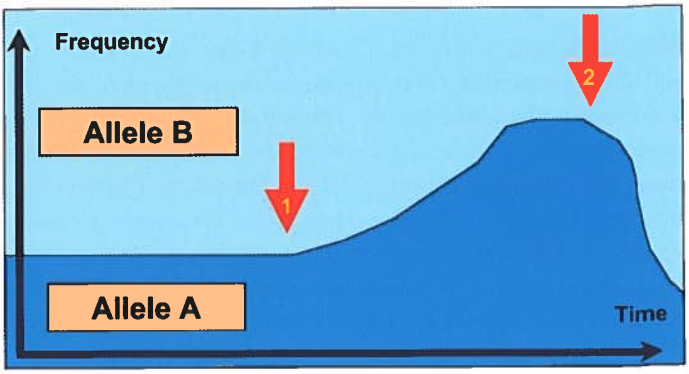

Figure 12. *An explanation of the basic principle for the identification of evolutionary selection pressures. Over a long period of time we have measured the allele frequencies of a single genetic variant in the human genome. For this, we could in theory make use of the DNA isolated from all the skeletons sampled for the Catharinakerk project. This would cover a period of nearly 800 years or 32 human generations (assuming a generation time of 25 years). In our earliest samples there is a certain balance between the frequencies of the two alleles of a single polymorphism. At a certain moment in time (indicated by the 1st red arrow) a yet unknown factor results in an sudden increase of the frequency of allele A, at the expense of the frequency of allele B. At a later moment in time, this process was reversed leading in the most recent samples to a allele frequency distribution roughly similar to the starting point. Without DNA samples collected over a long time, we can not observe such processes. What the factor at time 1 was could perhaps be deduced from detailed research on the skeletons. So far, this approach is purely theoretical, but the Catharinakerk project is very likely to change this completely, as it is the first project of such kind.*

CONCLUSIONS

Although only time will tell, I have no doubt that DNA research will revolutionise archaeology in many respects. First, and foremost, from now on, any self-respecting archaeologists should take stringent precautions while excavating ancient human remains. This involves more than simply wearing surgical gloves. For this, the Catharinakerk project is the golden standard. Human skeletal samples excavated otherwise will be contaminated. This is a given fact. As such, these samples should be given a low priority for DNA research. Careful thought should be given to these issues prior to any excavation where human remains are to be expected. This should also include preserving a DNA sample of everyone participating in the excavation. If discovered accidentally, the complete excavation should be stopped, at least with respect to isolating the skeleton. Finally, sufficient and appropriate storage capacity should be created, in order to properly retain samples for future DNA analyses.

The new DNA sequencing techniques are ruthlessly sensitive. This has the major advantage that any sufficiently intact DNA molecule will be retrieved and can be converted into a DNA sequence. Of course, the major disadvantage will be the detection of even minute (think of single cells!) amounts of contemporary DNA contamination. For this, briefly touching or keeping the skull in front of you is sufficient. If levels of contamination are high in comparison to the levels of authentic DNA, it will still be nearly impossible to decipher which is what. As such, these new DNA techniques definitively mark the end of the good old days with a nice skull in your dirty hands!

The exciting possibilities to use ancient DNA to learn more about our genetic past will for sure pave the way for archaeoepidemiology. For me personally, in this respect the Catharinakerk project is like a dream come true. What we will learn from this in terms of relevance for medical genetics it is no longer limited by the techniques and the samples. It is a simply a matter of sufficient funds and time.

Peter de Knijff
Leiden University Medical Center

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to the Board of "De stichting Nederlands museum voor anthropologie en praehistorie" for their kind invitation to present this 29th Kroon lecture, which I see as a very prestigious event. Ancient DNA is getting more and more popular, and the technology enabling it improves substantially every year. For the text I have borrowed heavily from the many projects I am currently involved with. Data are collected by hard work of all the people of the laboratory I am supervising: the Forensic Laboratory for DNA research, Department of Human Genetics, Leiden University Medical Center. The Himalayan project would not have been possible without the intense and constructive collaboration with Prof. Mark Jobling and Dr. Emma Parkin (University of Leicester, UK), Dr. Chris Tyler-Smith and Denise Carvalho-Silva (Sanger Institute – Welcome Trust, Hixton, UK), Prof. Guido Barbujani (University of Ferrara, IT), and Prof. George van Driem, Leiden University. The archaeogenetics project Catharinakerk, Eindhoven, marks the realisation of one of the scientific dreams of Drs. Nico Arts and myself and would not have been possible by the many volunteers involved in this project, the moral and financial support of the former Eindhoven alderman Mrs. Drs. M. Schreurs, and financial support provided by the former Dutch Minister of Science and Education Mrs. M. van der Hoeven. I am solely responsible for everything in writing and in graphics. All mistakes are mine, and I will not correspond about them.

BIBLIOGRAPHY

GILBERT MT, BANDELT HJ, HOFREITER M, BARNES I. Assessing ancient DNA studies. Trends Ecol Evol 2005; 20: 541-4.

GREEN RE, KRAUSE J, PTAK SE, BRIGGS EW, RONAN MT, SIMONS JF, DU L, EGHOLM M, ROTHBERG JM, PAUNOVIC M, PAABO S. Analysis of one million base pairs of Neanderthal DNA. Nature 2006; 444: 330-336.

HAMMER MF, SPURDLE AB, KARAFET T, BONNER MR, WOOD,ET, NOVELLETTO A, MALASPINA P, MITCHELL RJ, HORAI S, JENKINS T, et al. The geographic distribution of human Y chromosome variation. Genetics 1997; 145: 787–805.

HEYER E, PUYMIRAT J, DIELTJES P, BAKKER E, DE KNIJFF P. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 1997; 6: 799-803.

JOBLING MA, TYLER-SMITH C. New uses for new haplotypes the human Y chromosome, disease and selection. Trends Genet 2000; 16: 356–362.

DE KNIJFF P, KAYSER M, CAGLIÀ A, CORACH D, FRETWELL N, GEHRIG C, GRAZIOSI G, HEIDORN F, HERRMANN S, HERZOG B, HIDDING M, HONDA K, JOBLING M, KRAWCZAK M, LEIM K, MEUSER S, MEYER E, OESTERREICH W, PANDYA A, PARSON W, PENACINO G, PICCININI A, PEREZ LEZAUN A, PRINZ M, SCHMITT C, SCHNEIDER PM, SZIBOR R, TEIFEL GREDING J, WEICHHOLD G, ROEWER L. Chromosome Y microsatellites: population genetic and evolutionary aspects. Int J Legal Med 1997; 110: 134-140.

DE KNIJFF P. Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 2000; 67: 1055–1061.

KRINGS M, STONE A, SCHMITZ RW, KRAINITZKI H, STONEKING M, PAABO S. Neandertal DNA Sequences and the Origin of Modern Humans. Cell 1997; 90: 19–30.

MARGULIES M, EGHOLD M. et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005; 437:326-7

MELLARS P. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science. 2006; 313:796-800.

MELLARS P. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. Proc Natl Acad Sci U S A; 2006 103: 9381-9386. Erratum in: Proc Natl Acad Sci U S A 2006; 103:13560.

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. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 2006; 311:392-394.

RICHARDS M, BANDELT HJ, KIVISILD T, OPPENHEIMER S. A model for the dispersal of modern humans out of Africa. In: Bandelt HJ, Macaulay V, Richards M. Eds. Human mitochondrial DNA and the evolution of homo sapiens. Springer-Verlag. Berli, Heidelberg. 2006.

ROEWER L, KAYSER M, DIELTJES P, NAGY M, BAKKER E, KRAWCZAK M, DE KNIJFF P. Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related populations. Hum Mol Genet 1996; 5:1029-1033. (with corrigendum in Hum Mol Genet 1997; 6:828).

SENGUPTA S, ZHIVOTOVSKY LA, KING R, MEHDI SQ, EDMONDS CA, CHOW CE, LIN AA, MITRA M, SIL SK, RAMESH A, USHA RANI MV, THAKUR CM, CAVALLI-SFORZA LL, MAJUMDER PP, UNDERHILL PA. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. Am J Hum Genet 2006;78: 202-221.

UNDERHILL PA, SHEN P, LIN AA, JIN L, PASSARINO G, YANG, WH, KAUFFMAN E, BONNE-TAMIR B, BERTRANPETIT J, FRANCALACCI P, et al. Y chromosome sequence variation and the history of human populations. Nat Genet 2000; 26: 358–361.

UNDERHILL PA, PASSARINO G, LIN AA, SHEN P, MIRAZON LAHR M, FOLEY RA, OEFNER PJ, CAVALLI-SFORZA LL. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 2001; 65:43-62.

UNDERHILL PA. Inferring human history: clues from Y-chromosome haplotypes. Cold Spring Harb Symp Quant Biol 2003; 68:487-93.

YCC. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 2002; 12: 339–348.

KROONVOORDRACHTEN

REDACTIE: W.H. METZ

Amsterdams Archeologisch Centrum
van de Universiteit van Amsterdam

Printed by

JOH. ENSCHEDÉ AMSTERDAM BV